LINEAR-IN-FLUX-EXPRESSION (LIFE) APPROACH TO DYNAMIC BIOLOGICAL NETWORKS

By

NATHANIEL J. MERRILL

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Computational and Integrative Biology

Written under the direction of

Professor Benedetto Piccoli

And approved by

_____

Benedetto Piccoli

_____

Kwangwon Lee

_____

Angélica González

_____

Monique Chyba

Camden, New Jersey

May 2020

ABSTRACT OF THE DISSERTATION

Linear-in-Flux-Expression (LIFE) Approach to Dynamic Biological Networks

by NATHANIEL J. MERRILL

Dissertation Director:

Benedetto Piccoli

This work analyzes the dynamics of three distinct classes of biological systems. The first is metabolic networks. The methodology named LIFE (Linear-in-Flux-Expression) was developed with the purpose of studying and analyzing large metabolic systems. With LIFE, the number of model parameters is reduced by accounting for correlations among the components of the system. These systems can be associated to graphs. General results on the stability of LIFE systems are discussed, particularly we formulate necessary conditions on the graph's structure to ensure the stability of the dynamics. Moreover, stability analysis from related fields, such as Markov chains, network flows, and compartmental systems, can also be applied. Control of LIFE systems through the addition of drugs as well as modifying intakes is discussed. A generalized graph object which incorporates hyperedges and uberedges is used to apply LIFE to metabolic networks, in particular to Mycobacterium tuberculosis (MTB). Results from LIFE simulations on MTB carbon metabolism are presented via simulations. Finally, the method allows us to rank 4-drug combinations in terms of their effectiveness in destabilizing MTB metabolic networks, thus killing the bacterium.

The second class of systems is models for circadian rhythm. One of the essential characteristics of an authentic circadian clock is that the free-running period sustains

an approximately 24-hour cycle. The dynamics of the circadian clock is modified by an external stimulus, called a zeitgeber. This modification process is known as entrainment and operates to reset the phase and period of the circadian clock. When analyzing the phase of entrainment of many individuals, it is often assumed that an organism with a short period will have a phase advance, and a prolonged period will have a phase delay; however, this does not explain all known experimental data, so a Two-Step Entrainment model was developed. This work analyzes how parameters of the model affect the dynamics and presents results fitting the Two-Step Entrainment model to human data.

The third class of systems consists of ecological networks. The interactions of species are often described via a network. Construction of networks in paleoecology is challenging due to the lack of observations of interactions, as well as biases in the preservation of species. The links of species in these networks must be inferred based on properties such as body size, similarities to living species, genetic information (when possible), and other known characteristics. Studying how paleo-networks have changed and adapted through time could assist in predicting how current ecological communities might react to environmental stressors. This work reconstructs networks from arthropod data found in rodent middens. The dynamics of these networks over 20,000 years is analyzed, and network metrics such as connectance are compared to modern networks.

Acknowledgements

Thanks to the invaluable support from my wife, my advisor, and my committee.

Contents

Introduction

Biological systems are inherently dynamic. Examples of dynamic networks are 1.) interactions of metabolites in metabolic networks, 2.) cycling and entrainment of biological clocks in circadian rhythm, and 3.)the emergence and disappearance of species in a paleoecological web. In this work, new methods for analyzing the dynamics of these systems are explored.

These systems can be modeled by constructing networks that describe the interactions of the variables. An advantage of network representation is that there is already a rich theory of network analysis that can be used. Here we use this developed theory, along with novel methods such as Linear-in-Flux-Expression (LIFE) and the two-step entrainment model.

Metabolic Systems

Models in Quantitative Systems Pharmacology (QSP) [3, 108, 121, 129] aim to gain information about drug discovery and optimization *in silico*. A potential advantage of this is obtaining valuable knowledge before more expensive stages, such as animal testing. Traditional methods represent a metabolic network as a directed graph, with edges representing reactions and vertices representing metabolites. Fluxes associated with the edges represent reaction rates. There are various challenges at modeling, including the complexity of the involved networks. Linear techniques such as Flux Balance Analysis, Markov Chains, Zero Deficiency Theory, and Laplacian Dynamics, are often used due the large scale systems [4, 20, 27, 45, 60, 65]. These techniques have had many successful uses, and have proved efficient due to their scalability in addressing such problems [53, 74, 85, 106, 117].. Despite this success, QSP commonly assume that all fluxes are independent [2, 64, 126]. Independence of fluxes is a reasonable assumption under special circumstances, but not when dealing with

general networks. Moreover, it does not recognize the robustness found in metabolic networks.

Nonlinear dynamics in metabolites are inherent in metabolic networks. For example, consider the action of enzymes, genes, or other metabolites to regulate a reaction; this would necessarily correspond to a nonlinear term for the metabolites involved. Recently Flux Balance Analysis methods were extended to include nonlinear metabolite dynamics. Linear-in-Flux-Expression (LIFE) [96] allows the leveraging of correlations among the fluxes of the model at a steady state. The stoichiometric matrix is used to create a system of Ordinary Differential Equations and study the dynamics of the system. In the LIFE approach, the stoichiometric matrix differs from standard definition as in [107, 137], in that it does not have restrictions on the dynamics of the metabolites. We compare the LIFE method with other systems biology approaches, highlighting which theory can be directly applied to this novel approach. The LIFE method allows to identify several general conditions ensuring the stability of the system. These conditions correspond to very natural assumptions for biological systems. For instance, every node must have a path to an excretion is a necessary condition for a LIFE system to have an equilibrium. To better describe biological networks, features not present in traditional methods were added. A new mathematical object called a metabolic graph is defined.

A metabolic graph differs from a directed graph in three main extensions: 1. The inclusion of intakes and excretions to an external environment. Many biological systems have inflows and outflows. To include these flows, we introduced virtual nodes. Virtual nodes represent the outside environment and function as a source and sink to the system. 2. Hyperedges. Some biological reactions include more than one substrate or product, and so it is necessary to include edges which have multiple starting or ending nodes [18, 138]. 3. Uberedges. The action of some metabolites or enzymes will promote or inhibit a specific reaction. This corresponds to a generalized

edge that begins at a node but ends at an edge. Mathematically we refer to these types of edges as uberedges [68].

Metabolic graphs are a new object without developed theory. Here, we begin to advance this theory, including proving a version of the Max-flow-min-cut theorem [47] for metabolic graphs and providing general conditions for stability. Due to their complexity, many systems may be unstable or support multiple equilibria. Using metabolic graphs, we apply LIFE to *Mycobacterium tuberculosis* (MTB) Central Carbon Metabolism. *Mycobacterium tuberculosis* is known to endure hostile environments and is very difficult to treat [10, 26, 51, 109]. Using LIFE, we simulate the effect of drug combinations on MTB metabolic networks to analyze the overall effect that different treatments may have. The network is first analyzed on its own, then the addition of drugs is done by adding uberedges to reactions known to have been affected by the drug.

Circadian Rhythm

Circadian rhythms are changes in an organism that follow a nearly 24-hour cycle. These rhythms are found in almost all organisms and are significant in predicting future stressors [120] . The biological clocks governing circadian rhythms usually entrain to an external signal called a zeitgeber. For circadian rhythms, the most common zeitgeber is the day/night cycle of the sun. The process of entrainment resets the clock, synchronizing the internal rhythm to the external environment.

A new model of circadian rhythm was developed in [86]. This model does not explain all biological resetting mechanisms for circadian rhythms but describes the effect of these mechanisms. As such it can be applied to various species, including humans and *Neurospora crassa*. Previous models were guided by a common assumption that a long period would only lead to phase delay and a short period would lead only to phase advance [1]. Although it is a widely accepted hypothesis, there exist examples

in nature that do not follow this assumption [36, 37, 58, 82] . We thus decided to develop a more general model capable to produce data *in silico* corresponding to the behavior of such natural organisms. This gave rise to a two-step entrainment model which can generate the wide range of results necessary to a wide range of data.

The potential of this model is analyzed to clearly explain the effects of each parameter, the reachable phase of entrainment values, and the range of entrainment. A piece-wise analytic solution model is found which provides faster simulations, thus greatly enhancing the ability to simulate many different conditions. We then use the two-step entrainment model to fit phase data in humans from the work [58]. Unlike other works, here we fit the entire phase trajectory, and not just the final phase. The capacity to fit phase data, including that which does not follow the common assumption, is an important step forward for circadian modeling.

Ecological Networks

One of the most fundamental elements of ecology is the interactions of species. Characterizing these interactions is an important step in understanding how the community functions. Ecological networks are dynamic and vary greatly over time and space, and should not be treated as static objects [115]. Several significant questions in ecology relate to how communities will change in response to climate and urbanization. To better understand how communities may change in the future it is useful to study how communities have changed in the past.

Analyzing interaction networks and their metrics helps to determine stability and key species of the network. The data used in this work comes from rodent middens which are between 400 and 40,000 years old. Because the species reactions cannot be directly observed, the main difficulty becomes reconstructing realistic networks based on this information. The methods of reconstruction are discussed and the reconstructed networks are analyzed for comparison to modern networks as well as

how they change through time.

Chapter 1

Equilibria for Large Metabolic Systems and the LIFE Approach

1.1 Introduction

The rate of drug development has increased in recent years. With the improved understanding of the clinically relevant differences among patient biology, there is a growing need to develop treatments in the context of a specific patient. Quantitative Systems Pharmacology (QSP) is an ideal tool for designing drugs and dosing regimens with a consideration of a patient's biology [48, 108]. In QSP, mathematical models of biological systems are implemented in-silico. The effects of a treatment can vary between individual patients. The reasons for this variability are not yet well understood, however with QSP models we may gain understanding by testing a drug on a Virtual Patient (briefly VP), an in-silico representation of a person's response to a drug.

QSP models can be tested on several VPs that sample the space of patient biological networks, and the result of a simulation using a VP better predicts the response to drug for a patient with similar biology to the VP. QSP models have been used for

applications in modeling cholesterol, HIV, and arthritis among others [64, 94, 126]. These predictions compare the expected effectiveness of the drug with the current established methods as well as predict the safety of new drug candidates. It is important to note that the initial levels of metabolites (or more general, chemical compounds) of patients receiving treatment can vary greatly, and even patients with similar baseline levels may respond differently. These variations show that, for QSP models to effectively predict patient response to a drug, there must be a wide range of VPs that represent the variety of patient responses [2, 48, 129].

QSP models generate VPs from several different parameterizations. The purpose of multiple parameterizations is to produce a wide range of responses that closely matches the range of clinical data. The parameters selected are typically parameters that have the greatest impact on the model and that vary across the patient population. These parameters will be chosen to fit the desired range, and then qualified by using a clinical dataset to test the model [2, 64, 50, 126]. The parameters used in QSP models are often assumed to have little to no correlation, or correlation is ignored completely [2, 64, 126]. Some patients may initially have similar baseline chemical compound levels yet respond very differently to treatment, and understanding how a perturbation of parameters effects the system will help predict patient response.

Traditional methods such as [102] focuses on linearity of the systems of Ordinary Differential Equations (ODEs), representing the metabolic network, with respect to the state of metabolites. Other methods represent the system of ODEs as a constant matrix multiplied by a reaction rate vector containing metabolite values [137]. LIFE methodology relies on linearity with respect to the fluxes, thus allowing for non-linearity in metabolites. Now, for given metabolite levels x one can define the kernel $K(x)$ as the null space of the stoichiometric matrix of metabolite reactions, seen as a linear map from the space of fluxes to the space of metabolites. This paper focuses on understanding the map $x \to K(x)$ from the space of metabolites to the

space of subspaces of fluxes, also called Grassmannian. More precisely, we address two main problems: the first is understanding the intersection of the kernels corresponding to different levels of metabolites, while the second is finding all metabolite levels such that a given flux belongs to the corresponding kernels. A number of results for these two problems are presented in terms of properties of the graph representing the metabolic network. Such results are illustrated with an example from the human cholesterol metabolism and a simple toy network. We also present a practical application for the LIFE approach; it can be used for QSP simulations, particularly how VPs can be fit to clinical data using an optimization process.

1.2 The LIFE Approach

1.2.1 LIFE model

We indicate by $x \in \mathbb{R}^n$ the metabolite variables and by $f \in \mathbb{R}^m$ the flux variables. A general system of ODEs which governs the quantities of x and f is written as

$$\begin{aligned} \frac{dx}{dt} &= F(x, f), \\ \frac{df}{dt} &= G(x, f), \end{aligned} \tag{1.1}$$

where $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ and $G : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$. In [60, 77], the authors show that the dynamics described by G evolve over a much smaller time-scale than F. This is referred to as "time-scale separation". Based on time-scale separation arguments of metabolic systems, we approximate the dynamics of the fluxes with $G \approx 0$, and our work focuses on the dynamics of the metabolites (F), with the fluxes playing the role of constant parameters.

Assuming $G \equiv 0$ and F linear in x, the usual method of writing the system of Ordinary Differential Equations(ODEs) (1.1) governing metabolism is given by (see

[60]):

$$\frac{dx}{dt} = J(f) \cdot x, \tag{1.2}$$

where $J(f)$ is an $n \times n$ matrix depending on the fluxes of the system.
Our method is also based on the assumption $G \equiv 0$, but asks for linearity of F with respect to the fluxes rather than to the metabolites. Such assumption is more often encountered when dealing with metabolic networks [96]. We call the Linear-in-Flux-Expression (LIFE) approach the idea of using linearity with respect to fluxes to write the dynamics as:

$$\frac{dx}{dt} = S(x) \cdot f, \tag{1.3}$$

where f is the column vector of fluxes and $S : \mathbb{R}^n \to M_{n\times m}$ is called the *stoichiometric matrix*. One constructs the stoichiometric matrix from the metabolites and the reactions that comprise a biochemical system. Each reaction corresponds to a flux f that connects two distinct metabolites or represents an intake or an excretion from the network. Each row of S corresponds to a metabolite and each column of S corresponds to a flux.

We use generalized idea of a *directed graph*, where we allow inflows to a graph from a general source, and outflows from the graph to a general sink. We say *graph* for brevity in this paper.

Definition 1. The *indegree* of a node is the number of directed edges for which the node is the terminal vertex. The *outdegree* of a node is the number of directed edges for which the node is the initial vertex.

Definition 2. A *source* of a graph is a directed edge with a node representing a compound only at the terminal end; the initial vertex has outdegree 1, indegree 0 and is not represented in our system. This is equivalent to an exchange reaction entering the system [107]. A *sink* of a graph is a directed edge with a node representing a compound only at the initial end. The terminal vertex has indegree 1 and outdegree

0, and is not represented in the system. This is equivalent to an exchange reaction leaving the system.

Definition 3. The stoichiometric matrix depends on a state variable x, and is denoted $S(x)$, (or S for brevity). $(S)_{ij} = s_{ij}$ can be defined from a graph. If the edge f_j has initial vertex x_i and terminal vertex x_k, then

$$\begin{cases} s_{ij} = -x_i \\ s_{kj} = x_i. \end{cases}$$

If the edge f_j is a source with terminal vertex x_k,

$$s_{kj} = 1.$$

If the edge f_j is a sink with initial vertex x_i,

$$s_{ij} = -x_i.$$

Definition 4. A *Weakly Connected Component* of a graph is a maximum subgraph such that an undirected path exists between every pair of nodes. A graph is weakly connected if there exists such a path between every pair of nodes.

Definition 5. The *grassmannian* $G(k, V)$ is the k-dimensional linear subspace of a space of dimension V.

The kernel of dimension d of a system is a subset of the $Grassmannian(d, m)$. We study the map $x \to K(x)$ as it relates to perturbations of stable systems.

Lemma 1. *Let x be the initial state for system (1.3), $f \in K(x)$. Assume that all eigenvalues of the jacobian matrix of the system at x have negative real part. Then*

there exists $\epsilon > 0$ *such that if* $y = x + \delta$, $|\delta| < \epsilon, y(\cdot)$ *is the solution starting at* y

$$\lim_{t \to +\infty} y(t) \in K^{-1}(f).$$

Proof. The assumption on the eigenvalues of the jacobian matrix imply the system is Lyapunov stable at x, see theorem 4.1.2 of [17], which implies $\lim_{t\to+\infty} S(y(t))f = 0$; we conclude $\lim_{t\to 0} y(t) \in K^{-1}(f)$. □

Lemma 1 motivates our investigation of $K^{-1}(f)$ and will determine candidate states to which a stable system will return after a perturbation.

Two main problems are investigated in this work.

- Problem 1: Characterize the intersections of the kernel for different states. For $x \neq y$ determine the intersection of the kernels $K(x) \cap K(y)$.
- Problem 2: Given $x, f \in K(x)$ compute $K^{-1}(f)$.

By exploring the map $x \to K(x)$ we will characterize $K(x) \cap K(\tilde{x})$ for some perturbation of the state, $\tilde{x} \neq x$. We show that for a fixed state x, $K(x) \cap K(\tilde{x})$ can have any dimension depending on $\tilde{x}$. That is, for some $\tilde{x}$ there is a $dim(K(x) \cap K(\tilde{x})) = 1$, and for some other $\tilde{x}$, $dim(K(x) \cap K(\tilde{x})) = 2$, etc.

1.2.2 LIFE approach for Virtual Patients

Traditional QSP approaches perturb the fluxes of a system and analyze the response. The LIFE method also perturbs the system, but assumes a steady state prior to the perturbation. We can utilize $K(x)$ to simulate VPs with the LIFE method, and optimize the fluxes to simulate metabolite trajectories that approximate clinical data. We find parameterizations of the system which minimize the distance between compound trajectories and measurements (Fig. 1.1). The first step of the procedure to fit our LIFE model is to generate a flux from $K(x)$. We sample coefficients to

use for a linear combination of kernel basis vectors. The coefficients of the basis for the kernel are hereafter called *parameters* and are denoted $a_i, i \in \{1, \ldots, k\}$ for $k = dim(K(x))$. Different parameters give a different sample from the kernel, and different samples produce different trajectories over time. We calculate the trajectory of metabolites $x(t)$ according to system (1.3), as well as solutions to a variational system: $v_i \in \mathbb{R}^n$ for $i \in \{1, \ldots, m\}$. More precisely, for $\hat{f}$, a flux sampled from $K(x)$, and $\frac{dx}{dt} = S(x) \cdot \hat{f}$

$$\frac{dv_i}{dt} = F_i(x, v_i, \hat{f})$$
$$F_i(y, z, \hat{f}) = (D(S(x) \cdot f) \cdot z)\Big|_{x=y, f=\hat{f}} + \frac{\partial (S(x) \cdot f)_{|x=y}}{\partial f_i}$$

We calculate trajectories of x and v by using a fourth order Runge-Kutta scheme. We use v to calculate the gradient of the cost function and use steepest decent method for minimizing this cost. For measurement of metabolites at time $t_j, (\bar{x}_j)$, we have

$$J = \sum_j \|x(t_j, a, x_0) - \bar{x}_j\|^2 \tag{1.4}$$

$$\frac{\partial J}{\partial f_i} = \sum_j \langle 2(\hat{x} - \bar{x}(t_j)), v_i(t_j) \rangle \tag{1.5}$$

where $v_i(t_j)$ is a value of v_i corresponding to time t_j. The derivative of the cost (1.5) (see proposition 1 and 2 in [22]) is used to selected new parameters a_i.

The steepest decent method can be implemented to update the coefficients.

$$a_i^{new} = a_i^{old} - g J_{f_j} \qquad \text{for } i \in \{1, \ldots, k\} \tag{1.6}$$

where $k = dim(K(x))$, with $g \in \mathbb{R}^+$ an optimization parameter.

The LIFE method requires that we sample fluxes from the kernel $K(x)$ of our stoichiometric matrix S, and that all metabolites are positive values. Therefore we

consider the intersection of the positive orthant with the $K(x)$. A convex combination of kernel basis vectors with positive entries will achieve this goal, however, it will only describe a subset of the kernel, in general. This problem was recognized by Palsson, [107]. In future work we will investigate this further.

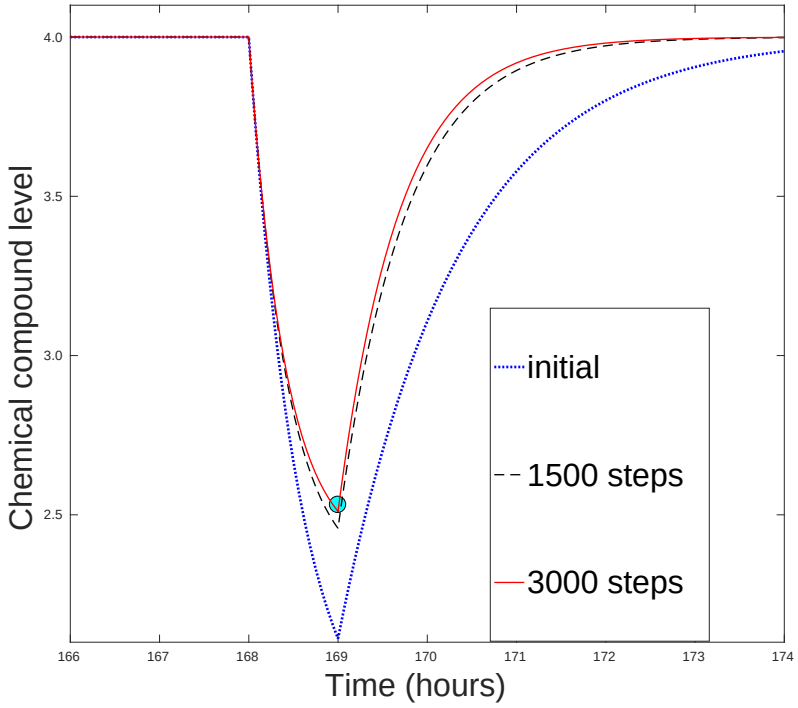


Figure 1.1: An example of optimization algorithm performed with the life method. Each step of the procedure minimizes cost in (1.4) according to measurements (cyan dot).

1.3 Key Example

A graph of a simple metabolic network is shown (Fig. 1.2). In this network are six metabolites $x = x_i, i \in \{1, ..., 6\}$. $\dot{x}_i$ indicates the derivative of metabolite, x_i. The fluxes are fixed and our system models the dynamics of the metabolites with a given flux.

In Fig. 1.2, the fluxes inside rectangles $\{f_1, f_2, f_3\}$ represent constant rates, whereas those in circles are first order rates. Specifically, the amount of x_1 molecules increases at a rate of f_1 per hour. Linearity in the flux space of the LIFE method facilitates the description for the kernel. We utilize fluxes from the kernel to analyze the system at steady state.

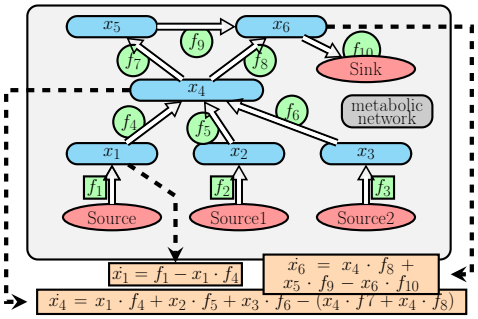


Figure 1.2: An example of a human cholesterol metabolic network with corresponding LIFE equations.

For this example, $S(x)$ is a 6×10 matrix, f is a vector composed of ten rate constants, from Fig. 1.2. A similar method for modeling biochemical networks is explained in [107], however S is not dependent on x in [107].

$$S(x) = \begin{pmatrix} 1 & 0 & 0 & -x_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -x_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -x_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_1 & x_2 & x_3 & -x_4 & -x_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & x_4 & 0 & -x_5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_4 & x_5 & -x_6 \end{pmatrix}. \tag{1.7}$$

Stochiometric matrix (1.7) governs the metabolites shown in Fig. 1.2. We may write this system of six ODEs from our sample in matrix form (1.7). One advantage of writing our system this way is we can calculate the kernel of the flux space for large

systems. The kernel of $S(x)$ is a set of flux vectors. We call an element of this set $\hat{f}$

$$\hat{f} = a_1 v_1 + a_2 v_2 + a_3 v_3 + a_4 v_4$$

$$= a_1 \begin{bmatrix} x_6 \\ 0 \\ 0 \\ \frac{x_6}{x_1} \\ 0 \\ 0 \\ 0 \\ \frac{x_6}{x_4} \\ 0 \\ 1 \end{bmatrix} + a_2 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \frac{x_5}{x_4} \\ \frac{-x_5}{x_4} \\ 1 \\ 0 \end{bmatrix} + a_3 \begin{bmatrix} -x_3 \\ 0 \\ x_3 \\ \frac{-x_3}{x_1} \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + a_4 \begin{bmatrix} -x_2 \\ x_2 \\ 0 \\ \frac{-x_2}{x_1} \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{1.8}$$

The kernel for equation (1.7) is given in equation (1.8). Note that there are four free variables, a_1, a_2, a_3, a_4, in this kernel for any fixed set of metabolite levels, x.

1.4 Results

Lemma 2. *Let $S \in M_{n\times m}$, $n < m$, be a stoichiometric matrix and G the associated directed graph. Assume G to be weakly connected with no sources or sinks. Denote by s_i the ith row of S. Then we have,*

$$\alpha_1 = \alpha_2 = \cdots = \alpha_n \iff \sum_{i=1}^{n} \alpha_i s_i = \vec{0}.$$

Proof. Because G has no sources and sinks S will have exactly two nonzero elements in each column. This is because each column represents a flow from one node to another.

$\Longleftarrow$) Fix a column j and let a, b be the rows with nonzero entries. Consider a linear combination of the rows of S such that

$$\sum_{i=1}^{n} \alpha_i s_i = \vec{0}. \tag{1.9}$$

Recall from definition (3) we have $s_{a,j} = -s_{b,j}$. Because a, b are the only nonzero entries in column j, the jth entry of $\sum_{i=1}^{n} \alpha_i s_i$ satisfies

$$\alpha_a s_{a,j} + \alpha_b s_{b,j} = 0 \implies \alpha_a = \alpha_b.$$

Because G is weakly connected there exists a path between any pair of nodes. Select two arbitrary nodes in the graph G and label them v,v'. Let W be the path between v and v' and label the nodes on the path W as $v = v_1, v_2, \ldots, v_{p-1}, v_p = v'$. Let j_i be the edge connecting v_i and v_{i+1}. Then for any i, the j_ith column satisfies

$$\begin{cases} s_{i,j_i} = -s_{i+1,j_i} \\ s_{k,j_i} = 0, \qquad \text{for } k \neq i, i+1. \end{cases}$$

Assume (1.9) then,

$$\alpha_i s_{i,j_i} + \alpha_{i+1} s_{i+1,j_i} = 0 \implies \alpha_i = \alpha_{i+1}.$$

Because j_i can represent any edge on path W, we have $\alpha_1 = \alpha_2, \alpha_2 = \alpha_3, \ldots, \alpha_{p-1} = \alpha_p \implies \alpha_1 = \alpha_2 = \cdots = \alpha_p$. Because v, v' were arbitrary nodes,

$$\sum_{i=1}^{n} \alpha_i s_i = \vec{0} \implies \alpha_1 = \alpha_2 = \cdots = \alpha_n. \tag{1.10}$$

$\Longrightarrow$) We assume that $\alpha_1 = \alpha_2 = \cdots = \alpha_n$. As before, fix a column j and let a, b be the rows with nonzero entries. From definition (3) we have $s_{a,j} = -s_{b,j}$. Now consider

the jth column of $\sum_{i=1}^{n} \alpha_i s_i$,

$$\sum_{i=1}^{n} \alpha_i s_{i,j} \tag{1.11}$$

which simplifies to

$$\alpha_a s_{a,j} + \alpha_b s_{b,j} = \alpha_a s_{a,j} - \alpha_b s_{a,j} = (\alpha_a - \alpha_b) s_{a,j} = 0$$

This is true for each column, which gives us $\sum_{i=1}^{n} \alpha_i s_i = \vec{0}$. □

Proposition 1. *Let $S \in M_{n\times m}$, $n < m$, be a stoichiometric matrix and G the associated directed graph. Assume G to be weakly connected with no sources or sinks. Then we have,*

$$Rank(S) = n - 1.$$

Proof. The ($\implies$) of lemma 2 implies $Rank(S) < n$.
Now we show that $Rank(S) \geq n - 1$. Consider the submatrix S^* constructed by removing the nth row from S. Then for s_i^* the ith row of S^*,

$$\sum_{i=1}^{n-1} \alpha_i s_i^* = \left(\sum_{i=1}^{n} \alpha_i s_i \right)_{\Big|_{\alpha_n=0}}. \tag{1.12}$$

by (1.12) and lemma 2 It follows that

$$\sum_{i=1}^{n} \alpha_i s_i^* = \vec{0} \implies \alpha_i = 0 \text{ for all } i \in \{1, \ldots, n-1\}.$$

Therefore, $Rank(S^*) = n-1 \implies n-1 \leq Rank(S) < n$ and so $Rank(S) = n-1$. □

A system with the properties of proposition 1 also satisfies the zero deficiency theorem of [45], which implies it has one equilibrium solution.

Proposition 2. *Let $S \in M_{n\times m}$, $n < m$, be a stoichiometric matrix and G the associated directed graph. Assume G to be weakly connected with at least one source and*

no sinks. Then we have,

$$Rank(S) = n.$$

Proof. First we show that for a graph G with a single source and no sinks, that for S, the stoichiometric matrix for G, $Rank(S) = n$. Let the source be called f_{m+1} and the terminal vertex of f_{m+1} be called x_1. Let G^* be the subgraph of G without the source, and S^* be the matrix for G^*. S^* is a submatrix of S excluding the column containing the source. We have $rank(S^*) = n - 1$. We can use elementary row operations to change a row in S^* without changing the rank of S^*. We replace the first row of S^* with $\sum_{i=1}^{n} s_i^* = \vec{0}$ and call this new matrix S^1, $rank(S^1) = rank(S^*) = n - 1$. Similarly, if we append a column of zeros to the right side of S_1^* the rank will not change. We call the matrix with the added column S^2, $rank(S^2) = rank(S^1) = rank(S^*) = n - 1$. S^2 is S with the first row of S set to $\vec{0}$. Now we replace the first row of S^2 with $(s_{1,1}, s_{1,2}, \ldots, s_{1,m-1}, s_{1,m} = 1)$ which gives us S. Because the first row is independent to all others:

$$rank(S) = rank(S^2) + 1 = n.$$

□

For ease of proof the graph contained no sinks. However, adding sinks to the graph will not change the rank of the S. This is because S is already full rank and adding a sink will append a new column to S. A graph with sources and no sinks is not realistic as it will have continuous accumulation of metabolites.

Proposition 3. *Let S be the stoichiometric matrix and G the associated directed graph. Assume G to be weakly connected with no sources or sinks. Consider the kernel of S, $K(x)$, and assume that $\tilde{x} = cx$ for some $c \in \mathbb{R}$. Then we have, $K(x) = K(\tilde{x})$.*

Proof. Let $\hat{f} \in K(x)$, and $\breve{f} \in K(\tilde{x})$ then

$$\begin{aligned} S(x)\hat{f} = 0 &\implies cS(x)\hat{f} = S(\tilde{x})\hat{f} = 0 \\ S(\tilde{x})\breve{f} = 0 &\implies cS(x)\breve{f} = 0. \end{aligned} \tag{1.13}$$

□

Proposition 4. *Let S be a stoichiometric matrix for a graph containing a directed path along three nodes, and the middle node has only one incoming and one outgoing edge. For a state x and different state $\tilde{x}$, if $\tilde{x} \neq cx$ for $c \in \mathbb{R}$ then*

$$K(x) \cap K(\tilde{x}) = \{\vec{0}\}. \tag{1.14}$$

Proof. G has directed path along three nodes, initial node x_1, middle node x_2, terminal node x_3; call edge connecting x_1 to x_2 as f_1 the other edge is f_2. Then the second row of S is $s_2 = (-x_1, x_2, 0, \ldots, 0)$ and

$$\begin{aligned} &S(x)f = 0 \implies f_1x_1 = f_2x_2 \implies f_1 = f_2\frac{x_2}{x_1} \\ &S(\tilde{x})f = 0 \implies f_1\tilde{x}_1 = f_2\tilde{x}_2 \implies f_1 = f_2\frac{\tilde{x}_2}{\tilde{x}_1} \\ &f \in K(x) \cap K(\tilde{x}), f \neq \{\vec{0}\} \implies f_2\frac{x_2}{x_1} = f_2\frac{\tilde{x}_2}{\tilde{x}_1} \implies \tilde{x} = cx. \end{aligned}$$

□

Proposition 5. *Let $S \in M_{n\times m}$, $n < m$, be a stoichiometric matrix and G the associated directed graph. Assume G to be weakly connected with one source and no sinks. Let $S^* \in M_{n\times m-1}$, be a submatrix of S where the source is removed. (WLOG let the source in G be represented by the last column of S). Consider the kernels of S and S^*, $K(x)$ and $K^*(x)$ respectively and let B^* be a basis of $K^*(x)$. Let B be the collection of vectors such that each $b \in B$ is equal to a $b^* \in B^*$ with a 0 appended as the last entry for each vector. Then B is a basis for $K(x)$.*

Proof. We prove:

1. for $b \in B, Sb = \vec{0}$ and so $b \in K(x)$.

2. B is an independent set with number of elements equal to dimension of $K(x)$.

Let e be an $n \times 1$ column vector containing a single 1 and the other entries 0.

$$S_{n\times(m+1)}b = \left(S^*|e\right)\begin{pmatrix} b^* \\ \cdots\cdots \\ 0 \end{pmatrix}.$$

Let b_j be the jth entry of vector b. For A_i, the ith entry of the solution to Sb.

$$A_i = \sum_{j=1}^{m} S_{ij}b_j = \sum_{j=1}^{m-1} S^*_{ij}b^*_j + S_{i\,m} \cdot 0 = 0.$$

Appending a 0 to each vector of a linearly independent set gives an linearly independent set. From propositions 1 and 2 we know that

$$Rank(S^*(x)) = n - 1, Rank(S(x)) = n.$$

The dimension of each kernel is the same ($dim(K^*(x)) = (m-1) - (n-1)$ and $dim(K(x)) = m - n$. The cardinality of B^* = cardinality of $B = dim(K(x)) = dim(K^*(x))$ because B is a basis and both kernels have the same dimension (though the dimension of their ambient space differs), we conclude that B is a basis for $K(x)$.

□

1.5 Example

Here we show a complete solution to problem 1. In this section we explore the kernel of an example network. The initial state of the kernel will be characterized, and

the intersection of this kernel with the kernels of perturbed metabolic states will be analyzed.

$S(x)$ is the stoichiometric matrix associated to the graph in Fig. 1.3

$$S(x) = \begin{pmatrix} -x_1 & 0 & 0 & x_4 & -x_1 & 0 \\ x_1 & -x_2 & 0 & 0 & 0 & -x_2 \\ 0 & x_2 & -x_3 & 0 & x_1 & 0 \\ 0 & 0 & x_3 & -x_4 & 0 & x_2 \end{pmatrix}.$$

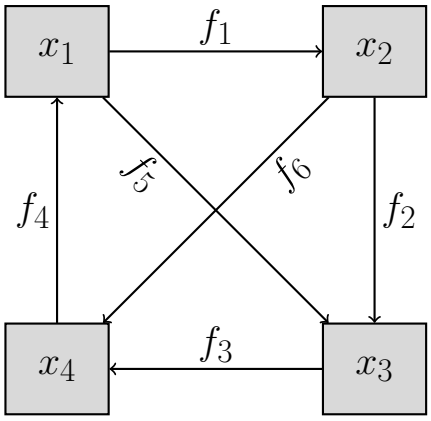


Figure 1.3: A directed graph representing a biochemical system.

From proposition 1 we have $rank(S) = 3$, which implies the dimension of the kernel is 3. The basis for the kernel is

$$(v_1|v_2|v_3)^T = \begin{pmatrix} 0 & -1 & -\frac{x_2}{x_3} & 0 & 0 & 1 \\ -1 & -\frac{x_1}{x_2} & 0 & 0 & 1 & 0 \\ \frac{x_4}{x_1} & \frac{x_4}{x_2} & \frac{x_4}{x_2} & 1 & 0 & 0 \end{pmatrix}.$$

A perturbation of $x \to \tilde{x}$ will alter the basis vectors and thus change the kernel $K(x) \to K(\tilde{x})$. $K(\tilde{x})$ may have some non trivial intersection with $K(x)$. Any flux in the perturbed kernel can be represented by the perturbed basis vectors. For all

$f \in K(\tilde{x}) : f = \tilde{\lambda}_1\tilde{v_1} + \tilde{\lambda}_2\tilde{v_2} + \tilde{\lambda}_3\tilde{v_3}$ where each $\tilde{v_i}$ represents a perturbed basis vector and each $\tilde{\lambda_i} \in \mathbb{R}$.

A flux $f \in K(x) \cap K(\tilde{x})$ can be found as a solution to the following equation:

$$\lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3 = \tilde{\lambda}_1\tilde{v_1} + \tilde{\lambda}_2\tilde{v_2} + \tilde{\lambda}_3\tilde{v_3}.$$

Comparing the equation by components, we have the conditions that must be satisfied for any flux in the intersection.

$$\lambda_1 = \tilde{\lambda}_1, \quad \lambda_2 = \tilde{\lambda}_2, \quad \lambda_3 = \tilde{\lambda}_3 \tag{1.15}$$

$$\begin{cases} \lambda_1\frac{x_4}{x_1} - \lambda_2 & = \tilde{\lambda}_1\frac{\tilde{x}_4}{\tilde{x}_1} - \tilde{\lambda_2} \\ \lambda_1\frac{x_4}{x_2} - \lambda_2\frac{x_1}{x_2} - \lambda_3 & = \tilde{\lambda}_1\frac{\tilde{x_4}}{\tilde{x}_2} - \tilde{\lambda}_2\frac{\tilde{x_1}}{\tilde{x}_2} - \tilde{\lambda_3} \\ \lambda_1\frac{x_4}{x_3} - \lambda_3\frac{x_2}{x_3} & = \tilde{\lambda}_1\frac{\tilde{x}_4}{\tilde{x}_3} - \tilde{\lambda}_3\frac{\tilde{x}_2}{\tilde{x}_3}. \end{cases} \tag{1.16}$$

With (1.15), we simplify system (1.16) to

$$\frac{x_4}{x_1} = \frac{\tilde{x}_4}{\tilde{x}_1} \tag{1.17}$$

$$\lambda_1\left(\frac{x_4}{x_2} - \frac{\tilde{x_4}}{\tilde{x_2}}\right) = \lambda_2\left(\frac{x_1}{x_2} - \frac{\tilde{x_1}}{\tilde{x_2}}\right) \tag{1.18}$$

$$\lambda_1\left(\frac{x_4}{x_3} - \frac{\tilde{x_4}}{\tilde{x_3}}\right) = \lambda_3\left(\frac{x_2}{x_3} - \frac{\tilde{x}_2}{\tilde{x}_3}\right). \tag{1.19}$$

Depending on which of the conditions are met the dimension of the intersection ($dim(K(x) \cap K(\tilde{x}))$) can be determined. Different perturbations of x will be considered that satisfy only some of these conditions. The following cases ((I) through (V)) show results specific to which condition are satisfied.

(I) Let $\tilde{x}$ be a perturbation such that (1.17) is not satisfied, the intersection will be trivial and $dim(K(x) \cap K(\tilde{x}) = 0$.

(II) Let $\tilde{x}$ be a perturbation which satisfies (1.17) and

$$\left(\frac{x_4}{x_2}-\frac{\tilde{x}_4}{\tilde{x}_2}\right)\left(\frac{x_1}{x_2}-\frac{\tilde{x}_1}{\tilde{x}_2}\right)\neq 0, \left(\frac{x_4}{x_3}-\frac{\tilde{x}_4}{\tilde{x}_3}\right)\left(\frac{x_2}{x_3}-\frac{\tilde{x}_2}{\tilde{x}_3}\right)\neq 0.$$

This allows equations (1.18) and (1.19) to be arranged in the following manner.

$$\lambda_2=\lambda_1\frac{\frac{x_4}{x_2}-\frac{\tilde{x}_4}{\tilde{x}_2}}{\frac{x_1}{x_2}-\frac{\tilde{x}_1}{\tilde{x}_2}}, \qquad \lambda_3=\lambda_1\frac{\frac{x_4}{x_3}-\frac{\tilde{x}_4}{\tilde{x}_3}}{\frac{x_2}{x_3}-\frac{\tilde{x}_2}{\tilde{x}_3}}.$$

This shows a relationship where both λ_2 and λ_3 depend on λ_1 and the metabolites x_i. λ_1 is the only free variable and so $dim(K(x)\cap K(\tilde{x}))=1$.

(III) Let $\tilde{x}$ satisfy (1.17). And also let $\left(\frac{x_4}{x_3}-\frac{\tilde{x}_4}{\tilde{x}_3}\right)\left(\frac{x_2}{x_3}-\frac{\tilde{x}_2}{\tilde{x}_3}\right)\neq 0$, $\frac{x_4}{x_2}-\frac{\tilde{x}_4}{\tilde{x}_2}=0$, $\frac{x_1}{x_2}-\frac{\tilde{x}_1}{\tilde{x}_2}=0$. Then (1.18) is satisfied regardless of the value of λ_2. λ_2 is a free variable in addition to λ_1 while λ_3 is still dependent on the state $\tilde{x}$ and λ_1. With two free variables $dim(K(x)\cap K(\tilde{x}))=2$.

(IV) Let $\tilde{x}$ satisfy (1.17) and also let

$$\frac{x_4}{x_3}-\frac{\tilde{x}_4}{\tilde{x}_3}=0, \frac{x_2}{x_3}-\frac{\tilde{x}_2}{\tilde{x}_3}=0 \tag{1.20}$$

$$\left(\frac{x_4}{x_2}-\frac{\tilde{x}_4}{\tilde{x}_2}\right)\left(\frac{x_1}{x_2}-\frac{\tilde{x}_1}{\tilde{x}_2}\right)\neq 0. \tag{1.21}$$

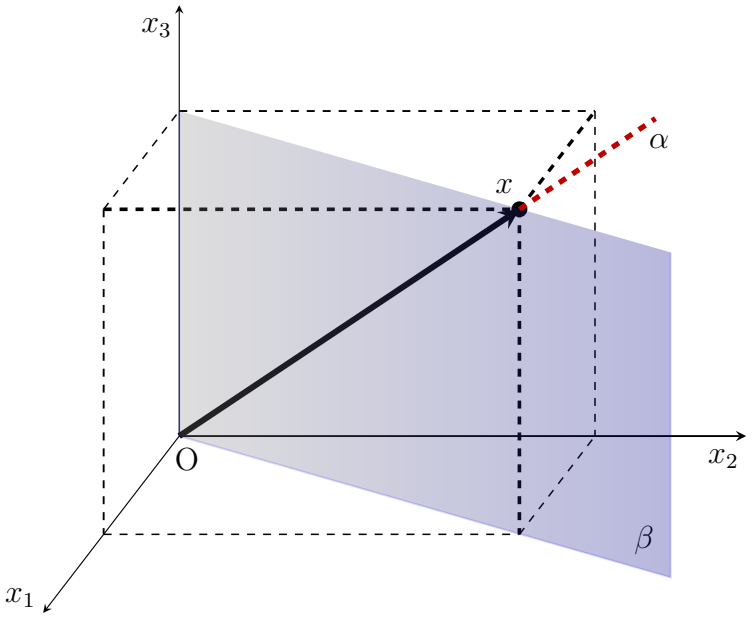


Figure 1.4: A 3-D representation of the metabolic state space which highlights states $\tilde{x}$ with nontrivial intersections with the kernel of the initial state (represented by x). A three dimensional representation is appropriate because (1.17) implies that x_1 and x_4 are proportional. The line α represents case (IV) where $\tilde{x}_1, \tilde{x}_2, \tilde{x}_3$ are proportional to x_1, x_2, x_3 respectively. States $\tilde{x}$ on this line will have kernels such that $dimension(K(x) \cap K(\tilde{x})) = 3$. The plane β represents case (III) where only $\tilde{x}_1, \tilde{x}_2$ are proportional to x_1, x_2.

Upon further inspection, however, we find that (1.17) and (1.20) implies $\left(\frac{x_4}{x_2} - \frac{\tilde{x}_4}{\tilde{x}_2}\right) = \left(\frac{x_1}{x_2} - \frac{\tilde{x}_1}{\tilde{x}_2}\right) = 0$ which contradicts (1.21). Thus the perturbation given by case (IV) doesn't exist.

(V) Let $\tilde{x}$ satisfy (1.17). And let

$$\frac{x_4}{x_2} - \frac{\tilde{x}_4}{\tilde{x}_2} = \frac{x_1}{x_2} - \frac{\tilde{x}_1}{\tilde{x}_2} = \frac{x_4}{x_3} - \frac{\tilde{x}_4}{\tilde{x}_3} = \frac{x_2}{x_3} - \frac{\tilde{x}_2}{\tilde{x}_3} = 0.$$

Equations (1.17), (1.18) and (1.19) are satisfied for any value of λ_1, λ_2 and λ_3. With three free variables $dim(K(x) \cap K(\tilde{x})) = 3$. Fig. 1.4 shows states $\tilde{x}$ for which $K(x) \cap K(\tilde{x})$ is non trivial. The reference state x is shown, and the entire space represents other states $\tilde{x}$ such that equation (1.17) is satisfied.

1.6 Conclusions

This work exploits the linearity of some metabolic systems with respect to fluxes. Namely, we are able to see systematic relationships among fluxes at equilibrium as opposed to treating all fluxes as independent. Propositions concerning the rank of our stoichiometric matrix are presented, from which the dimension of the kernel may be easily deducted.

Chapter 2

Stability of Metabolic Networks via LIFE

2.1 Introduction

Quantitative Systems Pharmacology (QSP) aims to gain more information about a potential drug treatment on a human patient before the more expensive stages of development begin [108]. QSP models allow us to perform in silico experiments on a simulated metabolic system that predicts the response of perturbing a flux. A drug may be metabolized differently by various patients, and modelers working in pharmacology must anticipate these differences. Building a profile of how the drug affects different classes of simulated patients will help the developers of new drugs understand the viability of a treatment and acquire insight into the mechanisms by which the drug acts.

A recent advancement in QSP modeling called Linear-in-Flux-Expression (LIFE) is a method of analyzing systems of Ordinary Differential Equations (ODEs) [96, 97]. Originally, LIFE was designed to analyze metabolic systems, which are composed of Fluxes and Metabolites. Fluxes in the metabolic system are the rates of chemical

reactions in the human body, and they determine the dynamics on the metabolites, which are the various chemical compounds involved in metabolism. Modeling these systems depends on choosing fluxes, which are difficult to measure directly, so that the system effectively simulates human metabolism.

To implement LIFE on a metabolic network, the network must be written as a directed graph [97]. The edges of the graph represent the reaction rates (*fluxes*), and the vertices represent quantities of chemical compounds (*metabolites*). From the graph we construct the stoichiometric matrix of the system. This stoichiometric matrix is not the classical one mentioned by [107, 137]. The LIFE method is also different from QSP models whose dynamics traditionally depend on a matrix containing information about the flux of the system. In these classical QSP models the dynamics of the metabolites are linear with respect to the metabolites. By contrast, systems using the LIFE method are linear in fluxes and have a stoichiometric matrix that is dependent on the metabolites.

Initially, the LIFE method was developed using the human cholesterol metabolism network [96]. LIFE enables us to simply describe the correlations among the fluxes of the model at steady state. There are generally many correlations among fluxes, and maintaining these correlations leads to a more consistent response to perturbing the fluxes in the system. This was advantageous to QSP modelers, who previously analyzed flux perturbations with little to no consideration to relationships among fluxes [2]. Now, we expand our study of these systems, showing that with few assumptions, systems that are linear in the flux is stable.

The LIFE method evolved from methods in systems biology [107]. Systems biology, in conjunction with network flows [61, 47], Markov chains [27], laplacian dynamics [102], control theory [17], and compartmental systems [20, 65] allow us to better understand biological networks on which pharmacology models are based. The field of compartmental systems focuses on models based on directed graphs. Vertices of

the graph represent quantities whose dynamics are determined by the edges of the graph, which represent fluxes among compartments. Markov chains study dynamics on directed graphs as well, but by contrast, this field focuses on stochastic processes. Control theory studies the way an external agent can alter the natural evolution of a system, given a set of admissible controls. In pharmacology, metabolism follows its natural evolution, and drugs serve as our controls. These fields have much to contribute to systems pharmacology, and we summarize useful results. We identify assumptions which are usually satisfied by real metabolic networks, guaranteeing stability of the metabolic system at a unique equilibrium.

The paper is organized as follows. In section 2.2, we describe the model system for the LIFE approach in the form of a system of ODEs associated to a metabolic networks, then show existence of positive solutions and provide results of equilibria under general assumptions. Also special classes of LIFE systems are introduced. Section 2.3 investigates the flow vectors for which a given metabolite vector x is an equilibrium of the network, including the extreme pathways approach. On the other side, Section 2.4 studies the opposite problem: find the metabolite vectors which are equilibria of the network for a fixed flow vector. This is done first investigating the relationships between linear LIFE systems and Markov chains, Laplacian dynamics and linear compartmental systems. Then we deal with special classes of nonlinear LIFE systems. Finally, a comparison between zero-deficiency theory is discussed. The paper ends with conclusions in Section 2.5 and an Appendix containing examples.

2.2 System model

2.2.1 Notation and preliminaries

We indicate by $\mathbb{R}_+ = [0, +\infty)$ the set of positive real numbers, by $\mathbb{R}^n$ the Euclidean real space of dimension n and by $M_{n\times m}$ the set of $n \times m$ matrices with real en-

tries. Given a matrix S, we indicate by S^T its transpose. Given $d_1, \ldots, d_n \in \mathbb{R}$, $\mathrm{diag}(d_1, \ldots, d_n)$ is the diagonal matrix with entries d_i on the diagonal. We denote by $\mathbf{1}$ a column vector with all entries equal to 1, of size clear from the context.

We introduce some terminology commonly used in graph theory. A directed graph is a couple $G = (V, E)$, with $V = \{v_1, \ldots, v_n\}$ the set of vertices and $E \subset V \times V$ the set of edges. For a graph with n vertices and m edges, ordering the edges lexicographically, the *incidence matrix* is a matrix, $\Gamma \in M_{n\times m}$ such that $\Gamma_{ij} = 1$ if the jth edge is (v_k, v_i) for some vertex v_k, $\Gamma_{ij} = -1$ if the jth edge is (v_i, v_k) for some vertex v_k, and $\Gamma_{ij} = 0$ otherwise. A path is a sequence of distinct vertices $v_{i_1} \cdots v_{i_k}$, with $(v_{i_j}, v_{i_{j+1}}) \in E$ for $j = 1, \ldots, k-1$. A graph is *strongly connected* if there exists a path between every pair of vertices. A *strongly connected component* of a directed graph is a maximal strongly connected subgraph.

A *terminal component* of a directed graph $G = (V, E)$ is a strongly connected component $G' = (V', E')$, with $V' \subset V$, $E' \subset E$, such that there exists no edge $e = (v', v)$, with $v' \in V'$ and $v \in V \setminus V'$. An undirected path is a sequence of distinct vertices $v_{i_1}, \cdots, v_{i_k}$, with either $(v_{i_j}, v_{i_{j+1}}) \in E$ or $(v_{i_{j+1}}, v_{i_j}) \in E$ for $j = 1, \ldots, k-1$. A directed graph is *weakly connected* if there exists an undirected path between every pair of vertices. A *weakly connected component* of a directed graph is a maximal weakly connected subgraph. A directed graph $G = (V, E)$ is *weakly reversible* if every weakly connected component is also strongly connected.

2.2.2 LIFE systems

Recall the LIFE model from section 1.2.1, we now illustrate how to construct a directed graph from the metabolic network for the system (1.3). We represent metabolites with vertices $V = \{v_1, \ldots, v_n\}$. We construct a set of edges $E \subset V \times V$ to represent reactions; each edge is associated to a flux from one metabolite to another, notice that we do not have loops. To represent intakes and excretions, we introduce

two virtual vertices, v_0 and v_{n+1}, not associated with any metabolite but rather representing the external environment. We denote by I, X the set of vertices attached to v_0, v_{n+1}, the vertices in I and X are called *intake vertices* and *excretion vertices*, respectively. We also introduce edges (v_0, w) with $w \in I \subset V$ representing intakes, and (w, v_{n+1}), $w \in X \subset V$ representing excretions. We use the extended graph $\tilde{G} = (\tilde{V}, \tilde{E})$ defined by $\tilde{V} = V \cup \{v_0, v_{n+1}\} = \{v_0, v_1, \ldots, v_n, v_{n+1}\}$ and $\tilde{E}$ collecting edges in E together with intake and excretion edges. The rows of the matrix S can be indexed by vertices in V and the columns by edges in $\tilde{E}$, thus we write S_{ve} for the entry corresponding to vertex v and edge e. Moreover we denote by x_v the metabolite corresponding to vertex v and by f_e the flux corresponding to edge e. All columns of S have zero sum, except those corresponding to intakes and excretions, which have positive and negative sum, respectively. In simple words, the entry $S_{ve}(x)$ is a function quantifying how much mass moves along edge e. We notice that, in systems biology, researchers often refer to $J(f)$ of (1.2) as the stoichiometric matrix, see [107].

Example 1. To illustrate the concepts of graph with virtual vertices and stoichiometric matrix related to a metabolic network, we provide a toy example with linear dynamics. Consider the system given by the following stoichiometric matrix and fluxes vector:

$$S(x) = \begin{pmatrix} 1 & -x_1 & -x_1 & 0 & 0 & 0 & x_4 \\ 0 & x_1 & 0 & -x_2 & 0 & 0 & 0 \\ 0 & 0 & x_1 & x_2 & -x_3 & -x_3 & 0 \\ 0 & 0 & 0 & 0 & x_3 & 0 & -x_4 \end{pmatrix}, f = \begin{pmatrix} f_{(v_0,v_1)} \\ f_{(v_1,v_2)} \\ f_{(v_1,v_3)} \\ f_{(v_2,v_3)} \\ f_{(v_3,v_4)} \\ f_{(v_3,v_5)} \\ f_{(v_4,v_1)} \end{pmatrix}.$$

Then the corresponding graph is represented in Figure 2.1.

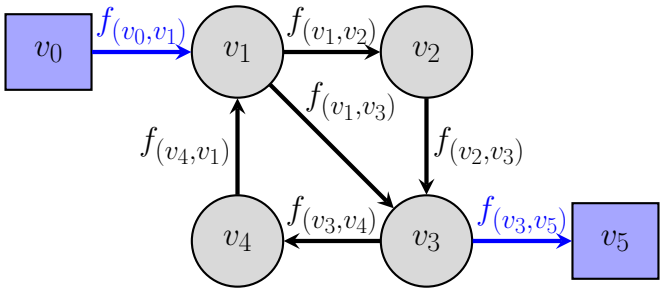


Figure 2.1: A directed graph $\tilde{G} = (\tilde{V}, \tilde{E})$ representing a biochemical system. The rectangles indicate virtual vertices and the subgraph of circular vertices and edges connecting them is $G = (V, E)$.

Remark 1. It is worth a reminder that our stoichiometric matrix is different from the traditional one defined by [107, 137], in which entries are stoichiometric coefficients, i.e. do not depend on metabolites.

To correctly represent the reactions corresponding to fluxes (which take always strictly positive values), we assume:

(A) For $x \in (\mathbb{R}_+)^n$, it holds

$$S_{ve}(x) = \begin{cases} H_e(x) > 0 & \text{if } e = (w, v),\ w \in V \text{ and } x_v > 0 \text{ or } e = (v_0, v),\ v \in I \\ -H_e(x) < 0 & \text{if } e = (v, w),\ w \in V \text{ and } x_v > 0 \text{ or } e = (v, v_{n+1}),\ v \in X \text{ and } x_v > 0 \\ 0 & \text{otherwise,} \end{cases}$$

where $H_e : \mathbb{R}^n \to \mathbb{R}$ is a positive continuous function.

Notice that Assumption (A) implies that, for each $v \in V$,

$$\sum_{v \in V} S_{ve}(x) = \begin{cases} H_e(x) & e = (v_0, \bar{v}),\ \bar{v} \in I, \\ -H_e(x) & e = (\bar{v}, v_{n+1}),\ \bar{v} \in X, \\ 0 & \text{otherwise,} \end{cases} \tag{2.1}$$

namely all columns of S have zero sum, except those corresponding to intakes and excretions, which have positive and negative sum, respectively. Under Assumption (A), the dynamics (1.3) can be interpreted as mass conservation law. Indeed, re-writing (1.3) entrywise and using (A), we have

$$\dot{x}_v = \sum_{e\in\tilde{E}} S_{ve}(x) f_e = \sum_{w:(w,v)\in\tilde{E}} H_{(w,v)}(x) f_{(w,v)} - \sum_{w:(v,w)\in\tilde{E}} H_{(v,w)}(x) f_{(v,w)}\,, \tag{2.2}$$

which is the mass balance for metabolite x_v: its variation is given by the sum of the incoming flows, minus the sum of the outgoing flows. This is the analogous Kirchhoff's current law for electrical networks, with the difference that currents are allowed to take negative values as well, while here metabolite variables are non-negative.

The total mass in the system is $m = \sum_{v\in V} x_v$. From (2.2) we have

$$\dot{m} = \sum_{v\in I} H_{(v_0,v)}(x) f_{(v_0,v)} - \sum_{v\in X} H_{(v,v_{n+1})}(x) f_{(v,v_{n+1})}\,.$$

Clearly, in the case without intakes nor excretions, $\dot{m} = 0$, i.e. the total mass of a closed system is constant in time.

Another remark which is useful is that, under Assumption (A), $S(x)$ can be re-written as $S(x) = \Gamma D(x)$, where $D(x)$ is a diagonal matrix of size $m \times m$, with diagonal entries given by $H_e(x)$'s, and Γ is obtained from the incidence matrix of $\tilde{G}$ by removing the first and last rows (corrseponding to v_0 and v_{n+1}). In the particular case without intakes nor excretions, Γ is the incidence matrix of G.

In the remainder of this section we study the dynamics (1.3) under the very general Assumption (A), while later in the paper we add other assumptions, restricting our attention to systems for which stronger statements can be obtained. A first important general property is that positivity of solution is guaranteed:

Proposition 6. *Consider a system* (1.3) *satisfying (A) and the Cauchy problem with*

initial datum $x_v(0) = x_0^v$. *Assume that* S *is locally Lipschitz. If* $f_e > 0$ *for every* $e \in E$ *and* $x_0^v \geq 0$ *for every* $v \in V$*, then there exists a local solution* $x_v(\cdot)$ *defined on* $[0, T]$*,* $T > 0$*, and* $x_v(t) \geq 0$ *for every* $t \in [0, T]$.

Proof. Existence follows from Lipschitz condition, while positivity of solution follows from the invariance of the set $\{x : x_v \geq 0\}$. □

In the next Proposition we show that existence of nontrivial equilibria implies some structure on the network: every vertex v for which there is a directed path from some $w \in I$ to v, must also have a directed path from v to some $y \in X$. This result refines the space of networks with which we are concerned. More precisely:

Proposition 7. *Consider a system* (1.3) *satisfying (A). Assume there exists an equilibrium* $\bar{x} \in (\mathbb{R}_+)^n$ *for a flux vector* f *such that* $f_e > 0$ *for every* $e \in \tilde{E}$*. Then for every vertex* $v \in V$ *for which there exists a path from* I *to* v*, there exists a path from* v *to* X.

Proof. Assume there exists an equilibrium $\bar{x}$ as in the statement and, by contradiction, a vertex v for which there exists a path from $w \in I$ to v, but there exists no $y \in X$ to which v is connected. Since there is no path from v to X, either v belongs to a terminal component with no excretion, or there is a path from v to a terminal component with no excretion. Denote by $G_T = (V_T, E_T)$ such a terminal component. Since there are a path from $w \in I$ to v and a (possibly trivial) path from v to V_T, then there is also a path from v_0 to V_T. Denote by $v_0, v_1 = w, \ldots, v_{\ell-1}, v_\ell \in V_T$ one such a path, such that $v_{\ell-1} \notin V_T$ (possibly the path is a single edge, in case $w \in V_T$).

It is easy to show that $x_{v_i} > 0$ for all $i = 1, \ldots, \ell$, as follows. Considering $e = (v_0, v_1)$, by (A) we have $S_{v_0,e}(\bar{x}) = H_e(\bar{x}) > 0$. Similarly, for every $e' = (v_1, w')$, $x_{v_1} = 0$ implies $S_{v_1,e'}(x) = 0$. This means we must have $\bar{x}_{v_1} > 0$, otherwise we would have $\dot{x}_{v_1} \geq f_e H_e(\bar{x}) > 0$ (where $e = (v_0, v_1)$), contradicting $\bar{x}$ being an equilibrium. Having proved that $\bar{x}_{v_1} > 0$, we can proceed by induction: for $i = 1, \ldots, \ell - 1$, $\bar{x}_{v_i} > 0$

implies $\bar{x}_{v_{i+1}} > 0$. The argument is the same as above, with a slight modification: looking at $e = (v_i, v_{i+1})$, $S_{v_i,e}(\bar{x}) = H_e(\bar{x}) > 0$ thanks to Assumption (A) together with $\bar{x}_i > 0$, while above we were in the case of an intake.

Finally we have a terminal component $G_T = (V_T, E_T)$ with no excretion, and an edge $(v_{\ell-1}, v_\ell)$ with $v_\ell \in V_T$ and $v_{\ell-1} \notin V_T$, such that either $v_{\ell-1} = v_0$ or $\bar{x}_{v_{\ell-1}} > 0$. In either case, considering $e = (v_{\ell-1}, v_\ell)$, by (A) we have $S_{v_{\ell-1},e}(\bar{x}) = H_e(\bar{x}) > 0$. Now consider the variation of mass in the component G_T: since there is no outgoing edge from G_T, and there is at least the incoming edge e, we have $\frac{\mathrm{d}}{\mathrm{d}t}\sum_{v\in V_T} x_v = \sum_{v\in V_T} \dot{x}_v \geq H_e f_e > 0$, contradicting the fact that $\bar{x}$ is an equilibrium. □

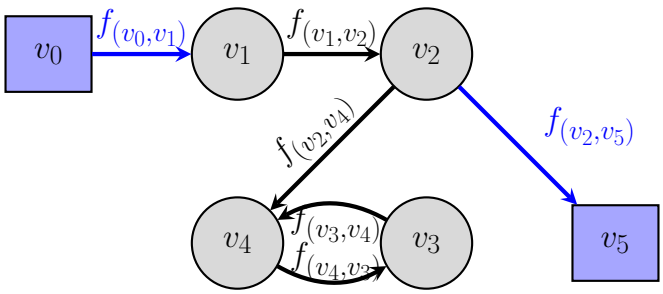


Figure 2.2: A directed graph $\tilde{G} = (\tilde{V}, \tilde{E})$ illustrating Proposition 7. Vertices v_3 and v_4 form a terminal component. There exists a path from v_0 to v_4 yet there is no path from v_4 to v_5

The system associated to the graph in Figure 2.2 provides an explicit example of the contradiction argument of Proposition 7. The vertices v_3 and v_4 violate the condition of Proposition 7, since there is a path from the intake vertex v_1 to v_3 and v_4, and there is no path from v_3 and v_4 to the excretion vertex v_2; hence, the system does not admit any equilibrium. The proof argument, specialized to this example, is to notice that v_3 and v_4 form a terminal component with no excretion, and to look at the path v_0, v_1, v_2, v_4. Assuming that there is an equilibrium $\bar{x}$, one shows first that $\bar{x}_{v_1} > 0$ due to the intake, and then that $\bar{x}_{v_1} > 0$ implies also $\bar{x}_{v_2} > 0$. Notice that $\dot{x}_{v_3} + \dot{x}_{v_4} \geq S_{v_2,(v_2,v_4)} f_{v_2,v_4}$ (in this particular example, equality is actually true), and the latter is > 0 since $\bar{x}_{v_2} > 0$, thus contradicting the fact that $\bar{x}$ is an equilibrium:

the mass $x_{v_3} + x_{v_4}$ grows unbounded.

The same system of Figure 2.2 shows the necessity of the assumption that $H_e(x) = 0$ when $x_v = 0$, $e = (v, w)$ for Proposition 7 to hold. Indeed assume that $H_{(v_2,v_5)}(x) = 1$ for every x such that $x_{v_2} = 0$ and that the other functions H_e, $e = (v, w)$, satisfy $H_e(x) = x_v$. Let f be a flux vector with strictly positive components such that $f_{(v_1,v_2)} = f_{(v_2,v_5)}$. Then $\bar{x} = (\frac{f_{(v_0,v_1)}}{f_{(v_1,v_2)}}, 0, \frac{f_{(v_4,v_3)}}{f_{(v_3,v_4)}}, 1)$ is an equilibrium.

Another example not satisfying the assumptions of Proposition 7 is shown in Figure 2.3. This system admits an equilibrium under Assumption (A) even though the vertex v_3 does not have a path to v_5. This is because there does not exist a path from v_0 to v_3.

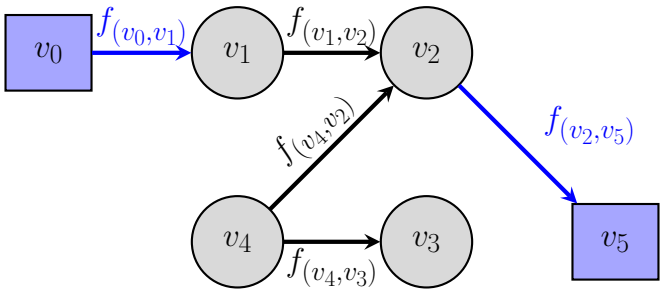


Figure 2.3: A directed graph where vertices v_3 and v_4 do not have a path from v_0 and also have no path to v_5. For an equilibrium, $\bar{x}$, of this system under Assumption (A), $\bar{x}_{v_4} = 0$ and $\bar{x}_{v_3} \geq x_{v_3}(0)$.

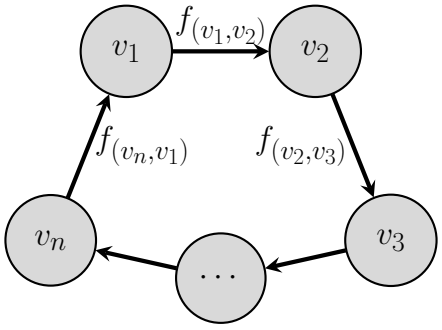


Figure 2.4: A directed cycle graph $G = (V, E)$ with n vertices and no intakes nor excretions. On such a LIFE system one can prescribe any desired dynamics.

For general systems (1.3) with Assumption (A) it is not possible to prove other

general conclusions about equilibria, beside Propositions 6 and 7. Indeed consider the simple metabolic network given in Figure 2.4. Under Assumption (A), the dynamics is written as:

$$\begin{cases} \dot{x}_{v_1} &= -f_{(v_1,v_2)} \cdot H_{(v_1,v_2)}(x) + f_{(v_n,v_1)} \cdot H_{(v_n,v_1)}(x) \\ \dot{x}_{v_2} &= -f_{(v_2,v_3)} \cdot H_{(v_2,v_3)}(x) + f_{(v_1,v_2)} \cdot H_{(v_1,v_2)}(x) \\ \dot{x}_{v_3} &= -f_{(v_3,v_4)} \cdot H_{(v_3,v_4)}(x) + f_{(v_2,v_3)} \cdot H_{(v_2,v_3)}(x) \\ &\vdots \\ \dot{x}_{v_n} &= -f_{(v_n,v_1)} \cdot H_{(v_n,v_1)}(x) + f_{(v_{n-1},v_n)} \cdot H_{(v_{n-1},v_n)}(x) \end{cases} . \tag{2.3}$$

We want to show that for any dynamical system in $\mathbb{R}^n$ on a compact set there exists an equivalent dynamics defined on the cycle graph of Figure 2.4. In other words, Asumption (A) is so general that we give Proposition 8 to show arbitrary dynamics can be defined, and we focus on more specialized cases (Assumption (B), and (C)). More precisely for every general dynamics

$$\begin{cases} \dot{x}_{v_1} &= F_1(x_{v_1}, x_{v_2}, \ldots, x_{v_n}) \\ \dot{x}_{v_2} &= F_2(x_{v_1}, x_{v_2}, \ldots, x_{v_n}) \\ &\vdots \\ \dot{x}_{v_n} &= F_n(x_{v_1}, x_{v_2}, \ldots, x_{v_n}) \end{cases} . \tag{2.4}$$

we look for a choice of the functions H_e and the fluxes f_e realizing such equivalence. Note that from (2.1), we have $x_{v_n} = C - x_1 - \cdots - x_{n-1}$ which implies that

$$\dot{x}_{v_n} = -\dot{x}_{v_1} - \cdots - \dot{x}_{v_{n-1}} = \sum_{i=1}^{n-1} -F_i. \tag{2.5}$$

Define the set $T = \cup_{i=1}^{n} \{x : x_j \geq 0 \text{ for all } j,\ x_i = 0 \text{ and } x_1 + \cdots + x_n \leq 1\}$ then we

have the following:

Proposition 8. *Consider system* (2.4) *and assume* $F_1 = \cdots = F_n = 0$ *on the set* T. *Then there exist functions* H_e *and fluxes* f_e *such that the dynamics* (2.3) *is equivalent to* (2.4) *on the bounded set delimited by* T.

Proof. Assign the functions $H_{(v_i,v_j)}(x)$ according to the following rule:

$$\begin{aligned} H_{(v_i,v_{i+1})}(x) &= \sum_{k\le i}[F_k]_- + \sum_{\ell>i}[F_\ell]_+ \quad \text{if } \ i<n, \\ H_{(v_n,v_1)}(x) &= \sum_{k=1}^{n}[F_k]_+, \end{aligned} \tag{2.6}$$

where $[F]_+ = \max\{F, 0\}$ and $[F]_- = \max\{-F, 0\}$. It is easy to verify that $\dot{x}_{v_i} = F_i(x)$ for $i < n$ and $\dot{x}_{v_n} = -\sum_{i=1}^{n} F_i = F_n(x)$ using (2.5). Moreover, because of the assumption on F_i's, the functions H_e satisfy (A), thus we are done. □

2.2.3 Special LIFE systems

Here we introduce a special class of systems of type (1.3) with simplified dynamics. We consider Assumption (A), and we impose further restriction on the functions $H_e(x)$: for an edge $e = (v, w)$, we assume H_e to depend on x_v only, and moreover we impose the scalar function $H_e(x_v)$ to be strictly increasing. More precisely, Assumption (B) is the following.

(B) It holds

$$S_{ve}(x) = \begin{cases} -H_e(x_v) & e = (v,w),\ v \in V, w \in V \cup \{v_{n+1}\} \\ H_e(x_w) & e = (w,v),\ w \in V \\ 1 & e = (v_0, v)\ v \in I \\ 0 & \text{otherwise,} \end{cases} \tag{2.7}$$

and each H_e is a positive, differentiable, and strictly increasing function $H_e : \mathbb{R} \to \mathbb{R}$, with $H_e(0) = 0$.

A typical example of a system verifying (B) is given by metabolic networks with Hill functions representing reactions, i.e. $H_e(x_v) = \frac{x_v^{p_e}}{K + x_v^{p_e}}$ with $p_e \in \mathbb{N}$, where $K > 0$ is the dissociation constant.

A further simplification occurs in the case where $H_e(x_v)$ is the same function $H_v(x_v)$ for all edges e having v as an initial vertex, i.e. such that $e = (v, w)$ for some w. This gives Assumption (C), as follows.

(C) It holds

$$S_{ve}(x) = \begin{cases} -H_v(x_v) & e = (v, w),\ v \in V, w \in V \cup \{v_{n+1}\} \\ H_w(x_w) & e = (w, v),\ w \in V \\ 1 & e = (v_0, v)\ v \in I \\ 0 & \text{otherwise,} \end{cases}$$

and each H_v is a positive, differentiable, and strictly increasing function $H_v : \mathbb{R} \to \mathbb{R}$, with $H_v(0) = 0$.

Under Assumption (C), the system $\dot{x} = S(x)f$ can be equivalently re-written as

$$\dot{x} = J(f)h(x) + \phi\,, \tag{2.8}$$

where $J(f) \in M_{n\times n}$ is defined by

$$J_{ij}(f) = \begin{cases} f_{(v_j, v_i)} & \text{if } (v_j, v_i) \in E \\ -\sum_{w:(v_i,w)\in\tilde{E}} f_{(v_i,w)} & \text{if} j = i \\ 0 & \text{otherwise,} \end{cases}$$

where $h(x)$ is a vector of size n given by $h_i(x) = H_{v_i}(x_{v_i})$ and ϕ is a vector of size n given by $\phi_i = f_{(v_0,v_i)}$ if $(v_0, v_i) \in \tilde{E}$, and $\phi_i = 0$ otherwise.

Finally, the simplest class of LIFE models we consider are *linear* systems, namely systems satisfying Assumption (C), where each $H_v(x_v)$ is the identity function $H_v(x_v) = x_v$.

Example 2.1 (continued). This example is a linear LIFE system. We can equivalently re-write its dynamics $\dot{x} = S(x)f$ as $\dot{x} = J(f)x + \phi$, with

$$J(f) = \begin{pmatrix} -f_{(v_1,v_2)} - f_{(v_1,v_3)} & 0 & 0 & f_{(v_4,v_1)} \\ f_{(v_1,v_2)} & -f_{(v_2,v_3)} & 0 & 0 \\ f_{(v_1,v_3)} & f_{(v_2,v_3)} & -f_{(v_3,v_4)} - f_{(v_3,v_5)} & 0 \\ 0 & 0 & f_{(v_3,v_4)} & -f_{(v_4,v_1)} \end{pmatrix}, \ \phi = \begin{pmatrix} f_{(v_0,v_1)} \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

2.3 Equilibria for fixed metabolites

In this section we consider equilibrium solutions of the system (1.3) satisfying Assumption (A). In general one is interested in conditions guaranteeing existence of an equilibrium and also in conditions necessary for uniqueness and stability of such an equilibrium. There are two problems: for fixed metabolite concentrations x find all flux vectors f for which x is an equilibrium, i.e. $\dot{x} = S(x) \cdot f = 0$, and, vice versa, for a fixed flux vector f, find all x that are equilibria. In this section we focus on the first, while the latter is investigated in Section 2.4.

The set of flux vectors for which x is an equilibrium formed by all vectors f that solve the equation $S(x) \cdot f = 0$, i.e. the nullspace $\mathcal{N}(S(x))$ of $S(x)$. First, we discuss the dimension of $\mathcal{N}(S(x))$. Then, since fluxes must be positive (to have a correct biological meaning), we focus on describing the cone $\mathcal{N}(S(x)) \cap (\mathbb{R}_+)^m$. Recall that, under Assumption (A), we can rewrite $S(x)$ as $S(x) = \Gamma D(x)$ where $D(x) \in M_{m\times m}$ is

a diagonal matrix with $H_e(x)$'s as entries and Γ is obtained from the incidence matrix of $\tilde{G}$ by removing the first and last rows; in case there are no intakes nor excretions, then Γ is the incidence matrix of G. Assuming that x has strictly positive entries, (A) implies that all diagonal elements of $D(x)$ are strictly positive, so that $D(x)$ is invertible. Hence, for $\Gamma \cdot (D(x) \cdot f) = 0$ to have non trivial solutions, the nullspace of Γ must have dimension greater than zero.

We extend one of the results of [97] to get:

Proposition 9. *Consider the system* (1.3) *with no intakes nor excretions satisfying Assumption (A) and let G be the the associated graph. Let $x \in \mathbb{R}^n$ have strictly positive entries and $S(x) \in M_{n\times m}$ be the stoichiometric matrix, then*

$$\operatorname{rank}(S(x)) = n - \ell,$$

where ℓ is the number of weakly connected components of G.

Proof. We have $S(x) = \Gamma D(x)$. As shown in Proposition 4.3 of [12] the rank of an incidence matrix Γ is $\operatorname{rank}(\Gamma) = n - \ell$. Since $D(x)$ is full rank we have $\operatorname{rank}(S(x)) = n - \ell$. □

Next we consider systems that contain both intakes and excretions. It was shown in [97] that if intakes and excretions are added to a graph satisfying the conditions of Proposition 9 then $\operatorname{rank}(S(x)) = n$. We extend this result to get:

Proposition 10. *Consider the system* (1.3) *satisfying Assumption (A) and let G be the the associated graph. Let $x \in \mathbb{R}^n$ have strictly positive entries and $S(x) \in M_{n\times m}$ be the stoichiometric matrix, then* $\operatorname{rank}(S(x)) = n - k$*, where k is the number of weakly connected components containing neither intake nor excretion vertices.*

Proof. It was shown in [97] that if G is weakly connected and contains an intake vertex then $\operatorname{rank}(S(x)) = n$. The same argument from [97] can be also used for an

excretion vertex and so if G is weakly connected and contains an excretion vertex we also have $\text{rank}(S(x)) = n$. In the case where G is not weakly connected $S(x)$ can be rewritten as a block diagonal matrix

$$S(x) = \begin{pmatrix} S_1(x) & & 0 \\ & \ddots & \\ 0 & & S_\ell(x) \end{pmatrix}$$

where each diagonal element $S_i(x)$ is the stoichiometric matrix of a weakly connected component. If $S_i(x)$ contains an intake or excretion vertex then $rank(S_i(x)) = n_i$, where n_i is the number of metabolites in $S_i(x)$. If $S_i(x)$ contains neither intake nor excretion vertices then $rank(S_i(x)) = n_i - 1$. This implies that $S(x)$ has n_i linearly independent rows for each $S_i(x)$ with intake or excretion vertices and $n_i - 1$ linearly independent rows for each $S_i(x)$ with neither intake nor excretion vertices. Let k represents the number of $S_i(x)$ with neither intake nor excretion vertices, the total number of linearly independent rows in S is now expressed $(\sum_{i=1}^{\ell}(n_i)) - k = n - k$. Thus $\text{rank}(S(x)) = n - k$. □

Notice that the rank of $S(x)$ depends on the number of weakly connected components of the graph, which is the same irrespective of the orientation of edges. However, when we focus on the existence of non-trivial positive flows admitting an equilibrium with strictly positive entries, the orientation of edges does matter, as we can see e.g. from Proposition 7, where a necessary condition is given for existence of equilibria, in terms of existence of suitable paths.

2.3.1 Network flows

The problem of finding positive flows admitting an equilibrium with strictly positive entries has been extensively studied in the operations research literature under the

name of 'network flow' problems [61]. In network flow problems one considers a directed graph where edges represent flows between the vertices. The maximum flow which an edge can support is called the capacity of the edge. In addition to capacity each edge may also have a cost associated to it. Network flow problems assume that the system is at equilibrium with respect to the vertices i.e. that the flow entering and leaving a vertex must be the same. A flow that satisfies this assumption is called a feasible flow. Finding feasible flows is the same as finding flows in the cone $\mathcal{N}(S(x)) \cap (\mathbb{R}_+)^m$ with the further constraint that each flux must be less than its capacity.

A flow is a mapping from $f : E \to (\mathbb{R}_+)^m$ that satisfies $0 \leq f(v_i, v_j) \leq c(v_i, v_j)$ (where $c(v_i, v_j)$ is the capacity of the edge (v_i, v_j)) and $\sum_{(v_i,v_j)\in E} f(v_i, v_j) - \sum_{(v_j,v_k)\in E} f(v_j, v_k) = 0$ [93]. One of the most common network flow problems is to find the maximum flow of the network, i.e. the largest amount of total flow from a source to a sink. In [47] the authors consider a network which contains exactly one source (v_0) and one sink (v_{n+1}). They then prove the following result, known as max-flow min-cut theorem, which characterizes the maximum flow as the minimum cut cacapity, where a cut set is a set of edges whose removal disconnects the source from the sink, and the capacity of a cut is the sum of the capacities of its edges.

Proposition 11 (Max-Flow Min-Cut Theorem)**.** *The maximum flow value obtainable in a network is the minimum capacity of all cut sets which disconnect v_0 and v_{n+1}.*

An implication of the max-flow min-cut theorem is that a feasible flow exists if there is a path from v_0 and v_{n+1}. The following proposition extends this relationship:

Proposition 12. *Given a LIFE system with graph $\tilde{G}$ and stoichiometric matrix $S(x)$ satisfying (A), fix an $x \in (\mathbb{R}_+)^n$ with strictly positive entries and for every $v \in I$ fix $\bar{f}_{v_0,v} > 0$, i.e. fix arbitrary values for the intake flows. There exists $f \in \mathcal{N}(S(x)) \cap (\mathbb{R}_+)^m$ such that $f_{v_0,v} = \bar{f}_{v_0,v}$ for all $v \in I$ if and only if for each $v \in I$ there exists a path to X.*

Proof. Consider the maximum flow problem on the graph $\tilde{G}$, where edges (v_0, v) have capacity $\bar{f}_{v_0,v}$, and all other edges have infinite capacity. The feasible flows φ for this network are in one-to-one correspondence with the equilibrium flows $f \in \mathcal{N}(S(x)) \cap (\mathbb{R}_+)^m$ such that $f_{v_0,v} \leq \bar{f}_{v_0,v}$ for all $v \in I$; the correspondence is simply given by $f_{v_0,v} = \varphi_{v_0,v}/H_{v_0,v}(x)$ for all $v \in I$ and $f_{v,w} = \varphi_{v,w}/H_{v,w}(x)$ for all $w \in V$. If for all $v \in I$ there is a path from v to X (and hence a path from v to v_{n+1}), the minimum cut is the removal of all edges (v_0, v). The maximum flow φ^* then has $\varphi^*_{v_0,v} = \bar{f}_{v_0,v}$, thus also ensuring the existence of an equilibrium flow f^* satisfying the same.

If for some $v_i \in I$ there is no path from v to X, then all feasible flows φ satisfy $\varphi_{v_0,v} = 0$, and hence all equilibrium flows f satisfy $f_{v_0,v} = 0$ which contradicts the assumption. □

2.3.2 Extreme pathway algorithm for calculating the positive basis

There are many standard methods for computing a basis of the nullspace of a matrix [100], and hence to describe $\mathcal{N}(S(x))$. Since we are interested in positive fluxes, we are rather interested in the cone $\mathcal{N}(S(x)) \cap (\mathbb{R}_+)^m$. One method to describe this cone is to look for a positive basis, which is a minimal set of vectors generating the whole cone via linear combinations with positive scalars, called positive combinations. The use of positive combinations ensures that all generated vectors belong to the cone. Minimality is equivalent to ask for the vectors to be positively linearly independent, i.e. no vector can be expressed as a positive combination of the others. Since we allow only positive combinations (and not every linear combination), typically a positive basis has more vectors than a traditional basis. In [125] the authors prove that a flux cone has a positive basis, and that the vectors of the positive basis are unique up to multiplication by a positive scalar.

Finding a positive basis for a cone is not as easy as finding a traditional basis. In [107], Palsson describes a method to find a positive basis by using extreme pathways of a metabolic network. The extreme pathways are a unique set of positively independent flux vectors that represent the edges of the cone $\mathcal{N}(S(x)) \cap (\mathbb{R}_+)^m$ [107, 125].

Here we summarize a method to build extreme pathways described by [125]. We start considering the equation $V_0 \cdot S(x)^T = C_0$, and the solution given by $V_0 = I_m$, the $m \times m$ identity matrix, $S(x)$ the stoichiometric matrix and $C_0 = S(x)^T$. At each iteration new matrices $V_i \in M_{n_i \times m}$, $C_i \in M_{n_i \times n}$ are defined which satisfy $V_i \cdot S(x)^T = C_i$.

Algorithm 1. *The algorithm consists of the following steps:*

Step 1. *Select from C_i the first non-zero column (lexicographically), whose corresponding metabolite is neither an intake vertex nor an excretion vertex, say j_i. If no such column exists jump to Step 5.*

Step 2. *Add to C_i all possible new rows, obtained using positive linear combinations of two other rows, which have a zero on the column j_i. Then define C_{i+1} by removing the old rows used to generate the new ones. Notice that C_{i+1} has all zeros in column j_i.*

Step 3. *Define V_{i+1} by adding to V_i the rows generated by the same combination as those of Step 2 (to keep the validity of the equation $V \cdot S(x)^T = C$) and removing the old rows.*

Step 4. *Remove positively linearly dependent rows from V_{i+1}. Remove the corresponding rows from C_{i+1}.*

Step 5. *Repeat the steps 1-4 now considering first excretion vertices and then intake vertices until all columns are* 0.

If the algorithm stops at step m, then $C_m = 0$ (the matrix with all zero entries). The rows of V_m are the vectors of a positive basis. Notice also that the number of

rows n_i may increase or decrease during the various steps.

Originally in [107] the extreme pathways method was proposed for stoichiometric matrices not dependent on x. We apply it here to the most general case of systems (1.3) with Assumption (A), by fixing a desired metabolite variables vector $x \in (\mathbb{R}_+)^n$ and using the extreme pathways method to characterize all fluxes vectors with positive entries such that the corresponding LIFE system (1.3) admits x as an equilibrium.

To illustrate the process we report an example using the Reverse Cholesterol Transport Network (RCT) from [96]. The RCT network is shown in Figure 2.5.

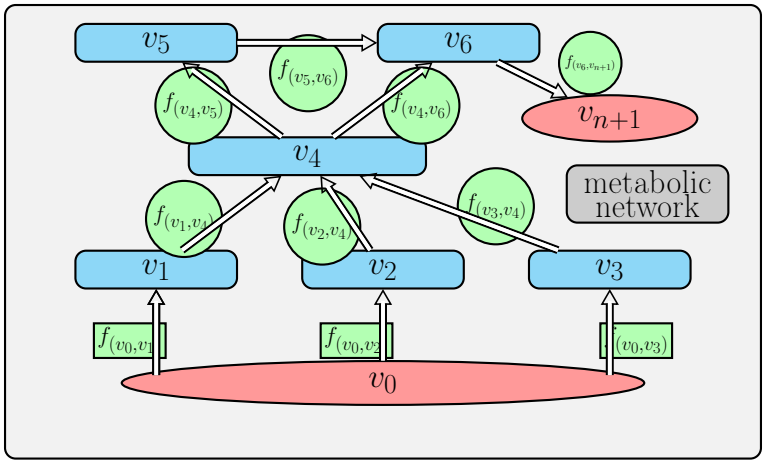


Figure 2.5: Reverse Cholesterol Transport Network from [96]. This network contains 6 vertices which represent metabolites, 10 edges which represent fluxes and 2 virtual vertices v_0, v_{n+1}. There are three intake vertices v_1, v_2, v_3 and 1 excretion vertex v_6.

Example 2. Finding Extreme Pathways of RCT network

We use Palsson's algorithm for finding extreme pathways on the stoichiometric

matrix from the RCT network. The stoichiometric matrix for RCT is given by:

$$S(x) = \begin{pmatrix} 1 & 0 & 0 & -x1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -x2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -x3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & x1 & x2 & x3 & -x4 & -x4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & x4 & 0 & -x5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & x4 & x5 & -x6 \end{pmatrix}$$

For space, the complete calculations are continued in the Appendix.

2.3.3 Extreme pathways provide a positive basis

Let us first recall Farkas' Lemma [52]. Here we use the notation $x \geq 0$ to indicate that every entry in the vector x is positive.

Lemma 3 (Farkas' Lemma)**.** *Let $A \in M_{n\times m}$ and $b \in R^n$. For the equation $Ax = b$ exactly one is true:*

1. *There exists $x \in \mathbb{R}^m$ such that $Ax = b$ and $x \geq 0$.*
2. *There exists $y \in \mathbb{R}^n$ such that $A^T y \geq 0$ and $b^T y < 0$.*

Proposition 13. *The extreme pathways of $S(x)$ form a positive basis of $\mathcal{N}(S(x)) \cap (\mathbb{R}_+)^m$.*

Proof. There are three statements which must be shown to prove this Proposition.

Statement 1: The extreme pathway vectors (v_i) are positively independent (systematically independent in terminology used in [107]).

Statement 2: The span of the extreme pathways vectors is contained in $(\mathcal{N}(S(x)) \cap (\mathbb{R}_+)^m)$ i.e. $\sum_{\lambda_i \geq 0} \lambda_i v_i \subset (\mathcal{N}(S(x)) \cap (\mathbb{R}_+)^m)$.

Statement 3: The extreme pathways vectors span $(\mathcal{N}(S(x)) \cap (\mathbb{R}_+)^m)$ i.e. $(\mathcal{N}(S(x)) \cap (\mathbb{R}_+)^m) \subset \sum_{\lambda_i \geq 0} \lambda_i v_i$.

Statement 1 follows from Step 4 of the algorithm.

Let us now prove Statement 2. Since the vectors corresponding to extreme pathways are positive combinations of rows of matrices with positive entries, then $v_i \in (\mathbb{R}_+)^m$ for all i. We have $V_n \cdot S(x)^T = C_n = 0$, equivalently $S(x) \cdot V_n^T = C_n^T = 0$. This implies that the columns of V_n^T in the nullspace of $S(x)$, thus the rows of V_n in the nullspace of $S(x)$. Since both $\mathcal{N}(S(x)$ and $(\mathbb{R}_+)^m$ are closed for positive linear combinations we are done.

Finally it remains to prove statement 3, in order to show $(\mathcal{N}(S(x)) \cap (\mathbb{R}_+)^m) \subset \sum_{\lambda_i \geq 0} \lambda_i v_i$ we proceed as follows. First we show that after one iteration of the algorithm the rows of V_1 form a positive basis for vectors w such that the j_1-th element of $S(x)w$ equals zero. After n iterations, we show that V_n is a positive basis for vectors $w \in \mathcal{N}(S(x))$.

For simplicity of notation, we assume $j_1 = 1$. Define W_1 to be the set of vectors w such that the first entry of $S(x)w$ equals zero. It is clear from the construction of V_1 that its rows belong to W_1. It remains to be shown that the rows of V_1 span W_1. Since $V_1 \cdot S(x)^T = C_1$, thus we can alternatively show that the columns of C_1^T span the subset of the range of $S(x)$ with first entry equal to zero. In other words, we must show that if $\sum_{\lambda_i \geq 0} \lambda_i c_{i,1} = 0$, where $c_{i,1}$ represents the first entry in the c_i-th row of C_0, that the sum $\sum_{\lambda_i \geq 0} \lambda_i c_i = 0$ can also be represented using rows from C_1. We assume an ordering such that $c_{i,1} \geq 0$ for $i \leq \ell$ and $c_{i,1} < 0$ for $i > \ell$, each row in C_1 can be represented by $c_i + \alpha_{i,j} c_j$ where $i \leq \ell, j > \ell$ and $\alpha_{i,j} = \frac{-c_{i,1}}{c_{j,1}} \geq 0$. We need to find $\mu_{i,j}$ such that,

$$\sum \mu_{i,j}(c_i + \alpha_{i,j} c_j) = \sum_{i=1}^{n} \lambda_i c_i. \tag{2.9}$$

We can split the sum on the right hand side by considering the positive and

negative entries separately,

$$\sum \mu_{i,j}(c_i + \alpha_{i,j}c_j) = \sum_{i=1}^{\ell} \lambda_i c_i + \sum_{i=\ell+1}^{n} \lambda_i c_i.$$

For λ_i, $i \in \{\ell+1, \ldots, n\}$ we can write:

$$\sum_{k=1}^{\ell} \mu_{k,i}\alpha_{k,i} = \lambda_i$$

which gives:

$$\sum_{k=1}^{\ell} \mu_{k,i}\ (-c_{k,1}) = c_{i,1}\lambda_i,\ i \in \{\ell+1, \ldots, n\}. \tag{2.10}$$

For λ_i, $i \in \{1, \ldots, \ell\}$ we can write:

$$\sum_{k=\ell+1}^{n} \mu_{i,k} = \lambda_i$$

which gives:

$$\sum_{k=\ell+1}^{n} \mu_{i,k}c_{i,1} = c_{i,1}\lambda_i,\ i \in \{1, \ldots, \ell\}. \tag{2.11}$$

The equations (2.10), (2.11) can be written in the following way,

$$\begin{matrix} \ell \text{ rows} \\ n-\ell \text{ rows} \end{matrix} \left\{ \begin{pmatrix} [c_{1,1}] & 0 & \cdots & 0 \\ 0 & [c_{2,1}] & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & [c_{\ell,1}] \\ -diag(c_{1,1}) & -diag(c_{2,1}) & \cdots & -diag(c_{\ell,1}) \end{pmatrix} \right. \cdot \begin{pmatrix} \mu_{1,\ell+1} \\ \mu_{1,\ell+2} \\ \vdots \\ \mu_{1,n} \\ \mu_{2,\ell+1} \\ \vdots \\ \mu_{2,n} \\ \vdots \\ \mu_{\ell,n} \end{pmatrix} = \begin{pmatrix} c_{1,1}\lambda_1 \\ c_{2,1}\lambda_2 \\ \vdots \\ c_{n,1}\lambda_n \end{pmatrix} \tag{2.12}$$

where $[c_{i,1}]$ represents a $1 \times (n-\ell)$ vector with all entries equal to $c_{i,1}$ and $diag(c_{i,1}) \in$

$M_{(n-\ell)\times(n-\ell)}$ a diagonal matrix with $c_{i,1}$ on the diagonal. Let A, b be the matrix on the left-hand side and the vector on the right-hand of equation (2.12), respectively. We can then apply Farkas' Lemma. Consider a vector $y \in \mathbb{R}^n$ such that $A^T \cdot y \geq 0$, then for the first row we have $c_{1,1}y_1 - c_{1,1}y_{\ell+1} \geq 0$. Further for any $i = 1, \ldots, \ell$ and $j = \ell + 1, \ldots, n$ we have $c_{i,1}y_i - c_{i,1}y_j \geq 0$ which implies $y_i \geq y_j$. Setting $\bar{y} = max(y_{\ell+1}, \ldots, y_n)$ we have $y_i \geq \bar{y}$ for $i = 1, \ldots, \ell$. Since all $c_{j,1} < 0$ for $j = \ell+1, \ldots, n$ and all $\lambda_j \geq 0$ we have that $y_j c_{j,1}\lambda_j \geq \bar{y}c_{j,1}\lambda_j$. Next we define $\bar{\mathbf{y}} = (\bar{y}, \ldots, \bar{y}) \in \mathbb{R}^n$ and note that $b^T y \geq b^T \bar{\mathbf{y}}$. Since $\sum_{\lambda_i \geq 0} \lambda_i c_{i,1} = 0$ we have $b^T \bar{\mathbf{y}} = 0$. Therefore condition 2 of Farkas' Lemma fails. This implies that condition 1 is true and thus there exists a set of positive μ's that solve the system (2.12). Since a set of μ's can be found we have that after the first iteration of the algorithm the rows of V_1 span W_1. By similar argument the rows of the V_i in subsequent iterations span the set of vectors w such that the first i entries of $S(x)w$ equal zero. After n iterations, the rows of V_n span $\mathcal{N}(S(x)) \cap (\mathbb{R}_+)^m$, thus we are done. □

Example 3. Comparison of the standard and positive bases

Using the RCT network, an example is presented which compares a standard basis of the nullspace and positive basis for the nullspace. Here we provide the calculations needed to show *Statement 2* and *Statement 3* of Proposition 13 for the RCT. The full details for this example are contained in the Appendix.

2.4 Equilibria and asymptotic behavior of metabolites for fixed fluxes

In this section we characterize equilibria and the stability of LIFE systems with a fixed flux vector. We start with linear systems, and then we consider special LIFE systems, with Assumptions (B) and (C). LIFE systems satisfying at least Assumption (A) are known as *compartmental systems* in the automatic control community [65, 20].

We build upon the well-established results on linear compartmental systems to get a full understanding of linear LIFE systems, as well as for special LIFE systems satisfying Assumption (C). Notice that weakly connected components of $\tilde{G}$ correspond to subsystems having no interaction with each other, so that we can study each weakly connected component separately from the others. Hence, throughout this section, we assume $\tilde{G}$ is weakly connected, without loss of generality.

2.4.1 Linear systems

In this section we study the properties of linear systems, i.e. special LIFE system (1.3) satisfying Assumption (C) with fluxes $H_v(x_v) = x_v$, or, equivalently, a system (2.8), with $h(x) = x$. We first focus on metabolic networks with no intakes nor excretions, recalling results from a rich literature from different communities. For the case with intakes and excretions, we make use of results for Compartmental Systems [21].

Linear system without intakes and excretions

In the case with no intakes nor excretions (also known as free closed system), the linear LIFE dynamics become $\dot{x} = J(f)x$. Notice that $J(f)$ is a *Metzler matrix* (i.e. has non-negative off-diagonal entries), and all its columns sum to zero ($\mathbf{1}^T J(f) = \mathbf{0}^T$).

This system is known as *Laplacian* dynamics, because $L = -J(f)^T$ is a weighted Laplacian of the graph G defined as follows. Given the graph G and given strictly positive weights f_e associated with its edges (for us, the weights are the fluxes), the *weighted adjacency matrix* A is defined by $A_{ij} = f_{(v_i,v_j)}$ if $(v_i, v_j) \in E$, and $A_{ij} = 0$ otherwise. The corresponding weighted Laplacian is $L = D - A$, where $D = \mathrm{diag}(A\mathbf{1})$ is the diagonal matrix containing the weighted out-degrees, i.e. row-sums of A.

The eigenvalues of the Laplacian, and in particular its eigenvalue 0 and the corresponding eigenspace which is actually the set of equilibria of $\dot{x} = J(f)x$, have been studied in graph theory (see e.g. [25]). The study of the spectrum of the Laplacian has

received extensive attention also in the automatic control community, on one hand because of the recent interest in the so-called consensus dynamics (see [20, Chapter 7]) $\dot{x} = -Lx$ (which are different from the dynamics $\dot{x} = J(f)x$, since $L = -J(f)^T$, but the eigenvalues of $-L$ and $J(f)$ are the same), and on the other hand because of the interest in the linear compartmental system $\dot{x} = J(f)x$ (see [20, Chapter 9]).

Laplacian dynamics have been introduced in the mathematical biology literature by [102], together with a complete study of their equilibria and convergence properties. Notice that in [102] the matrix $J(f)$ itself is called a Laplacian, while in the graph theory and control theory communities the name Laplacian refers to $L = -J(f)^T$.

The dynamics $\dot{x} = J(f)x$ is also very related to *continuous-time Markov chains* (see e.g. the textbook [27]). Thus we elaborate on this, a homogeneous continuous-time Markov chain over a finite state $V = \{1, \dots, n\}$ is a stochastic process $X(t)$ taking values in V, that satisfies the Markov property:
for all times $t_0 \leq t_1 \leq \cdots \leq t_h \leq t_{h+1}$,

$$\Pr(X_{t_{h+1}} = i_{h+1} \mid X_{t_0} = i_0, X_{t_1} = i_1, \dots, X_{t_h} = i_h) = \Pr(X_{t_{h+1}} = i_{h+1} \mid X_{t_h} = i_h),$$

and which is homogeneous in time:

$$\Pr(X_{t_{h+1}} = i_{h+1} \mid X_{t_h} = i_h) = \Pr(X_{t_{h+1}-t_h} = i_{h+1} \mid X_{t_0} = i_h) = P_{i_h, i_{h+1}}(t_{h+1} - t_h)\,.$$

$P(t)$ is the matrix whose (i,j)-th entry represents the probability of transition from state i to state j in an interval of time of length t. $P(t)$ is assumed to be right-differentiable, and with time derivative defined by $\dot{P}(t) = \lim_{h \to 0^+}(P(t+h) - P(t))/h$. The Markov chain is fully described by its *generator matrix* (or *transition rate matrix*) defined by $Q = \dot{P}(0)$. Indeed, $P(t)$ is related to Q by the *Kolmogorov forward equation* $\dot{P}(t) = P(t)Q$. The unique solution of the Cauchy problem $\dot{P}(t) = P(t)Q$ with $P(0) = I$ is $P(t) = e^{Qt}$. Denoting by $\pi(t)$ a column vector whose ith entry is

$\pi_i(t) = \Pr(X(t) = i)$, we have that $\pi^T(t) = \pi^T(0)P(t) = \pi^T(0)e^{Qt}$ and hence is a solution of $\dot{\pi}^T = \pi^T Q$ with given initial condition $\pi(0)$.

The generator matrix Q is a Metzler matrix whose rows sum to 0 ($Q\mathbf{1} = \mathbf{0}$), and hence there is an immediate equivalence between the dynamics $\dot{\pi}^T = \pi^T Q$ and the linear LIFE system $\dot{x} = J(f)x$, simply by taking $Q = J(f)^T$. In Markov chains, one looks at π being a probability vector, namely having positive entries and $\sum_i \pi_i = 1$. In LIFE systems, we are also interested in a vector x with positive entries, but the total mass could be arbitrary, so that one should set $\pi_i(0) = x_{v_i}(0)/m_0$, with $m_0 = \sum_i x_{v_i}(0)$.

Based on all this rich literature, we recall here the results about equilibria and their stability. The spectrum of $J(f)$ is characterized as follows (see e.g. [102, Proposition. 1]):

Proposition 14. *Assume there are no intakes nor excretions, then:*

- *All eigenvalues of $J(f)$ are either 0 or have strictly negative real part.*
- *The dimension of the nullspace of $J(f)$ is equal to the algebraic multiplicity of the 0 eigenvalue [1], and is equal to the number of terminal components in G.*
- *Moreover, denoting by $G_1, \dots, G_k$ the terminal components of G, there exists a basis of the nullspace of $J(f)$ composed of vectors $\pi_1, \dots, \pi_k$, such that vector π_i has strictly positive entries in correspondence of vertices in G_i, and has vanishing entries otherwise.*

It is customary to 'normalize' the vectors $\pi_1, \dots, \pi_k$ so that their entries sum to 1, so that they can be interpreted as probability vectors. Under this choice, the restriction of vector π_i to the terminal component G_i is the stationary distribution of the Markov chain restricted to such component, i.e. its unique equilibrium with mass

[1] The algebraic multiplicity of an eigenvalue is its multiplicity as a root of the characteristic polynomial, while its geometric multiplicity is the dimension of the corresponding eigenspace. In the case of the zero eigenvalue, the eigenspace is the nullspace of the matrix.

1 (uniqueness is obtained from the fact that the terminal component is a strongly connected component).

Notice that the dimension of the nullspace of an incidence matrix Γ (and hence of a stoichiometrix matrix $S(x)$) is related to the number of weakly connected components, while the dimension of the nullspace of a Laplacian matrix is related to strongly connected components, and not all components matter, but only the terminal ones.

In the case where G has a unique terminal component, the nullspace of $J(f)$ has dimension one: the equilibrium is unique, up to a multiplicative factor which is the initial total mass $m_0 = \sum_i x_{v_i}(0)$. In the case where G is strongly connected, the whole graph is a single strongly connected component. In this case, not only the nullspace of $J(f)$ has dimension one, but also it is generated by a vector whose entries are all strictly positive.

The spectral properties of $J(f)$ given in Proposition 14 fully characterize the asymptotic behavior of the linear system $\dot{x} = J(f)x$, by standard theory of linear systems, giving the following:

Proposition 15. *Consider the linear LIFE system with no intakes nor excretions and denote by $G_1, \ldots, G_k$ the terminal components of the associated graph G. The following properties hold:*

- *The total mass of the system $m = \sum_{v \in V} x_v = m_0$ is constant in time.*
- *From any positive initial condition $x(0)$, the system converges to an equilibrium, having strictly positive entries in correspondence of vertices of terminal components, and 0 elsewhere.*
- *Moreover, if there is a unique terminal component, then the equilibrium is uniquely determined by the initial total mass m_0.*

Linear system with intakes and excretions

We now focus on linear systems with intakes and/or excretions, using results from linear compartmental systems as summarized in [20, Chapter 9]. A first remark is that, having introduced a single v_0 from which all intake edges are originated, and a single v_{n+1} to which all excretion edges are headed, we have a single weakly connected component containing intakes and/or excretions. We restrict our attention to such a weakly connected component, while other possible components with no intakes nor excretions have a behavior described in Sect. 2.4.1.

The dynamics are given by $\dot{x} = J(f)x + \phi$, where vector ϕ represents the intakes. Due to excretions $J(f)$ does not have all column-sums equal to 0. This means that $-J(f)^T$ is not any more a Laplacian, but it is a *grounded Laplacian*. The term grounded Laplacian refers to a matrix obtained from a larger Laplacian matrix by deleting the row and column corresponding to a given vertex; the name 'grounded' has an interpretation for electrical networks, where this corresponds to connecting the given vertex to the ground. Consider the subgraph of $\tilde{G}$ where we have removed v_0 but not v_{n+1}, and consider the weighted Laplacian $L \in M_{(n+1)\times(n+1)}$ of such a graph (with weights equal to the fluxes), then define L_g by deleting the last row and last column (associated with v_{n+1}). The resulting grounded Laplacian is L_g is such that $J(f) = -L_g^T$.

The spectral properties of $J(f)$ are summarized as follows (see [20, Theorem 9.5 and Lemma 9.12]):

Proposition 16. *Consider a linear system with intakes and/or excretions, then:*

- *All eigenvalues of $J(f)$ are either 0 or have strictly negative real part.*
- *The dimension of the nullspace of $J(f)$ is equal to the algebraic multiplicity of the 0 eigenvalue, and is equal to the number of terminal components not containing any excretion.*

In particular, the following are equivalent:

(a) For every $v \in V$ there is a path from v to X.

(b) $J(f)$ is Hurwitz stable (i.e. all its eigenvalues have strictly negative real part).

(c) $J(f)$ is invertible.

Moreover, when $J(f)$ is invertible, all entries of $-J(f)^{-1}$ are positive; if G is strongly connected, they are strictly positive.

Equilibria of the dynamics $\dot{x} = J(f)x + \phi$ are the solutions of the linear system of equations $J(f)x = -\phi$. By Proposition 16, if all vertices v have a path to some excretions, then there is a unique equilibrium $\bar{x} = -J(f)^{-1}\phi$, and moreover all entries of $\bar{x}$ are positive. To study the general case, where vertices might or might not have a path to some excretions, it is convenient to partition the system into two subsystems, as follows. Partition the vertex set as $V = V_1 \cup V_2$, with V_1 the set of $v \in V$ such that there is a path from v to v_{n+1}. Without loss of generality, we can re-label vertices in V so as to have vertices $v_1, \ldots, v_r \in V_1$ and $v_{r+1}, \ldots, v_n \in V_2$. According to this decomposition, partition the vector x into two blocks x_1 corresponding to V_1 and x_2 corresponding to V_2, and similarly partition ϕ as ϕ_1, ϕ_2. Notice that there is no edge from V_2 to V_1, since an edge (w, v) with $v \in V_1$ implies that there is a path from v to v_{n+1} and also a path from w to v_{n+1}. Hence, we have:

$$J(f) = \begin{bmatrix} J_1 & 0 \\ J_{21} & J_2 \end{bmatrix},$$

and

$$\begin{cases} \dot{x}_1 = J_1 x_1 + \phi_1 \\ \dot{x}_2 = J_2 x_2 + J_{21} x_1 + \phi_2. \end{cases} \tag{2.13}$$

The first subsystem is called the *reduced system*, and its evolution is not affected by the second subsystem. The matrix J_1 is equal to the matrix $J(f)$ of the graph $\tilde{H}$ obtained from $\tilde{G}$ as follows: for any edge (v, w) with $v \in V_1$, $w \in V_2$, remove the edge (v, w) and add the edge (v, v_{n+1}) with flow $f_{v,v_{n+1}} = f_{v,w}$; then remove all vertices in V_2 and all corresponding edges. By definition of V_1, all vertices of $\tilde{H}$ have a path to v_{n+1}, and hence, by Proposition 16, J_1 is invertible and Hurwitz stable, and $-J_1^{-1}$ has positive entries.

For the second subsystem, it is easy to see that J_2 is the matrix $J(f)$ of the subgraph K of G corresponding to vertices in V_2. Hence, $L = -J_2^T$ is a Laplacian matrix, and by Proposition 14 its nullspace is generated by vectors $\pi_1, \ldots, \pi_k$, where $G_1, \ldots, G_k$ are the terminal components in K (same as the terminal components with no excretions in G), and each vector π_i has strictly positive entries corresponding to vertices in G_i and is 0 elsewhere. Choosing each π_i so that its entries sum to 1, the restriction of π_i to G_i is the stationary distribution of the corresponding Markov chain restricted to G_i, namely the Markov chain whose generator matrix is the transpose of the submatrix of $J(f)$ corresponding to G_i.

These remarks, together with standard tools of analysis of linear dynamical systems, lead to the following proposition:

Proposition 17 ([21], Theorem 9.13)**.** *Consider a weakly connected linear LIFE system with positive flows on all edges of $\tilde{G}$. From any positive initial condition, the reduced system (namely the subsystem connected to X) converges to its unique equilibrium with positive entries $\bar{x}_1 = -J_1^{-1}\phi_1$.*

If there are some vertices not connected to X, then:

- *If there is a terminal component G_T with no excretions such that there is a path from I to G_T, then the mass in G_T grows unbounded, and hence also $\lim_{t\to\infty} \|x_2(t)\| = +\infty$ and $\lim_{t\to\infty} \|x(t)\| = +\infty$.*

- *If for all terminal components with no excretions $G_1, \ldots, G_k$ there is no path from I to G_i, then the mass of the system remains bounded, and moreover $\lim_{t\to\infty} x_2(t)$ is some equilibrium point $\bar{x}_2$ (depending on the initial condition), such that all entries of $\bar{x}_2$ corresponding to non-terminal components are 0, while the restriction of $\bar{x}_2$ to a terminal component G_i is proportional to the stationary distribution of the Markov chain with generator matrix Q equal to the transpose of the submatrix of $J(f)$ corresponding to G_i.*

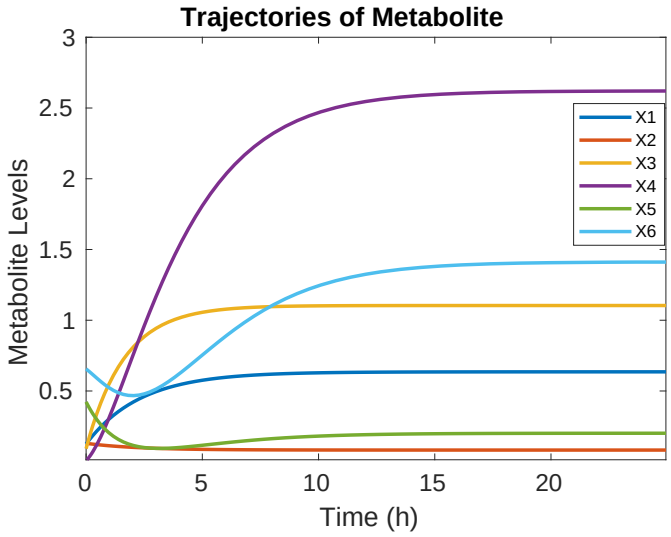


Figure 2.6: The trajectories of the values of metabolites over 25 hours.

Example 4. In this example a simulation is used to verify the calculated equilibrium found using Proposition 17. Here we use the RCT network shown in 2.5 and calculate the equilibrium. Because every vertex of the RCT is connected to X we have that

$\dot{x} = J(f)x + \phi$ with,

$$J(f) = \begin{pmatrix} -f_{(\mathrm{v}_1,\mathrm{v}_4)} & 0 & 0 & 0 & 0 & 0 \\ 0 & -f_{(\mathrm{v}_2\mathrm{v}_4)} & 0 & 0 & 0 & 0 \\ 0 & 0 & -f_{(\mathrm{v}_3\mathrm{v}_4)} & 0 & 0 & 0 \\ f_{(\mathrm{v}_1\mathrm{v}_4)} & f_{(\mathrm{v}_2\mathrm{v}_4)} & f_{(\mathrm{v}_3\mathrm{v}_4)} & -f_{(\mathrm{v}_4\mathrm{v}_5)} - f_{(\mathrm{v}_4\mathrm{v}_6)} & 0 & 0 \\ 0 & 0 & 0 & f_{(\mathrm{v}_4\mathrm{v}_5)} & -f_{(\mathrm{v}_5\mathrm{v}_6)} & 0 \\ 0 & 0 & 0 & f_{(\mathrm{v}_4\mathrm{v}_6)} & f_{(\mathrm{v}_5\mathrm{v}_6)} & -f_{(\mathrm{v}_6\mathrm{v}_{n+1})} \end{pmatrix},$$

$$\phi = \begin{pmatrix} f_{(\mathrm{v}_0,\mathrm{v}_1)} \\ f_{(\mathrm{v}_0,\mathrm{v}_2)} \\ f_{(\mathrm{v}_0,\mathrm{v}_3)} \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{2.14}$$

The vector of fluxes and initial metabolite values were randomized to obtain

$$f = \begin{pmatrix} f_{(\mathrm{v}_0,\mathrm{v}_1)} \\ f_{(\mathrm{v}_0,\mathrm{v}_2)} \\ f_{(\mathrm{v}_0,\mathrm{v}_3)} \\ f_{(\mathrm{v}_1,\mathrm{v}_4)} \\ f_{(\mathrm{v}_2,\mathrm{v}_4)} \\ f_{(\mathrm{v}_3,\mathrm{v}_4)} \\ f_{(\mathrm{v}_4,\mathrm{v}_5)} \\ f_{(\mathrm{v}_5,\mathrm{v}_6)} \\ f_{(\mathrm{v}_5,\mathrm{v}_6)} \\ f_{(\mathrm{v}_6,\mathrm{v}_{n+1})} \end{pmatrix} = \begin{pmatrix} 0.2729 \\ 0.0372 \\ 0.6733 \\ 0.4296 \\ 0.4517 \\ 0.6099 \\ 0.0594 \\ 0.3158 \\ 0.7727 \\ 0.6964 \end{pmatrix}, x_0 = \begin{pmatrix} 0.1253 \\ 0.1302 \\ 0.0924 \\ 0.0078 \\ 0.4231 \\ 0.6556 \end{pmatrix}. \tag{2.15}$$

Using these values the equilibrium was calculated to be $\bar{x} = (0.6354, 0.0824, 1.1040, 2.6211, 0.2015, 1.4122)$. The RCT network with the randomized initial val-

ues was simulated for 50 hours and the simulation results closely matched $\bar{x}$. The simulation results for the first 25 hours are shown in figure 2.6.

2.4.2 Special LIFE systems

In this section, we focus on special LIFE systems. We first recall some interesting results from [92] valid under Assumption (B), and then we exploit them together with the spectral properties of $J(f)$ described in Sect. 2.4.1 in order to fully characterize equilibria and convergence of special LIFE systems under Assumption (C). We refer the reader to [65] for some results valid under great generality, with assumptions less restrictive than Assumption (B). As for linear systems, we first notice that weakly connected components correspond to subsystems which have no influence on each other, and hence can be studied separately; if there are any weakly connected components with no intake nor excretion, they are subsystems with constant total mass.

Proposition 18. *([92, Theorem 6]) Consider the special LIFE system under Assumption (B) with no intakes nor excretions. The following properties hold:*

- *The total mass of the system $m = \sum_{v \in V} x_v$ is constant in time.*
- *From any positive initial condition $x(0)$, the system tends to the equilibrium set.*
- *Moreover, if there is a unique terminal component, then there exists a unique equilibrium with positive entries with the same mass as the initial mass, and the system converges to it.*

For the general case with intakes and/or excretions, the following result holds on the asymptotic behavior of the dynamics.

Proposition 19. *([92, Theorems 2 and 3]) For special LIFE systems under Assumption (B), with a positive initial condition,*

- *Trajectories are bounded if and only if there exists an equilibrium with positive entries;*

- *If trajectories are bounded, then they approach an equilibrium set for $t \to \infty$, and if moreover the equilibrium set consists of isolated points, then they converge to some equilibrium.*

A caveat reported in [92] is that the second item in Proposition 19 does not rule out the possibility to have non-periodic oscillatory trajectories that approach the equilibrium set lying outside it, and approaching it rotating infinitely many times; this can happen in the case of a connected compact equilibrium set.

Remark 2. Recall that, by Proposition 7, if there exists a vertex $v \in V$ such that there is a path from I to v and no path from v to X, then there exists no equilibrium. By the first item of Proposition 19, this further implies that all trajectories are unbounded.

The following results concern the existence and uniqueness of equilibria.

Proposition 20. *([92, Theorems 4 and 5]) For special LIFE systems under Assumption (B), the following holds.*

- *There exists an equilibrium with positive entries for arbitrary constant intakes if and only if for all $v \in V$ there is a path to X such that all edges in the path have $\lim_{x_v \to \infty} H_e(x_v) = +\infty$;*

- *If there exists an equilibrium with positive entries, and if there exists a path from all $v \in V$ to X, then the equilibrium is unique.*

Notice that in the first item, the 'only if' part is true only when we require existence of equilibria with positive entries for completely arbitrary intakes: arbitrary intake set I and arbitrary positive values of the corresponding fluxes; this strong condition is not necessary to have an equilibrium with positive entries for a given set I and a given value of the intake fluxes. Also notice that this statement considers fixed fluxes

f_e, differently from Proposition 10, where x is fixed and fluxes f_e are allowed to vary (except for intake fluxes); the crucial difference is that the product $H_e(x)f_e$ can be made arbitrarily large in the context of Proposition 10, even in the case where $H_e(x)$ is bounded.

A remark about the second item is that, by Proposition 19, the existence and uniqueness of the equilibrium with positive entries further implies that all trajectories converge to such equilibrium.

To apply the first item of Proposition 19 or the second item of Proposition 20, one needs to already have the knowledge about existence of an equilibrium with positive entries. This happens for example in the case where one starts by fixing a desired equilibrium with positive entries $\bar{x}$, and then applies the extreme pathways technique in order to design suitable fluxes ensuring that $\bar{x}$ is an equilibrium of the system. Under some assumptions on the graph, the above propositions then ensure uniqueness of the equilibrium, and its global asymptotic stability.

Special LIFE systems under Assumption (C)

In the remainder of this section, we consider Assumption (C). In this case, recall that the dynamics can be re-written as $\dot{x} = J(f)h(x) + \phi$. Different from linear systems, $h(x)$ can contain non-linearities, but the matrix $J(f)$ has the same definition as for linear systems, so that its spectral properties are described by Propositions 14 and 16. Also notice that equilibria, i.e. solutions of $J(f)h(x) = -\phi$, can be found by solving $J(f)h = -\phi$, where h is an unknown vector in $(\mathbb{R}_+)^n$, and then solving $h(x) = h$. The latter is equivalent to solving $H_{v_i}(x_{v_i}) = h_i$, for $i = 1, \ldots, n$.

We start by studying the case with no intakes nor excretions. In this case, by Proposition 14, $J(f)h = 0$ means $h \in \mathcal{N}(J(f))$, where the nullspace $\mathcal{N}(J(f))$ is generated by $\pi_1, \ldots, \pi_k$, associated with terminal components $G_1, \ldots, G_k$. Without loss of generality, we re-label vertices so that the first n_1 vertices encompass the first terminal

component G_1, the following n_2 vertices encompass the second terminal component G_2, and so on, up to the last n_k vertices encompass the last terminal component G_k, and finally the remaining vertices are not in any terminal component (say there are n_0 of them). Denote the corresponding subblocks of vector h as $h^{(1)}, \ldots, h^{(k)}$ and $h^{(0)}$ for the non-terminal ones, and similarly define $x^{(1)}, \ldots, x^{(k)}$ and $x^{(0)}$ for vector x. By Proposition 14, the nullspace of $J(f)$ is $\mathcal{N}(J(f)) = \{h \in (\mathbb{R}_+)^n : h^{(j)} = \alpha_j \tilde{\pi}_j, \alpha_j \in \mathbb{R}$ and $h^{(0)} = 0$, for $j \in \{1, \ldots, k\}\}$, where $\tilde{\pi}_i$ is the restriction of π to the component G_i, i.e. is a vector of size n_i with strictly positive entries, representing the stationary distribution of the Markov chain whose generator is the transpose of the restriction of $J(f)$ to G_i. Now we need to characterize the set of equilibria with positive entries $\bar{X} := \{x \in (\mathbb{R}_+)^n$ such that $h(x) \in \mathcal{N}(J(f))\}$. Recall that $H_v(x_v)$ are strictly increasing functions, being 0 when $x_v = 0$. Denote by R_v the range of H_v (for $x_v \geq 0$), notice that either $R_v = [0, h^{\max})$, or $R_v = [0, +\infty)$. Now denote by $\mathcal{H}_j$ the set of vectors $\alpha_j \tilde{\pi}_j$ such that $\alpha_j \geq 0$ and $[\alpha_j \tilde{\pi}_j]_v \in R_v$ for all v in G_j. Then denote by $H_j^{-1} h^{(j)}$ the vector obtained from $h^{(j)} \in \mathcal{H}_j$ by applying entry-wise the inverse functions H_v^{-1}. Finally we obtain $\bar{X} = \{x \in (\mathbb{R}_+)^n$ such that $x^{(j)} = H_j^{-1} h^{(j)}, h^{(j)} \in \mathcal{H}_j$ and $h^{(0)} = 0\}$.

Having characterized the set of equilibria with positive entries, now recall that Proposition 18 applies, and trajectories remain bounded, with total mass constant in time, and approach the above-described equilibrium set. We focus on the case with intakes and/or excretions. The definition of the reduced system and the partitioning in two subsystems introduced for linear systems applies also to special LIFE systems under Assumption (C), with the only difference that now the dynamics are non-linear:

$$\begin{cases} \dot{x}_1 = J_1 h_1(x_1) + \phi_1 \\ \dot{x}_2 = J_2 h_2(x_2) + J_{21} h_1(x_1) + \phi_2. \end{cases}$$

vectors $h_1(x_1)$ and $h_2(x_2)$ having replaced x_1 and x_2 in (2.13).

Hence, one can obtain the following analogous of Proposition 17.

Proposition 21. *Consider a weakly connected special LIFE system satisfying Assumption (C). Denote by R_v the range of the function $H_v(x_v)$ (for $x_v \geq 0$) and define $\bar{h} = -J_1^{-1}\phi_1$. If $\bar{h}_v \in R_v$ for all $v \in V_1$, then from any initial condition, the reduced system converges to its unique equilibrium $\bar{x}_1$ defined by $[\bar{x}_1]_v = H_v^{-1}(\bar{h}_v)$. Otherwise, the system has no equilibrium, and $\lim_{t\to\infty} \|x(t)\| = +\infty$.*

In the case where $x_1(t)$ converges to $\bar{x}_1$, if there are some vertices not connected to X, then:

- *If there exists a terminal component G_T with no excretions such that there is a path from I to G_T, then the mass in G_T grows unbounded, and hence also $\lim_{t\to\infty} \|x_2(t)\| = +\infty$ and $\lim_{t\to\infty} \|x(t)\| = +\infty$;*

- *If for all terminal components with no excretions $G_1, \ldots, G_k$ there is no path from I to G_i, then the mass of the system remains bounded, and moreover entries of $x_2(t)$ corresponding to non-terminal components converge to zero, while the restriction of $\bar{x}_2$ to a terminal component G_i approaches the equilibrium set constructed as follows. The restriction of J_2 to vertices in G_i has a nullspace generated by a single positive vector $\tilde{\pi}_i$; let H_i denote the subset of such nullspace given by vectors $h = \alpha\tilde{\pi}_i$ such that $h_v \in R_v$ for all vertices v of G_i; the equilibrium set is given by x's such that there exist $h \in H_i$ verifying $x_v = H_v^{-1}(h_i)$ for all vertices v of G_i.*

Proof. Recall that x_1 is an equilibrium for $\dot{x}_1 = J_1 h_1(x_1) + \phi_1$ if and only if $J_1 h_1(x_1) = -\phi_1$. By Proposition 16, J_1 is invertible and $-J_1^{-1}$ has positive entries; define $\bar{h}$ to be the unique solution to $J_1 h_1 = -\phi_1$. Notice that $\bar{h}$ has all positive entries. If all entries of $\bar{h}$ are within the range of the corresponding function $H_v(x_v)$, then we have a unique equilibrium with positive entries $\bar{x}_1$ obtained as in the statement of the

proposition, and hence by Proposition 19 $x_1(t)$ converges to $\bar{x}$. Otherwise, there exists no equilibrium, and hence, by Proposition 19, the mass of system grows indefinitely.

If there is a terminal component with no excretions but connected to some intakes, then by Proposition 17 there is no equilibrium, and hence by Proposition 19 the mass of system grows indefinitely. If all terminal components with no excretions are not connected to intakes, then the equilibrium set is obtained from the properties of the nullspace of J_2, given by Proposition 14. Moreover, Proposition 19 ensures that trajectories remain bounded and approach the equilibrium set. □

2.4.3 Zero-deficiency theory

In this section, we shortly recall the zero-deficiency theory, and compare it with our results on equilibria of LIFE systems. Zero-deficiency theory arises in the literature on chemical reaction networks, a seminal paper is [45]. Our short overview is based on [31].

A free closed chemical reaction network with m reactions between p complexes involving n species can be described by $\dot{x} = S\Gamma R(x)$ where $S \in M_{n\times p}$, $\Gamma \in M_{p\times m}$ is the incidence matrix of the network, and $R(x)$ a column vector of size m. The *deficiency* of the chemical reaction network is the difference of the dimensions of the nullspaces of $S\Gamma$ and Γ: $\delta = \dim\ker(S\Gamma) - \dim\ker(\Gamma)$, which is equivalent to $\delta = \text{rank}(\Gamma) - \text{rank}(S\Gamma)$ and to $\delta = p - \ell - \text{rank}(S\Gamma)$, where ℓ is the number of weakly connected components. The first equivalence is due to rank-nullity Theorem, and the second to the fact that $\text{rank}(\Gamma) = p - \ell$, since Γ is an incidence matrix (Proposition 4.3 of [12]). The reaction rate vector $R(x)$ is governed by the 'mass-action kinetics': $R(x) = K\Psi(x)$, where $K \in M_{m\times p}$ and $\Psi(x)$ is a vector of size p, defined as follows: $K_{ej} = k_e$ if complex j is the reactant complex of reaction e, $K_{ij} = 0$ otherwise; $\Psi_i(x) = \prod_{j=1}^{n} x_j^{s_{ji}}$. Hence, the dynamics can be equivalently re-written as $\dot{x} = S\Gamma K\Psi(x)$. The following results hold:

Proposition 22 (Zero-deficiency Theorem)**.** *Consider a free closed chemical reaction network with mass-action kinetics. If its deficiency is zero, then: there exists an equilibrium with strictly positive entries if and only if the system is weakly reversible (i.e. each weakly connected component is also strongly connected).*

Moreover, this strictly positive equilibrium is unique in each stoichiometric class (i.e. each weakly connected component has a space of equilibria which has dimension one, so that its equilibrium is unique up to a multiplicative constant representing the total mass in the component), and it is locally asymptotically stable.

Notice the close resemblance with the results for linear LIFE systems in the case with no intakes nor excretions, described in Propositions 14 and 15. It turns out that a particular class of closed free chemical reaction networks with mass-action kinetics and zero deficiency exactly coincides with linear LIFE systems with no intakes nor excretions. Indeed, let $p = n$, let S be the identity matrix of size n, let K be defined by $K_{ej} = f_e$ if edge e is of the form $e = (v_j, w)$ and $K_{ej} = 0$ otherwise; moreover choose exponents s_{ji} to be zero and ones so that $\Psi(x) = x$. This gives exactly a linear LIFE system with no intakes nor excretions. It is immediate to see that the deficiency is zero, since $S = I$ implies $\text{rank}(S\Gamma) = \text{rank}(\Gamma)$.

In the more general case, free closed chemical reaction networks with mass-action kinetics and zero deficiency are a class of LIFE systems with different assumptions than those considered in this paper, e.g., they might not even satisfy Assumption (A), due to the presence of a rectangular matrix S left-multiplying Γ, and often do not satisfy Assumption (B), due to the dependence of entries of $\Psi(x)$ on various entries of x.

2.5 Conclusion

For general LIFE systems (Assumption (A)), we show the existence of positive solutions. Stability criteria have been shown for the structure of a graph associated to LIFE systems. We show that further conclusions cannot be drawn in the general case for LIFE systems because arbitrary dynamics may be defined.

For the problem of understanding equilibria for a fix set of metabolites, we show that the rank of the stoichiometric matrix (and the nullspace of this matrix) are determined by structural properties of the graph associated to the system, discuss the effect of the associated graph having intake vertices and excretion vertices on the rank of the stoichiometric matrix, and give necessary conditions on the structure of the graph associated to LIFE systems for the existence of equilibria. More biologically relevant equilibria of LIFE systems are those with all positive metabolites and fluxes and we prove that the extreme pathways method for finding a positive basis describes all such equilibria.

The field of network flows contributes to LIFE systems by way of the min cut max flow theorem. We show the capacity of edges in network flows relates to saturation of functions corresponding to edges of LIFE systems. The method of extreme pathways to calculate a positive basis is proven to include the entire intersection of the nullspace with the positive orthant. This basis is essential to describing equilibria of LIFE systems.

Equilibria and asymptotic behavior of metabolites for fixed fluxes are studied. We show that under stricter assumptions (Assumption (C)) we determine the eigenvalues of the jacobian of the LIFE system as well as structure of the graph which admit certain equilibria. We analyze the case with intakes and excretions versus the cases without intakes or excretions.

Equilibria for LIFE systems with terminal components exist, but these equilibria are dependent upon initial mass of the system. Furthermore, conditions are given

for LIFE systems to tend to equilibrium from non-negative initial conditions. LIFE systems under Assumption (B) have bounded solutions if and only if there exits a non-negative equilibrium. Then for Assumption (B), we give conditions on the structure of the associated graph for which nontrivial equilibria exist.

We show that LIFE systems with no intakes or excretions have "zero deficiency." The rank of the stoichiometric matrix (defined in this work) gives information about the structure of the associated graph, specifically the connectivity of the graph and the existence of strongly connected components. Zero deficiency theory also provides information about the existence of equilibria with respect to this structure.

The structural conditions of graphs discussed in this work have implications on metabolic networks. We have included results about networks which are considered non-biological for the sake of completeness. These results allow a clear picture of the structure of metabolic systems which are capable of admitting a biologically relevant equilibrium. With this clear picture, we are able to determine metabolic networks for which it is most advantageous to analyze using Linear-In-Flux-Expression.

Appendix

This Appendix contains the calculations for examples 2 and3. Example 2 illustrates finding the extreme pathways of RCT using the algorithm described in Section 2.3.2. Example 3 compares a standard basis and positive basis for RCT and shows that when only the positive orthant is considered they describe the same space.

Example 2 continued. As shown previously, the stoichiometric matrix for the RCT

is given by:

$$S(x) = \begin{pmatrix} 1 & 0 & 0 & -x1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -x2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -x3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & x1 & x2 & x3 & -x4 & -x4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & x4 & 0 & -x5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & x4 & x5 & -x6 \end{pmatrix}$$

First let V_0 be the $m \times m$ identity matrix where m is the number of columns in S. Then define C_0 as follows, $V_0 \cdot S^T = C_0$. To better illustrate the row combinations in the algorithm a column is appended to C_0 which labels the rows.

We have

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \cdot S^T = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & R1 \\ 0 & 1 & 0 & 0 & 0 & 0 & R2 \\ 0 & 0 & 1 & 0 & 0 & 0 & R3 \\ -x_1 & 0 & 0 & x_1 & 0 & 0 & R4 \\ 0 & -x_2 & 0 & x_2 & 0 & 0 & R5 \\ 0 & 0 & -x_3 & x_3 & 0 & 0 & R6 \\ 0 & 0 & 0 & -x_4 & x_4 & 0 & R7 \\ 0 & 0 & 0 & -x_4 & 0 & x_4 & R8 \\ 0 & 0 & 0 & 0 & -x_5 & x_5 & R9 \\ 0 & 0 & 0 & 0 & 0 & -x_6 & R10 \end{pmatrix}$$

The next step is to identify any columns of C_0 that have no sources or sinks. For this example there are two columns, column 4 and column 5. Then we modify both V_0 and C_0 by first copying all the rows with a zero in the entry of the first identified column (column 4), then of the remaining rows add all possible positive combinations of two rows which produce a zero in this column. Note that V_0 is 10×10 and C_0 is 10×6.

This gives us

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{x_4}{x_1} & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{x_4}{x_1} & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{x_4}{x_2} & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{x_4}{x_2} & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{x_4}{x_3} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{x_4}{x_3} & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \cdot S^T = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & R1 \\ 0 & 1 & 0 & 0 & 0 & 0 & R2 \\ 0 & 0 & 1 & 0 & 0 & 0 & R3 \\ -x_4 & 0 & 0 & 0 & x_4 & 0 & \frac{x_4}{x_1}R4 + R7 \\ -x_4 & 0 & 0 & 0 & 0 & x_4 & \frac{x_4}{x_1}R4 + R8 \\ 0 & -x_4 & 0 & 0 & x_4 & 0 & \frac{x_4}{x_2}R5 + R7 \\ 0 & -x_4 & 0 & 0 & 0 & x_4 & \frac{x_4}{x_2}R5 + R8 \\ 0 & 0 & -x_4 & 0 & x_4 & 0 & \frac{x_4}{x_3}R6 + R7 \\ 0 & 0 & -x_4 & 0 & 0 & x_4 & \frac{x_4}{x_3}R6 + R8 \\ 0 & 0 & 0 & 0 & -x_5 & x_5 & R9 \\ 0 & 0 & 0 & 0 & 0 & -x_6 & R10 \end{pmatrix}$$

Note that V_1 is 11×10 and C_1 is 11×6.

Now doing the same for column 5 gives

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{x_4}{x_1} & 0 & 0 & 1 & 0 & \frac{x_4}{x_5} & 0 \\ 0 & 0 & 0 & \frac{x_4}{x_1} & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{x_4}{x_2} & 0 & 1 & 0 & \frac{x_4}{x_5} & 0 \\ 0 & 0 & 0 & 0 & \frac{x_4}{x_2} & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{x_4}{x_3} & 1 & 0 & \frac{x_4}{x_5} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{x_4}{x_3} & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \cdot S^T = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & R1 \\ 0 & 1 & 0 & 0 & 0 & 0 & R2 \\ 0 & 0 & 1 & 0 & 0 & 0 & R3 \\ -x_4 & 0 & 0 & 0 & 0 & x_4 & \frac{x_4}{x_1}R4 + R7 + \frac{x_4}{x_5}R9 \\ -x_4 & 0 & 0 & 0 & 0 & x_4 & \frac{x_4}{x_1}R4 + R8 \\ 0 & -x_4 & 0 & 0 & 0 & x_4 & \frac{x_4}{x_2}R5 + R7 + \frac{x_4}{x_5}R9 \\ 0 & -x_4 & 0 & 0 & 0 & x_4 & \frac{x_4}{x_2}R5 + R8 \\ 0 & 0 & -x_4 & 0 & 0 & x_4 & \frac{x_4}{x_3}R6 + R7 + \frac{x_4}{x_5}R9 \\ 0 & 0 & -x_4 & 0 & 0 & x_4 & \frac{x_4}{x_3}R6 + R8 \\ 0 & 0 & 0 & 0 & 0 & -x_6 & R10 \end{pmatrix}$$

Note that V_2 is 10×10 and C_2 is 10×6.

The same process is then followed for columns containing sources and sinks. Starting at the rightmost column add all possible positive combinations of two rows that create a 0 entry in this column. Do the same operations to both V_2 and C_2. Doing this for just the rightmost column we have

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{x_4}{x_1} & 0 & 0 & 1 & 0 & \frac{x_4}{x_5} & \frac{x_4}{x_6} \\ 0 & 0 & 0 & \frac{x_4}{x_1} & 0 & 0 & 0 & 1 & 0 & \frac{x_4}{x_6} \\ 0 & 0 & 0 & 0 & \frac{x_4}{x_2} & 0 & 1 & 0 & \frac{x_4}{x_5} & \frac{x_4}{x_6} \\ 0 & 0 & 0 & 0 & \frac{x_4}{x_2} & 0 & 0 & 1 & 0 & \frac{x_4}{x_6} \\ 0 & 0 & 0 & 0 & 0 & \frac{x_4}{x_3} & 1 & 0 & \frac{x_4}{x_5} & \frac{x_4}{x_6} \\ 0 & 0 & 0 & 0 & 0 & \frac{x_4}{x_3} & 0 & 1 & 0 & \frac{x_4}{x_6} \end{pmatrix} \cdot S^T = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & R1 \\ 0 & 1 & 0 & 0 & 0 & 0 & R2 \\ 0 & 0 & 1 & 0 & 0 & 0 & R3 \\ -x_4 & 0 & 0 & 0 & 0 & 0 & \frac{x_4}{x_1}R4 + R7 + \frac{x_4}{x_5}R9 + \frac{x_4}{x_6}R10 \\ -x_4 & 0 & 0 & 0 & 0 & 0 & \frac{x_4}{x_1}R4 + R8 + \frac{x_4}{x_6}R10 \\ 0 & -x_4 & 0 & 0 & 0 & 0 & \frac{x_4}{x_2}R5 + R7 + \frac{x_4}{x_5}R9 + \frac{x_4}{x_6}R10 \\ 0 & -x_4 & 0 & 0 & 0 & 0 & \frac{x_4}{x_2}R5 + R8 + \frac{x_4}{x_6}R10 \\ 0 & 0 & -x_4 & 0 & 0 & 0 & \frac{x_4}{x_3}R6 + R7 + \frac{x_4}{x_5}R9 + \frac{x_4}{x_6}R10 \\ 0 & 0 & -x_4 & 0 & 0 & 0 & \frac{x_4}{x_3}R6 + R8 + \frac{x_4}{x_6}R10 \end{pmatrix}$$

V_3 is 9×10 and C_3 is 9×6. After the third column we have the following,

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{x_4}{x_1} & 0 & 0 & 1 & 0 & \frac{x_4}{x_5} & \frac{x_4}{x_6} \\ 0 & 0 & 0 & \frac{x_4}{x_1} & 0 & 0 & 0 & 1 & 0 & \frac{x_4}{x_6} \\ 0 & 0 & 0 & 0 & \frac{x_4}{x_2} & 0 & 1 & 0 & \frac{x_4}{x_5} & \frac{x_4}{x_6} \\ 0 & 0 & 0 & 0 & \frac{x_4}{x_2} & 0 & 0 & 1 & 0 & \frac{x_4}{x_6} \\ 0 & 0 & x_4 & 0 & 0 & \frac{x_4}{x_3} & 1 & 0 & \frac{x_4}{x_5} & \frac{x_4}{x_6} \\ 0 & 0 & x_4 & 0 & 0 & \frac{x_4}{x_3} & 0 & 1 & 0 & \frac{x_4}{x_6} \end{pmatrix} \cdot S^T = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & R1 \\ 0 & 1 & 0 & 0 & 0 & 0 & R2 \\ -x_4 & 0 & 0 & 0 & 0 & 0 & \frac{x_4}{x_1}R4 + R7 + \frac{x_4}{x_5}R9 + \frac{x_4}{x_6}R10 \\ -x_4 & 0 & 0 & 0 & 0 & 0 & \frac{x_4}{x_1}R4 + R8 + \frac{x_4}{x_6}R10 \\ 0 & -x_4 & 0 & 0 & 0 & 0 & \frac{x_4}{x_2}R5 + R7 + \frac{x_4}{x_5}R9 + \frac{x_4}{x_6}R10 \\ 0 & -x_4 & 0 & 0 & 0 & 0 & \frac{x_4}{x_2}R5 + R8 + \frac{x_4}{x_6}R10 \\ 0 & 0 & 0 & 0 & 0 & 0 & x_4R3 + \frac{x_4}{x_3}R6 + R7 + \frac{x_4}{x_5}R9 + \frac{x_4}{x_6}R10 \\ 0 & 0 & 0 & 0 & 0 & 0 & x_4R3 + \frac{x_4}{x_3}R6 + R8 + \frac{x_4}{x_6}R10 \end{pmatrix}$$

V_4 is 8×10 and C_4 is 8×6. After the second column we have the following,

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{x_4}{x_1} & 0 & 0 & 1 & 0 & \frac{x_4}{x_5} & \frac{x_4}{x_6} \\ 0 & 0 & 0 & \frac{x_4}{x_1} & 0 & 0 & 0 & 1 & 0 & \frac{x_4}{x_6} \\ 0 & x_4 & 0 & 0 & \frac{x_4}{x_2} & 0 & 1 & 0 & \frac{x_4}{x_5} & \frac{x_4}{x_6} \\ 0 & x_4 & 0 & 0 & \frac{x_4}{x_2} & 0 & 0 & 1 & 0 & \frac{x_4}{x_6} \\ 0 & 0 & x_4 & 0 & 0 & \frac{x_4}{x_3} & 1 & 0 & \frac{x_4}{x_5} & \frac{x_4}{x_6} \\ 0 & 0 & x_4 & 0 & 0 & \frac{x_4}{x_3} & 0 & 1 & 0 & \frac{x_4}{x_6} \end{pmatrix} \cdot S^T = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & R1 \\ -x_4 & 0 & 0 & 0 & 0 & 0 & \frac{x_4}{x_1}R4 + R7 + \frac{x_4}{x_5}R9 + \frac{x_4}{x_6}R10 \\ -x_4 & 0 & 0 & 0 & 0 & 0 & \frac{x_4}{x_1}R4 + R8 + \frac{x_4}{x_6}R10 \\ 0 & 0 & 0 & 0 & 0 & 0 & x_4R2 + \frac{x_4}{x_2}R5 + R7 + \frac{x_4}{x_5}R9 + \frac{x_4}{x_6}R10 \\ 0 & 0 & 0 & 0 & 0 & 0 & x_4R2 + \frac{x_4}{x_2}R5 + R8 + \frac{x_4}{x_6}R10 \\ 0 & 0 & 0 & 0 & 0 & 0 & x_4R3 + \frac{x_4}{x_3}R6 + R7 + \frac{x_4}{x_5}R9 + \frac{x_4}{x_6}R10 \\ 0 & 0 & 0 & 0 & 0 & 0 & x_4R3 + \frac{x_4}{x_3}R6 + R8 + \frac{x_4}{x_6}R10 \end{pmatrix}$$

V_5 is 7×10 and C_5 is 7×6. And finally,

$$\begin{pmatrix} x_4 & 0 & 0 & \frac{x_4}{x_1} & 0 & 0 & 1 & 0 & \frac{x_4}{x_5} & \frac{x_4}{x_6} \\ x_4 & 0 & 0 & \frac{x_4}{x_1} & 0 & 0 & 0 & 1 & 0 & \frac{x_4}{x_6} \\ 0 & x_4 & 0 & 0 & \frac{x_4}{x_2} & 0 & 1 & 0 & \frac{x_4}{x_5} & \frac{x_4}{x_6} \\ 0 & x_4 & 0 & 0 & \frac{x_4}{x_2} & 0 & 0 & 1 & 0 & \frac{x_4}{x_6} \\ 0 & 0 & x_4 & 0 & 0 & \frac{x_4}{x_3} & 1 & 0 & \frac{x_4}{x_5} & \frac{x_4}{x_6} \\ 0 & 0 & x_4 & 0 & 0 & \frac{x_4}{x_3} & 0 & 1 & 0 & \frac{x_4}{x_6} \end{pmatrix} \cdot S^T = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & x_4R1 + \frac{x_4}{x_1}R4 + R7 + \frac{x_4}{x_5}R9 + \frac{x_4}{x_6}R10 \\ 0 & 0 & 0 & 0 & 0 & 0 & x_4R1 + \frac{x_4}{x_1}R4 + R8 + \frac{x_4}{x_6}R10 \\ 0 & 0 & 0 & 0 & 0 & 0 & x_4R2 + \frac{x_4}{x_2}R5 + R7 + \frac{x_4}{x_5}R9 + \frac{x_4}{x_6}R10 \\ 0 & 0 & 0 & 0 & 0 & 0 & x_4R2 + \frac{x_4}{x_2}R5 + R8 + \frac{x_4}{x_6}R10 \\ 0 & 0 & 0 & 0 & 0 & 0 & x_4R3 + \frac{x_4}{x_3}R6 + R7 + \frac{x_4}{x_5}R9 + \frac{x_4}{x_6}R10 \\ 0 & 0 & 0 & 0 & 0 & 0 & x_4R3 + \frac{x_4}{x_3}R6 + R8 + \frac{x_4}{x_6}R10 \end{pmatrix}$$

V_6 is 6×10 and C_6 is 6×6. The rows of V_6 are the basis for $\mathcal{N}(S(x))$. While the $\dim\mathcal{N}(S(x)) = 4$, there are 6 extreme pathway vectors. These vectors are positive, and positively linearly independent,

$$V_6 = \begin{pmatrix} x_4 & 0 & 0 & \frac{x_4}{x_1} & 0 & 0 & 1 & 0 & \frac{x_4}{x_5} & \frac{x_4}{x_6} \\ x_4 & 0 & 0 & \frac{x_4}{x_1} & 0 & 0 & 0 & 1 & 0 & \frac{x_4}{x_6} \\ 0 & x_4 & 0 & 0 & \frac{x_4}{x_2} & 0 & 1 & 0 & \frac{x_4}{x_5} & \frac{x_4}{x_6} \\ 0 & x_4 & 0 & 0 & \frac{x_4}{x_2} & 0 & 0 & 1 & 0 & \frac{x_4}{x_6} \\ 0 & 0 & x_4 & 0 & 0 & \frac{x_4}{x_3} & 1 & 0 & \frac{x_4}{x_5} & \frac{x_4}{x_6} \\ 0 & 0 & x_4 & 0 & 0 & \frac{x_4}{x_3} & 0 & 1 & 0 & \frac{x_4}{x_6} \end{pmatrix}.$$

Example 3 continued. The positive basis of S was found in example 2. This basis has six vectors despite the four dimensional nullspace. Also note that the basis vectors are not linearly independent, but positively linearly independent. The span

of the positive basis vectors is shown below,

$$b_1\begin{pmatrix} x_6 \\ 0 \\ 0 \\ \frac{x_6}{x_1} \\ 0 \\ 0 \\ 0 \\ \frac{x_6}{x_4} \\ 0 \\ 1 \end{pmatrix} + b_2\begin{pmatrix} x_6 \\ 0 \\ 0 \\ \frac{x_6}{x_1} \\ 0 \\ 0 \\ \frac{x_6}{x_4} \\ 0 \\ \frac{x_6}{x_5} \\ 1 \end{pmatrix} + b_3\begin{pmatrix} 0 \\ x_6 \\ 0 \\ 0 \\ \frac{x_6}{x_2} \\ 0 \\ 0 \\ \frac{x_6}{x_4} \\ 0 \\ 1 \end{pmatrix} + b_4\begin{pmatrix} 0 \\ x_6 \\ 0 \\ 0 \\ \frac{x_6}{x_2} \\ 0 \\ \frac{x_6}{x_4} \\ 0 \\ \frac{x_6}{x_5} \\ 1 \end{pmatrix} + b_5\begin{pmatrix} 0 \\ 0 \\ x_6 \\ 0 \\ 0 \\ \frac{x_6}{x_3} \\ 0 \\ \frac{x_6}{x_4} \\ 0 \\ 1 \end{pmatrix} + b_6\begin{pmatrix} 0 \\ 0 \\ x_6 \\ 0 \\ 0 \\ \frac{x_6}{x_3} \\ \frac{x_6}{x_4} \\ 0 \\ \frac{x_6}{x_5} \\ 1 \end{pmatrix}. \tag{2.16}$$

Next the positive basis vectors for the nullspace must be intersected with the positive orthant $\mathcal{N}(S(x)) \cap (\mathbb{R}_+)^m$. When this span is intersected with the positive orthant it is clear there is only one condition,

$$\text{For all i, } b_i \geq 0 \qquad b_i \in \mathbb{R}. \tag{2.17}$$

We refer to this span intersected with the positive orthant as B.

For the stoichiometric matrix $S(x)$ the basis for the null space was found. The

span of the four basis vectors is shown below,

$$a_1 \begin{pmatrix} x_6 \\ 0 \\ 0 \\ \frac{x_6}{x_1} \\ 0 \\ 0 \\ 0 \\ \frac{x_6}{x_4} \\ 0 \\ 1 \end{pmatrix} + a_2 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \frac{x_5}{x_4} \\ -\frac{x_5}{x_4} \\ 1 \\ 0 \end{pmatrix} + a_3 \begin{pmatrix} -x_3 \\ 0 \\ x_3 \\ -\frac{x_3}{x_1} \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + a_4 \begin{pmatrix} -x_2 \\ x_2 \\ 0 \\ -\frac{x_2}{x_1} \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{2.18}$$

To find the intersection of this span with the positive orthant, three conditions on the a_i must hold.

$$\begin{cases} a_i \geq 0, & \text{for } i \in \{1,2,3,4\} \\ a_1 x_6 \geq a_3 x_3 + a_4 x_2, & \\ a_1 x_6 \geq a_2 x_5. & \end{cases} \tag{2.19}$$

We refer to the intersection of this span with the positive orthant as C. We show that under the conditions given, $B = C$.

First it is shown that $B \subset C$ by showing that for an arbitrary set of $b_i \geq 0$, a_i's can be chosen to reach the same vector. Using the following substitutions for a_i it is

clear that if $b_i \geq 0$ the inequalities of (2.19) are satisfied

$$\begin{cases} a_1 = b_1 + b_2 + b_3 + b_4 + b_5 + b_6, \\ a_2 = (b_2 + b_4 + b_6)\frac{x_6}{x_5}, \\ a_3 = (b_5 + b_6)\frac{x_6}{x_3}, \\ a_4 = (b_3 + b_4)\frac{x_6}{x_2}. \end{cases} \tag{2.20}$$

This shows that any vector in B can be represented by vectors in C i.e. $B \subset C$.

Now it is shown that $C \subset B$. For arbitrary a's which satisfy (2.19), the following substitutions for b are used and the conditions of (2.17) are checked.

$$\begin{cases} b_1 = a_1 - a_3\frac{x_3}{x_6} - a_4\frac{x_2}{x_6} - a_2\frac{x_5}{x_6} + b_4 + b_6, \\ b_2 = a_2\frac{x_5}{x_6} - b_4 - b_6, \\ b_3 = a_4\frac{x_2}{x_6} - b_4, \\ b_5 = a_3\frac{x_3}{x_6} - b_6. \end{cases} \tag{2.21}$$

To insure that (2.17) is satisfied the following inequalities must hold,

$$a_1 + b_4 + b_6 \geq a_3\frac{x_3}{x_6} + a_4\frac{x_2}{x_6} + a_2\frac{x_5}{x_6}, \tag{2.22}$$

$$a_2\frac{x_5}{x_6} \geq b_4 + b_6, \tag{2.23}$$

$$a_4\frac{x_2}{x_6} \geq b_4, \tag{2.24}$$

$$a_3\frac{x_3}{x_6} \geq b_6. \tag{2.25}$$

If choices for b_4 and b_6 can be found which satisfy these inequalities then we will have $C \subset B$.

We have the following two conditions:

Condition 1. $a_4\frac{x_2}{x_6} + a_3\frac{x_3}{x_6} \leq a_2\frac{x_5}{x_6}$. Setting $b_4 = a_4\frac{x_2}{x_6}$ and $b_6 = a_3\frac{x_3}{x_6}$ immediately satisfies (2.23), (2.24), and (2.25). From (2.19) we have $a_1x_6 \geq a_2x_5$ which means that inequality (2.22) is satisfied.

Condition 2. $a_4\frac{x_2}{x_6} + a_3\frac{x_3}{x_6} > a_2\frac{x_5}{x_6}$.

Under Condition 2 three cases must be considered.

Case 1. $a_4\frac{x_2}{x_6} < a_2\frac{x_5}{x_6}$. In this case we set $b_4 = a_4\frac{x_2}{x_6}$ and $b_6 = a_2\frac{x_5}{x_6} - a_4\frac{x_2}{x_6}$. This immediately satisfies (2.23) and (2.24). Since $a_4\frac{x_2}{x_6} + a_3\frac{x_3}{x_6} > a_2\frac{x_5}{x_6}$ (2.25) is satisfied as well. Then from (2.19) we have $a_1x_6 \geq a_3x_3 + a_4x_2$, which means that (2.22) is satisfied.

Case 2. $a_4\frac{x_2}{x_6} > a_2\frac{x_5}{x_6}$ and $a_3\frac{x_3}{x_6} < a_2\frac{x_5}{x_6}$. In this case we set $b_6 = a_3\frac{x_3}{x_6}$ and $b_4 = a_2\frac{x_5}{x_6} - a_3\frac{x_3}{x_6}$, this satisfies (2.23), (2.24), (2.25). And from (2.19) we have $a_1x_6 \geq a_3x_3 + a_4x_2$, which means that (2.22) is also satisfied.

Case 3. $a_4\frac{x_2}{x_6} > a_2\frac{x_5}{x_6}$ and $a_3\frac{x_3}{x_6} > a_2\frac{x_5}{x_6}$. In this case we set $b_4 = b_6 = \frac{1}{2}a_2\frac{x_5}{x_6}$ which satisfies (2.23), (2.24), (2.25). Then from (2.19) we have $a_1x_6 \geq a_3x_3 + a_4x_2$, which means that (2.22) is also satisfied.

Since the $a's$ were arbitrary and $b's$ are found which satisfy (2.17) this gives us that $C \subset B$ as desired.

Chapter 3

Equilibria and control of metabolic networks with enhancers and inhibitors

3.1 Introduction

Models in Quantitative Systems Pharmacology (briefly QSP) [3, 108, 121, 129] have been used by the pharmaceutical industry in order to discover new drugs at less cost. In modeling metabolic networks, bio-chemical reactions are organized into a graph, with nodes representing metabolites and edges representing fluxes [96, 99, 107]. The dynamics is described by a *Stoichiometric matrix* encoding the kinetics of the biochemical reactions.

QSP models are commonly used with the assumption that all fluxes are independent or have insignificant correlation [2, 64, 126], which does not recognize the resilience of metabolic networks. Thus, despite the vast knowledge by Systems Biology (as well as other fields [107, 12, 25, 45, 65, 77, 92, 102]), researchers were unable to accurately simulate large metabolic networks [48, 129] as an inexpensive alternative for clinical

trials. A new method called *Linear-in-Flux-Expression* (briefly LIFE), where one rewrites the system as linear with respect to the fluxes [96], has shown potential to help pharmacology simulators recover crucial characteristics of metabolic networks. This approach has been investigated [99] and currently works for mild nonlinearities, however inhibitors, enhancers, and fully nonlinear dynamics are still not included.

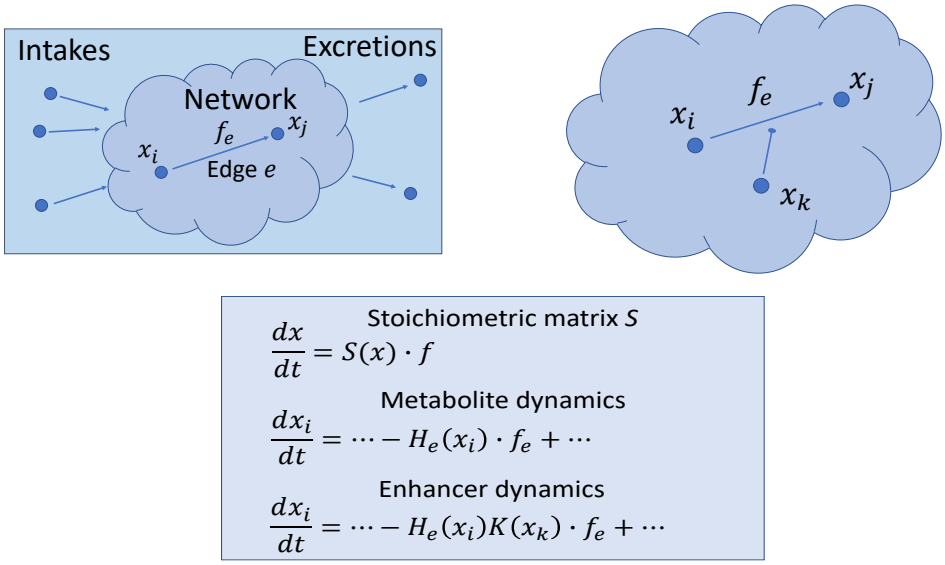


Figure 3.1: Top Left: network with intakes and excretions from and to the external environment. Nodes x_is are composed of metabolite concentrations, fluxes f_e represent biochemical reactions. Top Right: Enhancer x_k affects flux f_e. Bottom: network dynamics with Stoichiometric matrix $S(x)$ (first equation). Dynamics of a single metabolite x_i is affected by flux f_e via term H_e (second equation). Additional terms such as $K(x_k)$ are added to include inhibitors and enhancers (third equation).

The aim of this work is to apply results from previous studies to actual networks and extend our results to include control of the intakes and enhancer dynamics. Figure 3.1 depicts the main ideas. Metabolic networks naturally have intakes and excretions, while fluxes f_e are determined by kinetics of bio-chemical reactions. The LIFE approach consists of writing dynamics with Stoichiometric matrix $S(x)$ that describes the mass consumed and produced by the reactions. This matrix is dependent on the metabolites (as opposed to classical $S(f)$ dependent on fluxes) thus allowing nonlinear dynamics in metabolites. A general condition on network topology (connection

of every node to excretion, see [99]) guarantees existence and uniqueness of an equilibrium $\bar{x}_f$ for every flux vector $f = \{f_e\}$. Thus network dynamics is captured by the map $f \to \bar{x}_f$ with a multivalued inverse $x \to F(x)$ (see [92, Theorem 4]).

We first show various theories applied to a small yet significant part of human cholesterol metabolism, called "Reverse Cholesterol Transport" (briefly RCT). In particular, nonlinear dynamics implemented on RCT admits a unique stable equilibrium (for fixed fluxes and intakes). This allows to study drug discovery by control methods. More precisely, we consider both control of intakes (corresponding to a diet and/or supplements) and the introduction of enhancers and inhibitors (corresponding to drugs). Also, control of inhibitors/enhancers is compared versus control of the intakes.

The control of intakes is entirely similar to the problem of controlling inputs to compartmental systems. The main difference here is in the assumptions on the specific network dynamics. In particular, choosing Michaelis-Menten kinetic, which gives a nonlinear dynamics in metabolites, one has to face the problem of saturation. The latter may lead to non existence of equilibria. More precisely, since the kinetic can support only a maximal level of discharge for a given metabolite, too high intakes may cause the metabolite levels to increase indefinitely. By analyzing the map from intakes to equilibria, we are able to compute the set of admissible intakes (compatible with saturation levels) and thus reduce optimal control problems to finite-dimensional optimization ones.

The action of inhibitors and enhancers is defined for general networks: Inhibitors and enhancers augment other edges via multiplicative Michaelis-Menten kinetics. Existence, uniqueness and stability of equilibria is proved for the RCT example and conjectures for general networks under suitable assumptions. The latter are entirely similar to the case without enhancers and inhibitors, thus are expected to be satisfied by most natural metabolic networks. Lastly, we describe a process of drug discov-

ery by f augmenting the network with enhancers or inhibitors and finding optimal controls.

The paper is organized as follows. Section 3.2 introduces the LIFE methods and main definitions. Section 3.3 describes the Reverse Cholesterol Transport network and provide various results by applying well-established theories. Section 3.4 considers control problems for intakes, while Section 3.5 introduces the dynamics with enhancer and inhibitors. Section 3.6 illustrated drug discovery by mean of enhancer and inhibitors, and, finally, Section 3.7 contains conclusions.

3.2 Assumptions for Linear-in-flux-Expression

In this paper, our analysis is organized into sections with different assumptions on the entries S_{ve} of the stoichiometric matrix as function of the values of metabolites and fluxes. The most general class of systems we consider satisfies the following assumption. For $x \in (\mathbb{R}_+)^n$, it holds:

$$(A) \qquad S_{ve}(x) = \begin{cases} H_e(x) > 0 & \text{if } (e = (w, v),\ w \in V,\ x_w > 0) \text{ or } (e = (v_0, v),\ v \in I) \\ -H_e(x) < 0 & \text{if } x_v > 0 \text{ and } (e = (v, w), w \in V) \text{ or } (e = (v, v_{n+1}), v \in X)\,, \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

where $H_e : \mathbb{R}^n \to \mathbb{R}_+$ is continuous. Notice that Assumption (A) implies that, for each $v \in V$,

$$\sum_{v \in V} S_{ve}(x) = \begin{cases} H_e(x) & e = (v_0, \bar{v}),\ \bar{v} \in I, \\ -H_e(x) & e = (\bar{v}, v_{n+1}),\ \bar{v} \in X, \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$

All columns of S have zero sum, except those corresponding to intakes and excretions, which have positive and negative sum, respectively. Under Assumption (A), the dy-

namics (1.3) can be interpreted as mass conservation law. Here we introduce a special class of systems of type (1.3) with simplified dynamics. We consider Assumption (A), and we impose further restriction on the functions $H_e(x)$: For an edge $e = (v, w)$, we assume H_e to depend on x_v only, and moreover we impose the scalar function $H_e(x_v)$ to be strictly increasing:

$$(B) \quad S_{ve}(x) = \begin{cases} = -H_e(x_v) & e = (v, w),\ v \in V, w \in V \cup \{v_{n+1}\} \\ = H_e(x_w) & e = (w, v),\ w \in V \\ = 1 & e = (v_0, v)\ v \in I \\ = 0 & \text{otherwise,} \end{cases} \tag{3.3}$$

where $H_e : \mathbb{R} \to \mathbb{R}_+$ is differentiable, strictly increasing, and $H_e(0) = 0$.

A typical example of a system verifying (B) is given by metabolic networks with Hill functions representing reactions, i.e., $H_e(x_v) = \frac{x_v^{p_e}}{K + x_v^{p_e}}$ with $p_e \in \mathbb{N}$, and K is called the dissociation constant. LIFE approach allows a simple description of flux vector f which permits the system to be in equilibrium; such vectors comprise the null space of $S(x)$. In previous work [99], the authors showed how the theory of Laplacian dynamics, Markov chains, network flows, and compartmental systems apply to LIFE systems. The results tell us how the structure of the graph of the metabolic network affects existence, uniqueness, and stability of equilibria. An example of structure the authors investigated in [99] is a *terminal component* of a network, which is a part of the network not connected to excretions. The assumptions on $S(x)$ and the structure of the network allow also to compute equilibria related to a given flux vector f.

3.3 Dynamics theories at work on Reverse Cholesterol Transport

In this section, results from [99, 97] are applied to the example network of RCT [96]. RCT (see Figure 3.2) represents the mechanism of removal of cholesterol from plaques in arteries. Here, we apply theory from [20] to derive equilibria and stability of the RCT. The various mathematical approaches are applicable to this example as follows. Linear systems without intakes nor excretions are related to *continuous-time Markov chains* [27]. Linear systems with intakes and excretions from and to the external environment are known as *compartmental systems* [20, 65]. Nonlinear systems are studied using the results from [92].

The dynamics of the RCT network can be written as (1.3), with $S : \mathbb{R} \to M_{6\times 10}$. The matrix S is then the concatenation of S_{1-3} and S_{4-10} which are the submatrices given by the first three columns and last seven columns respectively.

$$S_{1-3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

and S_{4-10} is given by:

$$\begin{pmatrix} -H_{f_{e_4}}(x_{v_1}) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -H_{f_{e_5}}(x_{v_2}) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -H_{f_{e_6}}(x_{v_3}) & 0 & 0 & 0 & 0 \\ H_{f_{e_4}}(x_{v_1}) & H_{f_{e_5}}(x_{v_2}) & H_{f_{e_6}}(x_{v_3}) & -H_{f_{e_7}}(x_{v_4}) & -H_{f_{e_8}}(x_{v_4}) & 0 & 0 \\ 0 & 0 & 0 & H_{f_{e_7}}(x_{v_4}) & 0 & -H_{f_{e_9}}(x_{v_5}) & 0 \\ 0 & 0 & 0 & 0 & H_{f_{e_8}}(x_{v_4}) & H_{f_{e_9}}(x_{v_5}) & -H_{f_{e_{10}}}(x_{v_6}) \end{pmatrix},$$

while the flux vector $f \in \mathbb{R}^{10}$ is a 10×1 vector $(f_i)_{i\in\{1,\dots,10\}}$.

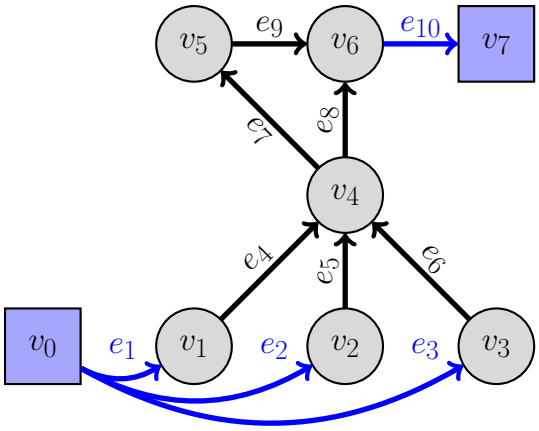


Figure 3.2: Reverse Cholesterol Transport Network. This network contains six vertices which represent metabolites, ten edges which represent fluxes and two virtual vertices v_0, v_7. There are three intake vertices v_1, v_2, v_3 and one excretion vertex v_6.

A further simplification is where $H_e(x_v)$ is the same function $H_v(x_v)$ for all edges e having v as an initial vertex. This gives Assumption (C), as follows. For all $x \in (\mathbb{R}_+)^n$,

$$\text{(C)} \quad S_{ve}(x) = \begin{cases} = -H_v(x_v) & e = (v, w),\ v \in V, w \in V \cup \{v_{n+1}\} \\ = H_w(x_w) & e = (w, v),\ w \in V \\ = 1 & e = (v_0, v)\ v \in I \\ = 0 & \text{otherwise,} \end{cases}$$

and each $H_v : \mathbb{R}_+ \to \mathbb{R}_+$, with $H_v(0) = 0$ is a differentiable and strictly increasing function. The system can be equivalently re-written as:

$$\dot{x} = J(f)h(x) + \phi, \tag{3.4}$$

where $h(x) = \begin{pmatrix} h_1(x_1), & \cdots, & h_n(x_n) \end{pmatrix}^T$, $J(f) \in M_{n\times n}$ (a $n \times n$ matrix with real entries), and $\phi \in \mathbb{R}^n$. The intake vector ϕ is written as a $n \times 1$ vector with a zero if there is no intake flux from the outside environment and ϕ_{e_i} the value of the flux if

there is an edge to v_i. For the RCT network, $J(f)$ and ϕ are given by:

$$J(f) = \begin{pmatrix} -f_{e_4} & 0 & 0 & 0 & 0 & 0 \\ 0 & -f_{e_5} & 0 & 0 & 0 & 0 \\ 0 & 0 & -f_{e_6} & 0 & 0 & 0 \\ f_{e_4} & f_{e_5} & f_{e_6} & -f_{e_7} - f_{e_8} & 0 & 0 \\ 0 & 0 & 0 & f_{e_7} & -f_{e_9} & 0 \\ 0 & 0 & 0 & f_{e_8} & f_{e_9} & -f_{e_{10}} \end{pmatrix}, \phi = \begin{pmatrix} f_{e_1} \\ f_{e_2} \\ f_{e_3} \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{3.5}$$

The following Proposition 23 (from [99]) shows that the existence of nontrivial equilibria implies some network structure. We define the sets I, X, as the set of vertices attached to v_0, v_7 respectively.

Proposition 23. *Consider system* (1.3) *satisfying (A). Assume there exists an equilibrium $\bar{x} \in (\mathbb{R}_+)^n$ for a flux vector f such that $f_e > 0$ for every $e \in E$. Then for every vertex $v \in V$ for which there exists a path from I to v, there exists a path from v to X.*

The max-flow min-cut theorem [47] implies that a positive flux vector exists if there is a path from v_0 and v_{n+1} as shown by the following proposition.

Proposition 24. *Consider system* (1.3) *satisfying (A), fix an $x \in (\mathbb{R}_+)^n$ and intake flow vector $\bar{\phi}$ with strictly positive entries. Then there exists $f \in (\mathbb{R}_+)^m$ in the null space of $S(x)$ if and only if for each $v \in I$ there exists a path to X.*

3.3.1 Linear RCT without intakes and excretions

Assume the RCT network dynamics is linear, then the flow of mass along an edge is proportional to the mass of the metabolite represented by the initial vertex. This means that Assumption (C) is satisfied with $h(x)$ from (3.4) such that $h(x) = x$. When the network is isolated from the external environment, the dynamics can be

written as $\dot{x} = \bar{J}(f)x$, where $\bar{J}(f)$ is a *Metzler matrix* (i.e., has non-negative off-diagonal entries), and all its columns sum to zero ($\mathbf{1}^T\bar{J}(f) = \mathbf{0}^T$).

Let G' be a graph associated to RCT network without intakes (remove two virtual vertices v_0, v_7 and remove blue edges e_1, e_2, e_3, e_{10} in Figure 3.2). We assume each edge is associated with a strictly positive weight $f_e > 0$. In the theory of *Laplacian dynamics* [102], the matrix $L = -\bar{J}(f)^T$ can be interpreted as a weighted Laplacian of G'. The weighted adjacency matrix A is composed of elements $A_{ij} = f_{(v_i,v_j)}$ if $(v_i, v_j) \in E$, and $A_{ij} = 0$ if there no directed edge from v_j to v_i. We defined a diagonal matrix D that contains the row-sum of A indicating the weights leaving from each vertex. The weighted Laplacian is given by $L = D - A$.

In the field of *consensus dynamics* [20, Theorem 6] systems are modeled using the weighted Laplacian $\dot{x} = -Lx$. The dynamics is different from the linear LIFE dynamics, but $\bar{J}(f)$ and $-L$ have the same eigenvalues. The study of the spectrum of the Laplacian has been investigated in the fields of graph theory and control [99]. The results of these fields apply to LIFE systems when we consider the metabolic network with all intakes and excretions removed. Moreover, the RCT network without intakes and excretions is related to a continuous-time Markov chain over a finite state $V = \{v_1, \ldots, v_6\}$. Thus results on Markov chains are applicable to LIFE systems. The spectrum of $\bar{J}(f)$ and the asymptotic behavior of the dynamics are described by the following proposition:

Proposition 25. *Assume there are no intakes nor excretions, then:*
(1) All eigenvalues of $\bar{J}(f)$ *are either* 0 *or have strictly negative real part.*
(2) The dimension of the nullspace of $\bar{J}(f)$ *is equal to the algebraic multiplicity of the* 0 *eigenvalue, and is equal to the number of terminal components in the graph.*
(3) From any positive initial condition, the system converges to an equilibrium, having strictly positive entries in correspondence of vertices of terminal components, and 0 *elsewhere.*

(4) The equilibrium is determined by the initial mass if there is a unique terminal component.

The RCT network without intakes and excretions has a unique terminal component made up of the singular vertex v_6. Satisfying (1) and (2) of Proposition 25, the eigenvalues of $\bar{J}(f)$ are: $\{0, -f_{e_4}, -f_{e_5}, -f_{e_6}, -f_{e_9}, -(f_{e_7} + f_{e_8})\}$. Furthermore (3) and (4) imply that all the initial mass in the RCT will converge to v_6.

3.3.2 Linear RCT with intakes and excretions

In the case the linear RCT network interact with the environment via intakes and/or excretions, the dynamics is given by $\dot{x} = J(f)x + \phi$, where $J(f)$ and ϕ is defined in (3.5). The associated graph G is given by Figure 3.2. A directed graph is called weakly connected if there exists an undirected path between each pair of vertices. A weakly connected component of a directed graph is defined as a maximal weakly connected subgraph. We notice that G is a weakly connected component including v_0 and v_7. The rows of $J(f)$ no longer sum to 0 because of a single element in the last column representing excretion. Thus $-J(f)^T$ does not satisfy the definition of weighted Laplacian. In [142] a term *grounded Laplacian* L_g is introduced. L_g can be constructed by deleting the row and column corresponding to a given set of vertices (called *grounded vertices*) from L. For RCT network, $-J(f)^T$ is a grounded Laplacian L_g (see [99]). The spectrum of $J(f)$ is described by:

Proposition 26. *Consider a linear system with intakes and/or excretions, then the following are equivalent:*

(a) For every $v \in V$ there is a path from v to X.

(b) $J(f)$ is Hurwitz stable (i.e., all its eigenvalues have strictly negative real part).

(c) $J(f)$ is invertible.

Moreover, when $J(f)$ is invertible, all entries of $-J(f)^{-1}$ are positive; if the graph is strongly connected, all entries of $-J(f)^{-1}$ are strictly positive.

It is clear from Figure 3.2 that the RCT network with intake and excretion satisfies (a). The Jacobian $J(f)$ can be seen in equation 3.5, and the eigenvalues are: $\{-f_{e_4}$, $-f_{e_5}, -f_{e_6}, -f_{e_9}, -f_{e_{10}}, -(f_{e_7} + f_{e_8})\}$, verifying (b).

3.3.3 Non-linear RCT

The non-linear RCT network satisfies Assumption (B) with $h(x)$ may be non-linear, e.g., $h(x) = \frac{Hx}{K_M+x}$ is Michaelis-Menten type of equation, where H is the saturation value, and K_M is the Michaelis constant.

Proposition 27. *([92, Theorem 6]) Consider the non-linear system under Assumption (B) with no intakes and excretions. The following properties hold:*
(1) The total mass of the system $m = \sum_{v\in V} x_v$ is constant in time.
(2) From any positive initial condition $x(0)$, the system tends to the equilibrium set.
(3) Moreover, if there is a unique terminal component, then there exists a unique equilibrium with positive entries with the same mass as the initial mass, and the system converges to it.

This means that despite the $h(x)$ functions being non-linear, the RCT network with no intakes and excretions will have constant total mass and the system tends to the equilibrium set. For the general case with intakes and excretions, we recall results from [92]:

Proposition 28. *([92, Theorems 2 and 3]) For a system with positive initial condition:*
(1) Trajectories are bounded if and only if there exists an equilibrium with positive entries.
(2) If trajectories are bounded, then they approach an equilibrium set for $t \to \infty$. If moreover the equilibrium set consists of isolated points, then they converge to some equilibrium.

The following results concern the existence and uniqueness of equilibria.

Proposition 29. *([92, Theorems 4 and 5]) Under Assumption (B) the following holds,*
(1) There exists an equilibrium with positive entries for arbitrary constant intakes if and only if for all $v \in V$ there is a path to X such that all edges in the path have $\lim_{x_v \to \infty} H_e(x_v) = +\infty$.
(2) If there exists an equilibrium with positive entries, and all $v \in V$ connect to X, then the equilibrium is unique.

For the RCT network, for every vertex $v \in I$ there is a path from v to X, therefore by Proposition 29 there is a unique equilibrium x_{eq} and by Proposition 28 we will approach the equilibrium as $t \to \infty$. Under the stricter Assumption (C) this equilibrium can be calculated analytically using equation 3.4 with $h(x)$ being given by:

$$\left(\frac{f_{e_1}}{f_{e_4}}, \quad \frac{f_{e_2}}{f_{e_5}}, \quad \frac{f_{e_3}}{f_{e_6}}, \quad \frac{f_{e_1}+f_{e_2}+f_{e_3}}{f_{e_7}+f_{e_8}}, \quad \frac{f_{e_7}(f_{e_1}+f_{e_2}+f_{e_3})}{f_{e_9}(f_{e_7}+f_{e_8})}, \quad \frac{f_{e_1}+f_{e_2}+f_{e_3}}{f_{e_{10}}} \right)^T. \tag{3.6}$$

The equilibrium x_{eq} can then be calculated by computing the inverse of each $h_i(x_i)$ function.

3.4 Control of intakes

In this section, we focus on the special LIFE system under Assumption (C). Typically f is fixed and x cannot be controlled directly, which leaves the intake vector ϕ as a natural choice for control targets. We are interested in intakes ϕ that lead to an equilibrium. In this section the inequality $x > 0$, respectively $x \geq 0$, with $x \in \mathbb{R}^n$, is to be interpreted as $x_i > 0$, respectively $x_i \geq 0$, for all $i = 1, \ldots, n$.

Proposition 30. *Consider system* (1.3) *satisfying Assumption (C), then the set* Φ *of intake vector* ϕ *giving rise to a positive equilibrium* $x_\phi > 0$ *is given by:*

$$\Phi = \{\phi | x_\phi := h^{-1}(-J(f)^{-1}\phi) > 0\}. \tag{3.7}$$

Proof. The dynamics of systems under Assumption (C) can be rewritten as (3.4). Thus the equilibrium of the system can be found by solving $J(f)h(x_\phi) + \phi = 0$. □

The control action is then captured by the map:

$$\phi \longrightarrow x_\phi = h^{-1}(-J(f)^{-1}\phi). \tag{3.8}$$

However, to obtain admissible equilibria, we need to take into account the saturation effect of nonlinear kinetics such as Michaelis-Menten:

Proposition 31. *Consider system* (1.3) *with Assumption (C) holding true and the function* $h(x) = (h_i(x_i))_{i\in\{1,\dots,n\}}$ *in* (3.4) *with* $h_i(x_i) : \mathbb{R}_+ \to \mathbb{R}_+$ *being strictly increasing and bounded by saturation values* $H = (H_i)_{i\in\{1,\dots,n\}}$,

$$h_i(x_i) \leq H_i, \text{ for all } i = \{1, \dots, n\}. \tag{3.9}$$

Then for the system to tend to an equilibrium the intake vector ϕ *must satisfy*

$$\phi = -J(f)h(x_\phi) \leq -J(f)H, \text{ for } \phi \in \Phi \text{ and } x_\phi \in \mathbb{R}_+. \tag{3.10}$$

Proof. Under Assumption (C) the system can be written as (3.4). From (3.4) and inequality (3.9), solving for ϕ we get (3.10). □

We consider the following control problem:

Definition 6. For a given system (1.3) satisfying Assumption (C), and a given desired equilibrium $x_{eq} \in (\mathbb{R}_+)^n$, find the optimal intake vector ϕ that drives the system as close as possible to x_{eq}.

The solution is provided by the intake vector ϕ which minimizes the Euclidean distance

$$||x_\phi - x_{eq}|| = \sqrt{\sum_{i=1}^{n}((x_{eq})_i - (x_\phi)_i)^2}. \tag{3.11}$$

An alternative control problem is to minimize the distance of the kinetic vector h instead of the metabolites:

Definition 7. For a given system (1.3) satisfying Assumption (C), and a given desired equilibrium $x_{eq} \in (\mathbb{R}_+)^n$, find the optimal intake vector ϕ that drives the system kinetic vector $h(x_\phi)$ as close as possible to $h(x_{eq})$.

The solution is given by the intake vector ϕ which minimizes the Euclidean distance

$$||h(x_\phi) - h(x_{eq})|| = \sqrt{\sum_{i=1}^{n}((h(x_{eq}))_i - (h(x_\phi))_i)^2}. \tag{3.12}$$

Since $h(x_{eq}) = -J(f)^{-1}\phi$ is linear with respect to ϕ, for the second problem we can use linear solvers to find the optimized ϕ. Therefore, the second control problem can be solved much more efficiently. Notice that, in general, the two control problems are not equivalent, since the solution minimizing (3.11) may be different than that minimizing (3.12). However, we do expect the solution to the control problem of Definition 7 to provide an ansatz toward the solution to the control problem of Definition 7. Indeed, even if the level sets of the two functions (3.11) and (3.12) are different, they are both convex and centered at the same point.

Let us now focus on the RCT network under Assumption (C) and explore the map (3.8): $h(x)$ is a 6×1 vector $h(x) = h_i(x_i)_{i\in\{1,\dots,6\}}$. We choose kinetics of Michaelis-Menten type with saturation value H_i ([66]), $h_i(x_i) = \frac{H_i x_i}{K_M + x_i}$,where K_M

is the Michaelis constant. Our intake vector ϕ is a 6×1 vector with last three components vanishing. Figure 3.4 illustrated the bounds on the kinetic and the intakes.

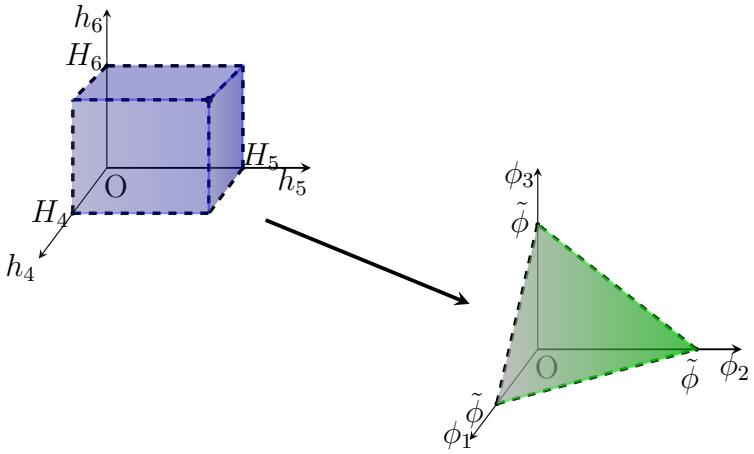


Figure 3.3: In blue the 3-dimensional cube, in the space (h_4, h_5, h_6) of kinetics of last three metabolites, satisfying the bounds given by the saturation values H_4, H_5 and H_6. Based on the conditions (3.13) and (3.14), the intake vector ϕ must belong to the tetrahedron bounded by the coordinate planes and the standard simplex (in green).

The matrices $\phi, J(f), h$, and H can be written in block form, then (3.10) becomes

$$\begin{pmatrix} \phi_e \\ \hline 0 \end{pmatrix} = \begin{pmatrix} \phi_{e_1} \\ \phi_{e_2} \\ \phi_{e_3} \\ \hline 0 \\ 0 \\ 0 \end{pmatrix} = \left(\begin{array}{ccc|ccc} f_{e_4} & 0 & 0 & 0 & 0 & 0 \\ 0 & f_{e_5} & 0 & 0 & 0 & 0 \\ 0 & 0 & f_{e_6} & 0 & 0 & 0 \\ \hline -f_{e_4} & -f_{e_5} & -f_{e_6} & f_{e_7}+f_{e_8} & 0 & 0 \\ 0 & 0 & 0 & -f_{e_7} & f_{e_9} & 0 \\ 0 & 0 & 0 & -f_{e_8} & -f_{e_9} & f_{e_{10}} \end{array}\right) \begin{pmatrix} h_1 \\ h_2 \\ h_3 \\ \hline h_4 \\ h_5 \\ h_6 \end{pmatrix}$$

$$= \left(\begin{array}{c|c} D & 0 \\ \hline J_3 & J_4 \end{array}\right) \begin{pmatrix} h_e \\ \hline h_d \end{pmatrix} \leq \left(\begin{array}{c|c} D & 0 \\ \hline J_3 & J_4 \end{array}\right) \begin{pmatrix} H_e \\ \hline H_d \end{pmatrix}.$$

So we have

$$\begin{cases} \phi_e &= Dh_e, \\ 0 &= J_3 h_e + J_4 h_d. \end{cases}$$

We define $\tilde{\phi} = \phi_{e_1} + \phi_{e_2} + \phi_{e_3}$ as the total mass flows in the system. Then

$$H_d \geq h_d = -J_4^{-1}(J_3 D^{-1}\phi_e) = \begin{pmatrix} \frac{1}{f_{e_7}+f_{e_8}} & 0 & 0 \\ \frac{f_{e_7}}{f_{e_9}(f_{e_7}+f_{e_8})} & \frac{1}{f_{e_9}} & 0 \\ \frac{1}{f_{e_{10}}} & \frac{1}{f_{e_{10}}} & \frac{1}{f_{e_{10}}} \end{pmatrix} \begin{pmatrix} \tilde{\phi} \\ 0 \\ 0 \end{pmatrix}.$$

Observing the structure of Figure 3.2 and that the total outflow of a vertex must equal the total inflow for an equilibrium state, we have

$$\phi_{e_i} \leq f_{e_{i+3}} H_i, i = 1, 2, 3, \tag{3.13}$$

$$\tilde{\phi} \leq min\left\{ H_4(f_{e_7}+f_{e_8}), H_5 \frac{f_{e_9}(f_{e_7}+f_{e_8})}{f_{e_7}}, H_6 f_{e_{10}} \right\}. \tag{3.14}$$

Having determined the bounds on ϕ for existence of an equilibrium, we explore the map (3.8) on the subset describe by the bounds.

Some parameters in the RCT network are chosen *a priori*, such as the fluxes f, the given functions $h(x)$, and the corresponding saturation values H. As a proof of concept, we randomly sample f and the initial values for x from the uniform distribution with range $(0, 1)$, randomly sample H_i from the uniform distribution with range $(0, 10)$, and randomly sample the desired equilibrium x_{eq} from uniform distribution $(0, 3)$. We use these randomly generated data to test the effectiveness of the optimization algorithm to drive the state to $x_{eq} = ((x_{eq})_i)_{i\in\{1,\ldots,6\}}$. x_{v_4} corresponds to the *High-density lipoprotein (HDL)*, x_{v_5} corresponds to the *Very low-density lipoprotein (VLDL)*, and x_{v_6} corresponds to the *low-density lipoprotein (LDL)*. According to [59], a high ratio indicates a higher risk of heart attack. A healthy state is considered to have lower ratio of $\frac{HDL+LDL}{LDL}$. The second control problem (Definition 7) can be solved using the following optimization algorithm:

Algorithm 2. *For assigned fluxes and kinetics:*

Step 1 : Calculate the Jacobian matrix J in (3.4) *and the matrix* $-J(f)^{-1}$.

Step 2 : Implement linear least-square method to find ϕ_{opti} *minimizing* (3.12) *with inequality constraint* $-J(f)^{-1}\phi \leq H$ *and lower bound* 0.

Step 3 : Calculate the cost and run simulation with optimized intakes ϕ_{opti} *to verify numerically the stability of dynamics.*

Notice that not all combinations of sampled parameters lead to equilibrium: Here we show an example leading with stability. To implement Algorithm 2, we use a linear least-square solver in MATLAB (2018b) for Step 2, finding the optimized ϕ_{opti} by minimizing the cost $\frac{1}{2}||h(x_\phi) - h(x_{eq}||_2^2 = \frac{1}{2}|| - J(f)^{-1}\phi - h(x_{eq}||_2^2$, which is half of the square of (3.12). The flux vector is given by:

$$f = \begin{pmatrix} 0.71, & 0.36, & 0.40, & 0.32, & 0.23, & 0.67, & 0.96, & 0.74, & 0.44, & 0.51 \end{pmatrix}^T \tag{3.15}$$

and the initial value of metabolites is:

$$\begin{pmatrix} x_{v_1 0} & x_{v_2 0} & x_{v_3 0} & x_{v_4 0} & x_{v_5 0} & x_{v_6 0} \end{pmatrix}^T = \begin{pmatrix} 0.87, & 0.80, & 0.35, & 0.76, & 0.62, & 0.21 \end{pmatrix}^T. \tag{3.16}$$

The kinetic h functions for all vertices are given by:

$$\begin{pmatrix} \frac{7.21 \cdot x_{v1}}{1+x_{v1}} & \frac{7.57 \cdot x_{v2}}{1+x_{v2}} & \frac{4.26 \cdot x_{v3}}{1+x_{v3}} & \frac{5.32 \cdot x_{v4}}{1+x_{v4}} & \frac{7.98 \cdot x_{v5}}{1+x_{v5}} & \frac{8.56 \cdot x_{v1}}{1+x_{v1}} \end{pmatrix}^T,$$

while the saturation levels for h are:

$$H = \begin{pmatrix} 7.21, & 7.57, & 4.26, & 5.32, & 7.98, & 8.56 \end{pmatrix}^T.$$

We randomly generate the equilibrium x_{eq} to be matched:

$$x_{eq} = \begin{pmatrix} 2.86, & 2.50, & 0.09, & 0.64, & 0.91, & 2.98 \end{pmatrix}^T$$

with

$$h(x_{eq}) = \begin{pmatrix} 5.34, & 5.41, & 0.36, & 2.07, & 3.81, & 6.40 \end{pmatrix}^T.$$

The $h(x)_{opti}$ and $(x)_{opti}$ optimized by Algorithm 2 are:

$$h(x)_{opti} = \begin{pmatrix} 5.34, & 5.40, & 0.35, & 1.89, & 4.12, & 6.26 \end{pmatrix}^T,$$

$$(x)_{opti} = \begin{pmatrix} 2.85, & 2.50, & 0.09, & 0.55, & 1.07, & 2.73 \end{pmatrix}^T.$$

while the optimized intake vector is $\phi_{opti} = \begin{pmatrix} \phi_{e_1}, & \phi_{e_2}, & \phi_{e_3} \end{pmatrix}^T = \begin{pmatrix} 1.73, & 1.24, & 0.24 \end{pmatrix}^T$. In the following, we check the conditions (3.13) and (3.14):

$$\begin{aligned} \phi_{e_1} &= 1.73 < f_{e_4} H_1 = 0.32 \times 7.21 = 2.31, \\ \phi_{e_2} &= 1.24 < f_{e_5} H_2 = 0.23 \times 7.57 = 1.74, \\ \phi_{e_3} &= 0.24 < f_{e_6} H_3 = 0.67 \times 4.26 = 2.85, \\ \tilde{\phi} = \phi_{e_1} + \phi_{e_2} + \phi_{e_3} &= 3.20 < min\{9.02, 6.21, 4.38\} = 4.38. \end{aligned}$$

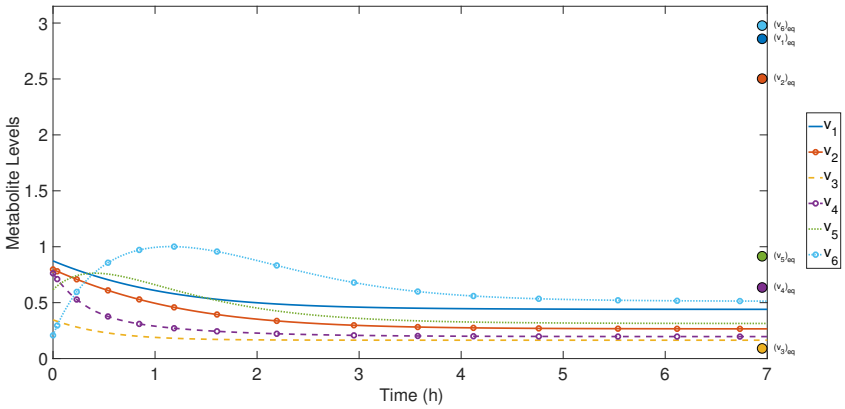


Figure 3.4: The RCT network with fixed fluxes f in (3.15) and initial conditions in (3.16) with no controls. The given equilibrium is $x_{eq} = (2.86, 2.50, 0.09, 0.64, 0.91, 2.98)^T$. The simulated equilibrium is $x_\phi = (0.440.27, 0.16, 0.20, 0.32, 0.52)^T$. The dots representing the given equilibrium levels for six vertices (corresponding to the trajectory with the same color). The cost is 4.1828.

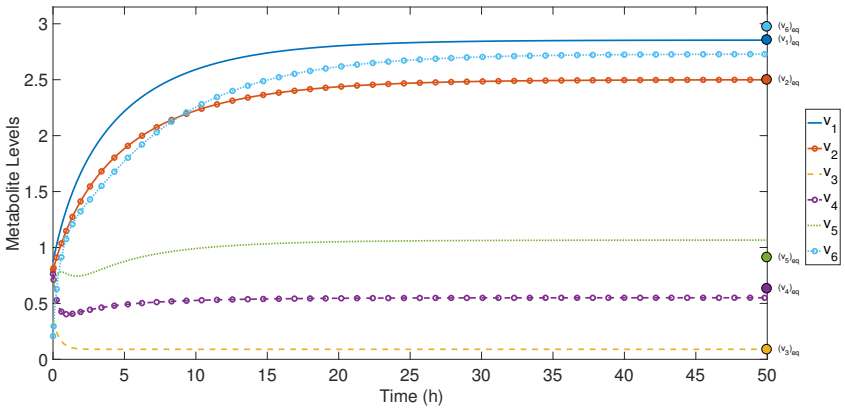


Figure 3.5: The RCT network with optimized intakes value (control with intakes). $(x_{eq})_{opti} = (2.85, 2.50, 0.09, 0.55, 1.07, 2.73)^T$. The cost is 0.3022.

We run the simulation with optimized intakes ϕ_{opti} and keep the other parameters constant. Figures 3.4 and 3.5 show simulations for the not controlled case and the one with optimized intakes. Notice that well approximate the desired equilibrium.

3.5 Metabolic networks with inhibitor and enhancer

In order to include inhibitors and enhancers we introduce a new Assumption (D):

$$\text{(D)}\quad S_{ve}(x) = \begin{cases} = -H_e(x_v)\cdot K_u(x_u) & e=(v,w),\ v\in V, w\in V\cup\{v_{n+1}\}, u\in V \\ = H_e(x_w)\cdot K_u(x_u) & e=(w,v),\ w\in V \\ = 1 & e=(v_0,v)\ v\in I \\ = 0 & \text{otherwise,} \end{cases} \tag{3.17}$$

where $H_v :^n\!\to_+$ is differentiable and strictly increasing while $H_v(0)=0$, $K_u:\mathbb{R}\to[0,+\infty)$ is differentiable, monotonic, and $K_u(0)=1$. Moreover, for enhancers K_u is increasing while for inhibitors K_u is decreasing.

Let us clarify the differences among the assumptions we discuss. Assumption (A) requires the function $H_e(x)$ to be defined for each edge, continuous, and dependent on the entire state x. Assumption (B) requires $H_e(x_v)$ to be defined for each edge, continuous, and dependent on the metabolite associated to the initial vertex of edge e. Assumption (C) requires $H_v(x_v)$ to be defined for the initial vertex of an edge, differentiable and strictly increasing, $H_v(0)=0$, and dependent on the metabolite associated to the initial vertex of an edge. Therefore all edges with the same initial vertex will be associated to the same function $H_v(x_v)$. For Assumption (D) the function $H_v(x_v)$ is defined just as in Assumption (C), however entries of $S_{ve}(x)$ consist of a product $H_v(x_v)\cdot K_u(x_u)$. The factor $K_u(x_u)$ models the action of inhibitors or enhancers. This represents nonlocal action, because the inhibitor or enhancer is not necessarily nearby (in topological sense) the affected metabolites. Assumption (A) is the most general with Assumption (B) implying Assumption (A), and Assumption (C) implying (B), i.e., $C \implies B \implies A$. Assumption (D) also implies Assumption (A), but is comparable with (B), i.e., $D \implies A, D \not\Rightarrow B, B \not\Rightarrow D$.

To illustrate the effect of enhancers and inhibitors on dynamics under Assumption (D), an example is shown on the RCT network of Figure 3.2, but with the addition of a single enhancer. It is assumed that the metabolite v_1 will act as an enhancer for the edge f_{e_7}. When the dynamics of this network are written , the stoichiometric matrix $S : \mathbb{R} \to M_{6\times 10}$ where $S = (S_{1-3}|S_{4-10})$ is different from Section 3.3 only by the non-zero entries of the seventh column, i.e., S_{4-10} is given by

$$\begin{pmatrix} -H_{v_1}(x_{v_1}) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -H_{v_2}(x_{v_2}) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -H_{v_3}(x_{v_3}) & 0 & 0 & 0 & 0 \\ H_{v_1}(x_{v_1}) & H_{v_2}(x_{v_2}) & H_{v_3}(x_{v_3}) & -K_{v_1}(x_{v_1}) \cdot H_{v_4}(x_{v_4}) & -H_{v_4}(x_{v_4}) & 0 & 0 \\ 0 & 0 & 0 & K_{v_1}(x_{v_1}) \cdot H_{v_4}(x_{v_4}) & 0 & -H_{v_5}(x_{v_5}) & 0 \\ 0 & 0 & 0 & 0 & H_{v_4}(x_{v_4}) & H_{v_5}(x_{v_5}) & -H_{v_6}(x_{v_6}) \end{pmatrix}.$$

3.5.1 Equilibria under Assumption (A) and (B)

The previous research focused on finding necessary and sufficient conditions for equilibria of (1.3), as well as the uniqueness and stability of such solutions, under Assumption (A) or (B). We aim at extending the results to the case of Assumption (D) as well. The Propositions 23 and 24 work also for Assumption (A) and as such are still applicable when inhibitors/enhancers are added as in Assumption (D). Proposition 23 gives necessary structural conditions to obtain an equilibrium x_{eq} for fixed flux vector f, specifically that any vertex with a directed path from an intake, also has a directed path to an excretion.

3.5.2 Dynamics under Assumption (D)

Without further assumptions on the system, there are few conclusions about these types of equilibria. In Section 3.3, additional properties summarized in Assumptions (B) and (C) were needed to prove uniqueness and stability for this type of equilibrium (see Propositions 28 and 29). To investigate equilibria for Assumption (D), we

consider system (1.2). For a single enhancer or inhibitor, we may write the dynamics as follows,

$$\dot{x} = J_1(f)h(x) + J_2(f)k(x) + \phi \tag{3.18}$$

where J_1 is the Jacobian matrix removing the terms which are affected by inhibitors or enhancers, $h(x)$ is a column vector of size n given by $h_i(x) = H_{v_i}(x_{v_i})$, J_2 is a matrix with nonzero entries only in the column affected by the enhancer or inhibitor, k is a column vector of size n representing inhibitors and enhancers given by $k_i(x) = K_{v_j}(x_{v_j}) \cdot H_{v_i}(x_{v_i})$ where v_i is the node from which the edge begins and v_j is the node which acts as an inhibitor or enhancer, and ϕ is a vector of size n given by $\phi_i = f_{e(v_0,v_i)}$ if $(v_0, v_i) \in \tilde{E}$ and $\phi_i = 0$ otherwise. Note that multiple enhancers or inhibitors may require additional matrices J_i. To see the possible necessity of added J_i matrices, consider two edges $e(v_i, v_j)$ and $e(v_i, v_k)$ of a network that are enhanced by v_m and v_n respectively. Since both edges begin at v_i they would be in the same column of J_2, however enhancer v_m is represented by $k_i(x) = K_{v_m}(x_{v_m}) \cdot H_{v_i}(x_{v_i})$, and enhancer v_n is represented by $k_i(x) = K_{v_n}(x_{v_n}) \cdot H_{v_i}(x_{v_i})$, because of this an additional J_3 matrix and k_2 would be necessary.

Proposition 32. *Consider system satisfying Assumption (D). For each vertex, if all of the outgoing edges from the vertex are affected by at most one enhancer or inhibitor, then the system can always be written as* (3.18).

When all the edges leaving a vertex v_i are affected by the same enhancer or inhibitor v_m, then $k_i(x) = K_{v_m}(x_{v_m}) \cdot H_{v_i}(x_{v_i})$, this allows the system to be written as (3.18).

Our main conjecture is the following.

Conjecture 1. *For a metabolic network under Assumption (D) containing enhancers and inhibitors, the following holds.*

- *An equilibrium with positive entries for arbitrary intakes can exist only if for all*

$v \in V$ there is a path to X such that all edges in the path have $\lim_{x_v \to \infty} H_e(x_v) = +\infty$.

Under Assumption (D) with enhancers but no inhibitors the following hold.

- *There exists an equilibrium with positive entries for arbitrary constant intakes if and only if for all $v \in V$ there is a path to X such that all edges in the path have* $\lim_{x_v \to \infty} H_e(x_v) = +\infty$.

- *If there exists an equilibrium with positive entries, and there is a path from all $v \in V$ to X, then the equilibrium is unique.*

It is expected that under Assumption (D) there exists an equilibrium if several conditions are verified. Here it is shown that the Reverse Cholesterol Transport network with the single enhancer (indicated previously) has a unique equilibrium. In addition, simulations show that this equilibrium is stable.

3.5.3 Unique equilibrium of the RCT network with a single enhancer

Consider the RCT network with enhancer, then the system can be written as (3.18) where $J_1(f)$, $J_2(f)$, $h(x)$, $k(x)$ and ϕ matrices are given by

$$J_1 = \begin{pmatrix} -f_{e_4} & 0 & 0 & 0 & 0 & 0 \\ 0 & -f_{e_5} & 0 & 0 & 0 & 0 \\ 0 & 0 & -f_{e_6} & 0 & 0 & 0 \\ f_{e_4} & f_{e_5} & f_{e_6} & -f_{e_8} & 0 & 0 \\ 0 & 0 & 0 & 0 & -f_{e_9} & 0 \\ 0 & 0 & 0 & f_{e_8} & f_{e_9} & -f_{e_{10}} \end{pmatrix}, h = \begin{pmatrix} H_{v_1}(x_{v_1}) \\ H_{v_2}(x_{v_2}) \\ H_{v_3}(x_{v_3}) \\ H_{v_4}(x_{v_4}) \\ H_{v_5}(x_{v_5}) \\ H_{v_6}(x_{v_6}) \end{pmatrix},$$

$$J_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -f_{e_7} & 0 & 0 \\ 0 & 0 & 0 & f_{e_7} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, k = \begin{pmatrix} 0 \\ 0 \\ 0 \\ [K_{v_1}(x_1)] \cdot H_{v_4}(x_{v_4}) \\ 0 \\ 0 \end{pmatrix}, \phi = \begin{pmatrix} f_{e_1} \\ f_{e_2} \\ f_{e_3} \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

We assume that the H_{v_i} functions satisfy the additional condition that

$$\lim_{x_v \to \infty} H_e(x_v) = +\infty. \tag{3.19}$$

Condition (3.19) is required so that we are able to satisfy equation (3.10).

In order for the system to be at equilibrium it must satisfy

$$0 = J_1(f)h(x) + J_2(f)k(x) + \phi, \tag{3.20}$$

which provides the following system of six equations,

$$\begin{pmatrix} -f_{e_4} & 0 & 0 & 0 & 0 & 0 \\ 0 & -f_{e_5} & 0 & 0 & 0 & 0 \\ 0 & 0 & -f_{e_6} & 0 & 0 & 0 \\ f_{e_4} & f_{e_5} & f_{e_6} & -f_{e_8} & 0 & 0 \\ 0 & 0 & 0 & 0 & -f_{e_9} & 0 \\ 0 & 0 & 0 & f_{e_8} & f_{e_9} & -f_{e_{10}} \end{pmatrix} \begin{pmatrix} H_{v_1}(x_{v_1}) \\ H_{v_2}(x_{v_2}) \\ H_{v_3}(x_{v_3}) \\ H_{v_4}(x_{v_4}) \\ H_{v_5}(x_{v_5}) \\ H_{v_6}(x_{v_6}) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ -f_{e_7} \cdot (-[K_{v_1}(x_{v_1}) H_{v_4}(x_{v_4})]) \\ f_{e_7} \cdot (-[K_{v_1}(x_{v_1}) H_{v_4}(x_{v_4})]) \\ 0 \end{pmatrix}.$$

Immediately notice that the first equation gives $-f_{v_0} H_{v_1}(x_{v_1}) = -\phi_1$. Since H_{v_1} is invertible and f_{e_4} and f_{e_1} are constants, this gives $x_{v_1} = H_{v_1}^{-1}(\frac{f_{e_1}}{f_{e_4}})$. Once the equilibrium value of x_{v_1} has been found, the equilibrium of $K_{v_1}(x_{v_1})$ is determined. Now let $f_{e_7}^* = f_{e_7} \times K_{v_1}(x_{v_1})$. Now we rewrite the system as $J(f)h(x) + \phi$ with,

$$J(f) = \begin{pmatrix} -f_{e_4} & 0 & 0 & 0 & 0 & 0 \\ 0 & -f_{e_5} & 0 & 0 & 0 & 0 \\ 0 & 0 & -f_{e_6} & 0 & 0 & 0 \\ f_{e_4} & f_{e_5} & f_{e_6} & -f_{e_7}^* - f_{e_8} & 0 & 0 \\ 0 & 0 & 0 & f_{e_7}^* & -f_{e_9} & 0 \\ 0 & 0 & 0 & f_{e_8} & f_{e_9} & -f_{e_{10}} \end{pmatrix}, \phi = \begin{pmatrix} f_{e_1} \\ f_{e_2} \\ f_{e_3} \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} H_{v_1}(x_{v_1}) \\ H_{v_2}(x_{v_2}) \\ H_{v_3}(x_{v_3}) \\ H_{v_4}(x_{v_4}) \\ H_{v_5}(x_{v_5}) \\ H_{v_6}(x_{v_6}) \end{pmatrix}. \quad (3.21)$$

This system satisfies the assumptions of Proposition 29 and we conclude that the RCT network has a unique equilibrium.

To show that the equilibrium solution for the RCT network is stable, simulations were performed with fixed randomized fluxes (sampled from a uniform distribution on interval [0,1]). Using the values of the flows the equilibrium was calculated, and then compared with the metabolite levels reached in simulation. The calculated equilibrium was reached in each simulation, suggesting that the RCT network is asymptotically stable. Figure 3.6 shows the results of one such simulation.

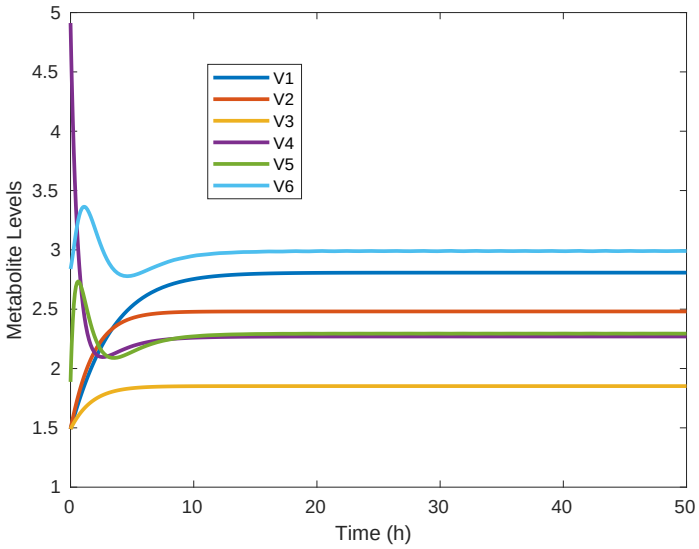


Figure 3.6: The trajectories of the values of metabolites over 50 hours. The simulation results closely matched the calculated equilibrium values of $x_1 = 2.8081$, $x_2 = 2.4811$, $x_3 = 1.8517$, $x_4 = 2.2699$, $x_5 = 2.2943$, $x_6 = 2.9901$.

3.6 Drug discovery by control methods

Existence and uniqueness of equilibria for networks with enhancer and inhibitors (I-E), as explored in Section 3.5, is the basis for studying drug discovery. Usually a drug affects a flux as an enhancer or inhibitor, thus a new drug can be represented as a specific extended network and flux vector f, see Figure 3.7. If networks including the drug treatment effect satisfy the assumptions for uniqueness and stability of equilibria, then we can predict the asymptotic state of the metabolic network, which is the unique equilibrium x_f corresponding to the flux vector f. In other words, one may reduce the drug discovery problem to an optimization one using the map $f \to x_f$ and studying those f corresponding to potential drugs. More precisely we proceed as follows:

Step 1. Consider the extended networks as in Figure 3.7 and study conditions on uniqueness and stability of equilibria using the methods of previous sections. As shown in Section 3.5, Step 1 is not trivial. Indeed, the I-E dynamics are non local,

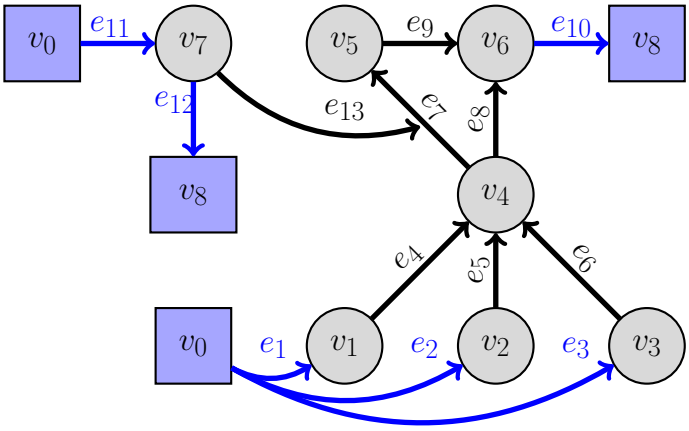


Figure 3.7: Reverse Cholesterol Transport Network with added enhancer or inhibitor metabolite v_7 which acts on edge e_7.

which means that the molecule affecting the specific reaction is not necessarily close in the network topology (i.e., may belong to a far away component of the network). For a given network $G = (E, V)$, we consider the extended version $\tilde{G} = (\tilde{E}, \tilde{V})$ as in Figure 3.7. More precisely, we add a new vertex representing the drug molecule and being both an intake and excretion vertex (assuming the drug is administered and there is a known excretion mechanism). Moreover, we add an edge from the drug molecule to the affected flux edge e. We denote by $f_{e,\pm}$ the flux vector corresponding to the extended network, where $\pm$ indicates the enhancer (+) or inhibitor (-) expected effect. Notice that the source flux of the drug molecule depends on the scheduled treatment, which may be given as pills, injections or others. For drug discovery purposes, we use an average flux, while for simulations we may use a time-varying one.

Step 2. Study the map $f_{e,\pm} \to x_{f_{e,\pm}}$ to minimize $\|\bar{x} - x_{f_{e,\pm}}\|$, where $\bar{x}$ is a desired "healthy" state. Step 1 ensures that the map $f_{e,\pm} \to x_{f_{e,\pm}}$ is well defined. Using the method of previous section, we can also compute explicitly this map under suitable assumptions. The set of states $R(x_{f_{e,\pm}})$ that can be reached using the admissible controls $f_{e,\pm}$ is called the reachable set, see [17]. In practice, we do not expect a unique desired state to be defined, but rather a set of conditions which determine a

desired or target set $\mathcal{X}$ of metabolic states deemed healthy. A successful treatment prescribes the fluxes $f_{e,\pm}$ which will drive the patient's current state x to some state $\bar{x}\in\mathcal{X}$. If the reachable set and desired states are disjoint, i.e., $R(x_{f_{e,\pm}})\cap\mathcal{X}=0$, then we look for the $x_{f_{e,\pm}}$ realizing the minimum distance from $\mathcal{X}$.

Step 3. Introduce dynamic optimization criteria on the trajectory to $x_{f_{e,\pm}}$ to find the best $f_{e,\pm}$. Consider the problem to minimize $\int_0^T L(t,x)dt$ where L measures the toxicity of the drug, see [24, 23]. In this case the dynamic of the drug is very important. It also becomes important to take the dosing regimen into consideration, which means using a time varying flux for the drug intake.

In the remainder of this section we give more details on **1-2**, while **3** is saved for future work. Let us give an explicit example of our approach:

Example 5. Figure 3.7 shows the RCT network with an added vertex v_7 which represents the drug. With the addition of the non-local enhancer or inhibitor the stoichiometric matrix of the RCT needs an additional row and several additional columns, and the flux vector f has the additional components $f_{(v_0,v_7)}$, $f_{(v_7,v_{n+1})}$. The new $S(x)$ matrix is a concatenation given by S_{1-3}, S_{4-10}, and S_{10-12}, is shown below,

$$S_{1-3}=\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\\0&0&0\\0&0&0\\0&0&0\\0&0&0\end{pmatrix},\quad S_{10-12}=\begin{pmatrix}0&0\\0&0\\0&0\\0&0\\0&0\\0&0\\1&-1\end{pmatrix},$$

$$S_{4-10} = \begin{pmatrix} -H(x_{v_1}) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -H(x_{v_2}) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -H(x_{v_3}) & 0 & 0 & 0 & 0 \\ H(x_{v_1}) & H(x_{v_2}) & H(x_{v_3}) & -K_{v_7}(x_{v_7}) \cdot H(x_{v_4}) & -H(x_{v_4}) & 0 & 0 \\ 0 & 0 & 0 & K_{v_7}(x_{v_7}) \cdot H(x_{v_4}) & 0 & -H(x_{v_5}) & 0 \\ 0 & 0 & 0 & 0 & H(x_{v_4}) & H(x_{v_5}) & -H(x_{v_6}) \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

where $k_{v_7} : \mathbb{R} \longrightarrow [0, +\infty)$ is differentiable, monotonic and $k_{v_7}(0) = 1$. For the enhancer the function $k_{v_7}(x_{v_7}) = 1 + \frac{x_{v_7}}{x_{v_7}+1}$ was used and for the inhibitor $k_{v_7}(x_{v_7}) = 1 - \frac{x_{v_7}}{x_{v_7}+1}$. The addition of the drug allows a control on the system based on the amount of drug used. It is assumed that the drug acts similarly to an enhancer in that the mass of the drug will not be added to the network, but the drug will facilitate the flow through the network by inhibiting or enhancing individual edges. Whether to use an inhibitor or an enhancer drug, as well as how much drug to use depends on the initial conditions of the network.

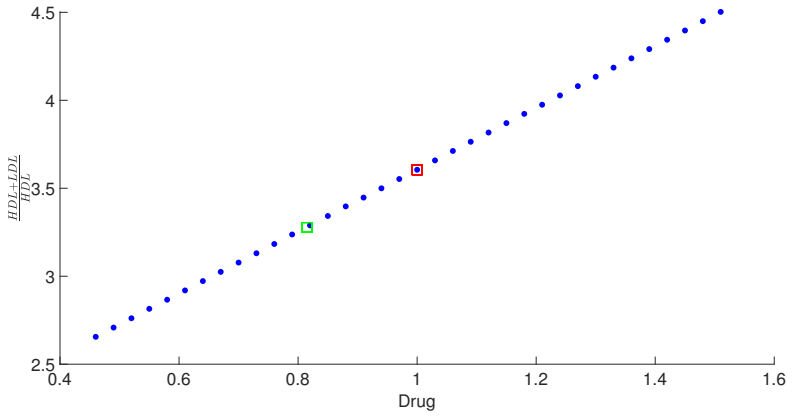


Figure 3.8: Graph showing the control of the system using drug as an inhibitor and enhancer. The x-axis of the graph shows the amount of inhibition or enhancement of edge e_7 from the drug, the value 1 corresponds with no control and is marked in red. The y-axis of the graph is the ratio between total cholesterol and HDL cholesterol, a lower ratio is preferred. The red box highlight the dot $(1.00, 3.60)$ which represents the ratio with no control. The green box highlight the dot $(0.82, 3.28)$ represent the drug level to be used to match the ration obtained by cutting in half the intakes.

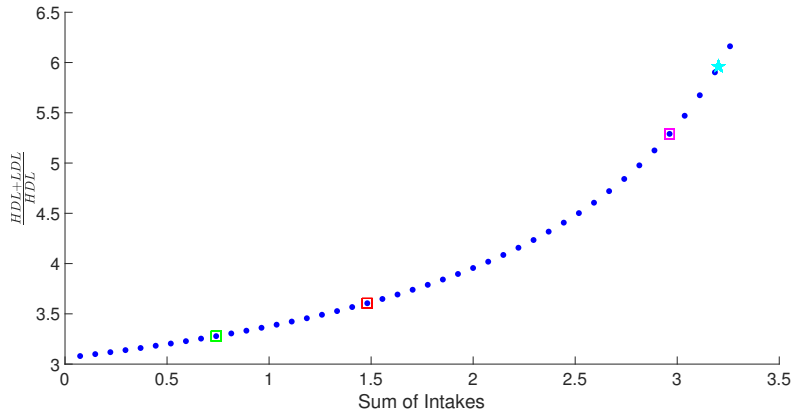


Figure 3.9: Graph showing the control of the system using the intakes as a control. The x-axis of the graph shows the total amount of intake. The y-axis of the graph is the ratio between total cholesterol and HDL cholesterol, a lower ratio is preferred. The red box highlight the dot $(1.48, 3.60)$ which represents the ratio with no control. The green box highlight the dot $(0.74, 3.28)$ which represents the ratio with half of the intakes. The magenta box highlight the dot $(2.96, 5.29)$ which represents the ratio with double intakes. The cyan pentagram highlight the dot $(3.20, 5.96)$ which represents the ratio with maximal intakes.

For cholesterol, the healthy state depends mostly on the ratio between total cholesterol and HDL cholesterol. The range of what is considered a healthy ratio depend on factors including age and gender, but in general a lower ratio of total cholesterol to HDL is preferred. In the case of the RCT network the healthy state $x_{f_{e,\pm}}$ depends on the value of v_4 (representing HDL) and v_6 (representing LDL), more precisely the ratio $\frac{v_4+v_6}{v_4}$. We examine whether an inhibitor or enhancer better lowers this ratio. To do this a set of fluxes were chosen from a uniform distribution so that each flux $f_i \in [0, 1]$. The fluxes f_{11} (corresponding to e_{11}) and f_{12} (corresponding to e_{12}) directly correspond to the drug and are treated separately, with f_{12} being fixed at 1, and f_{11} being the control variable. The result is rather intuitive, when the edge between v_4 and v_5 is enhanced the value of v_4 is lowered and the ratio increases, and when the edge is inhibited the value of v_4 increases and the ratio diminishes as seen in Figure 3.8. From the result it is clear that an inhibitor is the preferred type of drug for this example.

As explored in Section 3.4 the change in intakes also affects the state $x_{f_{e,\pm}}$. For comparison we also explored the effect of the intakes on the ratio $\frac{v_4+v_6}{v_4}$. Even when the control of intake is possible, large changes may not be realistic. Figure 3.9 shows how the ratio changes depending on the sum of the intake values. Comparing the control of drug to the control of intakes (see green boxes) we see that the same ratio is obtained by cutting in half intakes o using a drug that provides inhibition of the flow e_7 to 0.8 of its original value.

We show another example, where each metabolite has a desired target.

Example 6. Here we use the same initial conditions and target state from the example in Section 3.4, for a reminder Figure 3.4 shows an initial simulation with no control and highlights the desired equilibrium. Next, simulations are performed to obtain the desired equilibrium using the added drug from v_7 as a control. Two different k_{v_7} functions were used, one representing an enhancer drug and the other an inhibitor drug. As in the previous example, the enhancer function $k_{v_7}(x_{v_7}) = 1+\frac{x_{v_7}}{x_{v_7}+1}$ and inhibitor function $k_{v_7}(x_{v_7}) = 1 - \frac{x_{v_7}}{x_{v_7}+1}$ were used.

In the simulations it is assumed that the value of x_{v_7} can be easily controlled and kept at a constant value, this would correspond to a dosing regimen having constant intake of drug. Simulations were performed with both enhancers and inhibitors to find the best possible value of x_{v_7} for both types of drug. When only drug control was used it was found that using an inhibitor did not lower the cost between the optimized equilibrium and the desired equilibrium. Using an enhancer function the greater the value of x_{v_7}, the lower the cost. The upper bound of $x_{v_7} = 1$ was chosen, which corresponds to enhancement of the edge e_7 to 1.5 times its usual value and reduces the total cost to 4.1782. The results of this optimization are shown in Figure 3.10.

The control of the drug in this simulation has several limitations. For instance, only a single edge was targeted, whereas in pharmacology it is likely that many

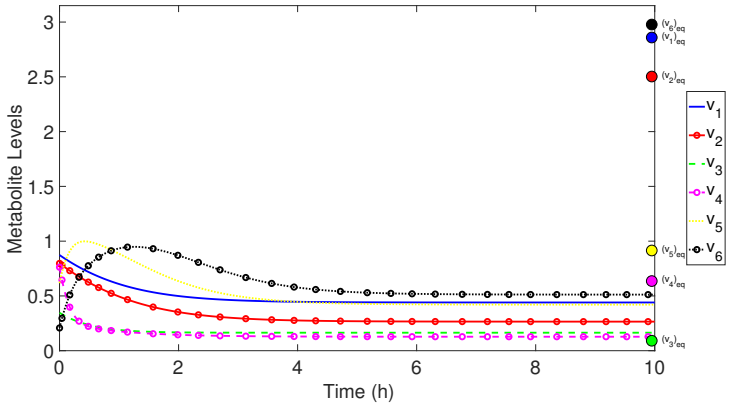


Figure 3.10: Control of the system only using the drug as an enhancer. The colored points show the desired equilibrium, while the lines show the actual trajectory of the metabolite values. The simulated equilibrium is $x_{drug} = (0.44, 0.27, 0.16, 0.15, 0.39, 0.51)$. The cost is 4.1782. The value of x_{v_7} is 1, which corresponds to $k(v_7) = 1.5$ enhance on edge e_7.

different targets may be considered and tested. Combining both methods, we control both the intakes and add a drug to act on e_7. Using optimized intakes, it was found that the inhibitor drug lowered the total cost between optimized equilibrium and the desired equilibrium. The drugs' optimal value was $x_{v_7} = 0.1765$, which corresponds to inhibition of the edge e_7 to 0.85 times its usual value. The results are shown in Figure 3.11.

Table 3.1: The total least-squares cost between obtained and target state of four different control schemes of RCT are displayed. The figures corresponding to the full trajectories are found in Figure 3.4, Figure 3.5, Figure 3.10, and Figure 3.11 respectively.

RCT systems	Cost
With no controls	4.1828
Control with intakes	0.3022
Control with drugs	4.1782
Control with intakes & drugs	0.2475

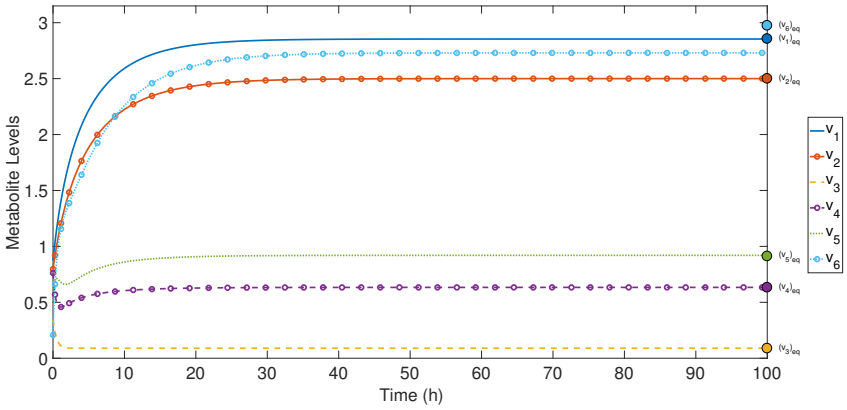


Figure 3.11: Control of the system by controlling the intakes and using the drug as an inhibitor. The simulated equilibrium is $x_{intakes\&drug} = (2.85, 2.50, 0.09, 0.63, 0.92, 2.73)$. The cost is 0.2475. The value of x_{v_7} is 0.1765, which corresponds to $k(v_7) = 0.85$ enhancement of e_7.

Three methods of control are performed; control using intakes (Section 3.4), control using drug, and finally control using both intakes and drug. The results of all three methods are summarized in Table 1. Controlling intakes rather than adding drug produced better results, although drug targeting of additional edges may further improve the drug results.

3.7 Conclusion

LIFE methodology was constructed to provide a new approach to modeling metabolic networks. Previously, in [99] the authors gave results which combined portions of several different theories and re-purposed them to shed light on simulating metabolic networks. The focus was on simple LIFE systems.

The intakes of a system are a reasonable target for control, and bio-molecules acting as inhibitors and enhancers are important structure which contributes to the stability of natural networks and often represent drug action. This work expanded the exploration of LIFE by investigating control problems: Controlling the intakes and controlling the inhibitor or enhancers of a network. Moreover, control methods are applied to a toy but significant example network of human metabolism called reverse cholesterol transport. To control the ratio of total cholesterol over HDL, the action of a single inhibitor was equivalent to a dramatic reduction of the intakes (cut by half). Our study on this network lead to conjectures about general networks. Future work will include addressing these conjectures to expand on LIFE methodology.

Chapter 4

Metabolic graphs, LIFE method and the modeling of drug action on *Mycobacterium tuberculosis*

4.1 Introduction

The modeling of metabolic networks plays a crucial role in systems biology and has many diverse applications, including in Quantitative Systems Pharmacology [108] for drug discovery and optimization of drug treatments. There are various challenges at modeling, including the complexity and dimensionality of the involved networks, often times comprising hundreds of metabolites and thousands of reactions, enzymes and genes. For this reason, only methods corresponding to linear dynamics presented the characteristic of scalability and computability to address such problems. In this area, Flux Balance Analysis (briefly FBA) plays a special role for its simplicity and many successful uses, see [53, 74, 85, 106, 117]. Other linear techniques proved efficient, such as zero deficiency theory, Markov chains, Laplacian dynamics [4, 20, 27, 45, 60, 65]. Beside the limitation of dealing only with linear dynamics, such methods mostly ne-

glected the nonlinear effects related to the action of enzymes, the regulation effect of genes and the action of drugs on genes. In mathematical terms, even if one accepts a linear dynamics in terms of the metabolites, the action of enzymes, genes and drugs affects fluxes among metabolites. For the example of a downregulation action of a gene, then reducing a flux would necessarily correspond to a nonlinear term in the dynamics of the involved metabolites. Even more, such action cannot be represented in the usual language of graph theory, thus requiring the use of more general theories. The present paper addresses such limitations using two main tools:

1. A new representation of metabolite dynamics called *Linear-In-Flux-Expression* (briefly LIFE) to allow nonlinearities in the dynamics.

2. The use of *metabolic graphs*, based on hyper and uber-graphs, to allow the representation of enzymes, genes and drug action.

The rest of this Introduction details the two main methods and provides a brief description of our main application: the action of antibiotics on the *Mycobacterium tuberculosis* (briefly MTB).

The paper is organized as follows. Section 4.2 discuss the LIFE approach and provide basic definition, while in Section 4.2.1 we illustrate the central carbon metabolism network of MTB. In particular the example shows the necessity of introducing the concept of metabolic graph which is done in Section 4.3. Extension of LIFE to metabolic graphs is done in Section 4.4 and the problem of existence and uniqueness of equilibria is explored in Section 4.4.1. Application of the methods to drug action on MTB (specifically synthesis of antibiotics) is illustrated in Section 4.5.

4.1.1 The LIFE method

Recently, flux balance analysis techniques were expanded to include nonlinear metabolite dynamics. Linear-in-flux-Expression allow using correlations among fluxes to our advantage. General conditions on network topology (connection of every node to ex-

cretion, see [99]) guarantee existence and uniqueness of an equilibrium $\bar{x}_f$ for every flux vector $f = \{f_e\}$, thus network asymptotic dynamics is captured by the map $f \to \bar{x}_f$. We will provide more detail below about the many results achievable by this approach combining several different methods for modeling chemical systems, including systems biology, zero deficiency theory, laplacian dynamics, and Markov chains [107, 4, 20, 65, 45, 60, 27]. We will also refer the reader to [5, 96, 99] for a general presentation of the LIFE approach.

4.1.2 Hyper, uber and metabolic graphs

As explained above FBA and other methods rely on representing the metabolic network as a directed graph, where edges represent biochemical reaction. There are (at least) three main limitations related to representing a complex metabolic network with a standard directed graph and these are:

1. Most networks include inflows and outflows (also called intakes and excretions in LIFE methodology) to the external environment or to other networks. Virtual nodes to represent such flow must be included or, alternatively, one must include directed edges with a node only on one end.

2. Some biochemical reactions necessarily involve more than two metabolites, e.g. when two or more compounds interact to form a set of other compounds. Therefore edges with multiple entering and exiting nodes must be included.

3. The action of enzymes, genes and drugs often times affect a specific reaction acting as enhancer or inhibitors. Such actions can be represented by edges joining a node to another edge.

All these extensions can be achieved by introducing appropriate generalization of the concept of graph. In particular hypergraphs [18, 138] contain hyperedges with multiple nodes, ubergraphs [68] include uberedges connecting node to other edges. We use this tool to defined a generalized graph, called metabolic graph addressing the limita-

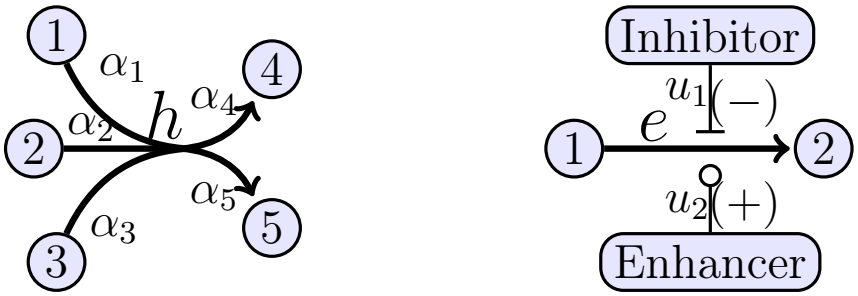


Figure 4.1: We define a *metabolic graph* to have two added features compared to simple directed graphs: 1. **Left** weighted hyperedge h will replace simple edges. These edges have weights assigned to each branch of the hyperedge respecting the stoichiometry of the corresponding reaction; and 2. **Right** enhancer or inhibitor dynamics acting on edge e. The inhibitor (enhancers) are included to model the action of molecules inhibiting (promoting) the enzyme for a reaction corresponding to edge e. The edges u_1, u_2 are called uberedges and connect a node to an edge.

tions 1-3. Figure 4.1 depicts the main idea behind the definition of metabolic graph. Metabolic graphs are hypergraphs which may include an uberedge which connects a node to a hyperedge.

4.1.3 Tuberculosis

The *Mycobacterium tuberculosis* (briefly MTB) has infected thirty percent of the world's population according to the World Health Organization(WHO). The WHO declared tuberculosis (brifely TB) a global emergency in 1993 [103]. The bacterium is known to endure hostile environments within the host organism through two main factors 1. genetically diverse sub-populations [26, 10] and 2. a sophisticated gene regulatory network(briefly GRN) [10, 51, 109]. Developing new TB drugs with improved effectiveness depends on modelling MTB in multiple metabolic states.

In the past, designing treatment for TB has been exceptionally difficult, often requiring drug cocktails of up to four different drugs. In addition to multiple drugs, treatment may last up to six months, and this duration along with complications due to drug side effects can negatively effect patient compliance. Two main approaches have been taken to improve TB treatment. the first approach is discovering more

effective dosing regimens of these drug cocktails [28, 95, 135]

The second approach is to anticipate the response of the GRN to treatment[133, 75, 32]. This allows one to exploit the changing metabolism of MTB, but requires accurate models of the response to drug treatment. Such models of MTB metabolism are difficult to construct due to the numerous complex effects of the GRN on the metabolic state as a consequence of environmental or internal chemical conditions.

In either approach, it is essential to model the individual drug interactions carefully so that a model can be built with a solid foundation. Modeling drug interactions in an intuitive way requires us to allow structures where molecules or metabolites inhibit or enhance a biochemical reaction. This action can be depicted as a node attached to a hyperedge (connecting nodes which represent multiple reactant metabolites to nodes representing products of the reaction).

4.2 Modeling metabolic networks with LIFE

Generalizing the dynamics of FBA, we focus on the following class of systems know as LIFE, see (1.3).

Metabolic networks contain exchange fluxes representing incoming mass from other parts of the network, or from the outside environment, and they are necessary for the existence of equilibria. For traditional metabolic networks described by a directed graph $G = (V, E)$, directed edges from a virtual node v_0 which acts as a source to some node in the graph are called intakes, and edges from a node in the graph to some virtual node v_{n+1} acting as a sink are called excretions.

Definition 8. Given a directed graph $G = (V, E)$, an edge $(v_0, v_1) \in E$ for the virtual node v_0, the edge (v_0, v_1) is called an intake, and v_1 is an intake node. For the virtual node v_{n+1} an edge (v_j, v_{n+1}) is called an excretion, and v_j is an excretion node. The set of intake nodes is denoted I, and the set of excretion nodes is denoted by J.

A general assumption, still allowing significant results, is the following:

$$(H1) \qquad S_{ve}(x) = \begin{cases} -F_e(x_v) & e = (v, w),\ v \in V, w \in V \cup \{v_{n+1}\} \\ F_e(x_w) & e = (w, v),\ w \in V \\ 1 & e = (v_0, v)\ v \in I \\ 0 & \text{otherwise,} \end{cases}$$

where $F_e : \mathbb{R}_+ \to \mathbb{R}_+$ is differentiable, strictly increasing, with $F_e(0) = 0$, for $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$. (H1) is a natural assumption on the system; the flow from a metabolite will depend only on that metabolite, but for a metabolite with multiple edges may have different kinetics with each reaction. This assumption also includes nonlinear kinetics such as Michaelis-Menten [79] corresponding to Hill functions $F_v(x_v) = \frac{x_v^p}{K+x_v^p}$ with $p \in \mathbb{N}$. Moreover, all columns of S have zero sum, except those corresponding to intakes and excretions, which have positive and negative sum, respectively. Therefore, the dynamics (1.3) can be interpreted as mass conservation law. To be biologically meaningful, we further restrict to equilibria for which all components of f and x are positive, thus rendering the problem nonlinear even for linear dynamics. A previous result shows that under (H1), the existence of equilibria depends on the structure of the network.

Proposition 33. *Consider a system* (1.3) *satisfying (H1). Assume there exists an equilibrium* $\bar{x} \in (\mathbb{R}_+)^n$ *for a flux vector* f *such that* $f_e > 0$ *for every* $e \in \tilde{E}$*. Then for every node* $v \in V$ *for which there exists a path from* I *to* v*, there exists a path from* v *to* J*.*

This proposition is proven in [99]. We extend these results to metabolic networks including enhancers and inhibitors as explained in Section 4.3.

4.2.1 Central Carbon Metabolism in *Mycobacterium tuberculosis*

The dynamics described in the previous section can be used to model the flows of a metabolic network. The entries of the stoichiometric matrix describe the mass flowing along an edge of the network per time. The flux for each edge represents the speed of this flow. We present an example metabolic network from [128] to motivate expanding the LIFE method from section 4.2.

The Metabolic network shown in Figure 4.2 is the central carbon metabolism of MTB derived from an infection in mouse. The network shown is a metabolic network of MTB in a state of growing bacilli. This network depicts carbon from lipid and sugar catabolism used by MTB for generating energy and biosynthetic precurors required for growth. Specifically glucose is a product of catabolism and the cell uses this energy to synthesize enzymes for the pentose phosphate pathway and provide ribose 5-P for nuclotide synthesis. Glycolysis yields metabolites phosphoenolpyruvate(PEP), Pyruvate, and acetyl-CoA. Experiments showed that MTB preferentially uses fatty acids as a carbon source. When MTB is in the "non-growing" state within mouse infection, some of the genes required for this process are known to be upregulated. These observations led to the proposal that MTB switches its carbon source from sugars to fatty acids during the persistent phase of infection.
Simple directed edges, as in standard graphs, model most reactions in carbon metabolism. However, the complete dynamics involves several hyperedges, i.e. generalized edges connecting more than two nodes. The hyperedge labeled $h_{1,2}$ consists of two distinct directed hyperedges, each connecting three metabolites involved in a single chemical reaction (the first having two reactants and the second having two products). The details about the reactions are taken from the KEGG database [71]. In particular, h_1 describes a chemical reaction producing isocitrate from Oxaloacetate, and acetyl CoA, and h_2 uses isocitrate to produce acetyl CoA and oxaloacetate. The third hy-

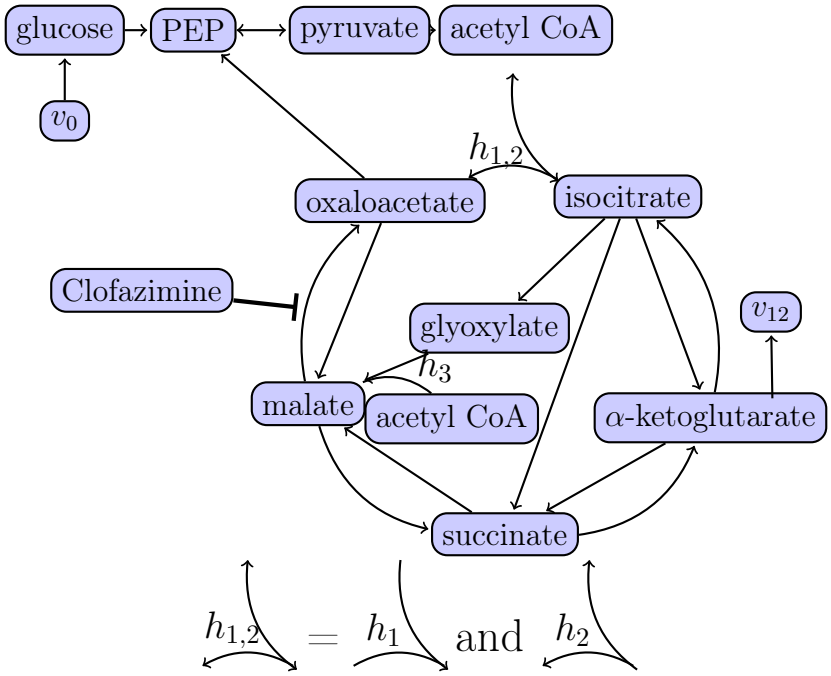


Figure 4.2: The central carbon metabolism of *Mycobacterium tuberculosis* with TB drug action shown. The well known drug Clofazimine inhibits the reaction which oxidizes malate. The edges labeled $h_{1,2}$ represents two different hyperedges, and h_3 is a third hyperedge connecting three metabolites.

peredge h_3 models acetyl CoA reacting with glyoxylate to produce malate. We also included the drug Clofazimine, which acts as an inhibitor on the malate-oxaloacetate reaction.

The need to describe reactions depicted by the hyperedges and inhibitor in Figure 4.2 is what brings us to metabolic graphs and extending the LIFE method accordingly.

4.3 Metabolic graphs

To model intakes, excretions, general biochemical reactions, inhibitors and enhancers acting on a metabolic network, we will introduce a new mathematical object called a *metabolic graph*. Metabolice graphs are a subclass of ubergraph, a generalization of hypergraphs introduced in [68] (here we use a different definition of ubergraph stated in the Appendix).

Directed graph are commonly used to model metabolic networks, with nodes representing metabolites and edges representing the biochemical reactions. A hypergraph is a more general structure where an edge can connect more than two nodes. For the following definitions, let $V = \{v_1, \ldots, v_n\}$ be a set of nodes and let $\mathcal{P}(V)$ be the power set of V.

Definition 9. A *hyperedge* h is a set of nodes connected to each other, i.e. $h \in \mathcal{P}(V) \setminus \{\emptyset\}$. Note that the set $h = \{v_i\}$ indicates a loop edge connecting v_i to itself.

Because chemical reactions of a metabolic network indicate a direction of flow, we require hypergraphs to include direction.

Definition 10. A *directed hyperedge* is an ordered pair of two subsets of nodes, i.e. $h = (X, Y)$ with $X \in \mathcal{P}(V)$ or $X = \{v_0\}$, $Y \in \mathcal{P}(V \cup \{v_{n+1}\})$, where v_0 (v_{n+1}) is a virtual node called the source (sink). Elements of X (Y) are called initial nodes (terminal nodes) for the hyperedge h. The set of directed hyperedges is denoted $\mathcal{H}$.

When using a hyperedge to model reacting metabolites forming a product, we encode the stoichiometry relationship on the a hyperedge via edge weights. It is convenient to have notation for the cardinality of the sets of initial and terminal nodes for a hyperedge.

Definition 11. Given a directed hyperedge $h = (X, Y)$, with $X, Y \in \mathcal{P}(V)$, the *indegree* of h is defined as

$$d_{\text{in}}(h) = |X|, \tag{4.1}$$

and the *outdegree* of h is defined as

$$d_{\text{out}}(h) = |Y| \tag{4.2}$$

where $|\cdot|$ indicates the cardinality of a set.

Definition 12. A weighted directed hyperedge is a couple $\mathcal{H} \ni h = (X, Y)$ with $X \in \mathcal{P}(V)$ or $X = \{v_0\}$, $Y \in \mathcal{P}(V \cup \{v_{n+1}\})$, and corresponding weights $\Psi_h : h \mapsto (\Psi_h^{\text{out}}, \Psi_h^{\text{in}})$ where $\Psi_h^{\text{out}} : X \mapsto \mathbb{R}_+$ and $\Psi_h^{\text{in}} : Y \mapsto \mathbb{R}_+$.

We are now ready to describe a mathematical representation of inhibitors and enhancers. More precisely, we want to consider metabolites influencing a given reaction. This can be captured by new type of generalized edges, called an *uberedge* linking nodes to directed hyper edges.

Definition 13. An e/i-uberedge is a couple $u = (v, h)$ with $v \in V, h \in \mathcal{H}$. We denote the set of e/i-uberedges by $\mathcal{U}$.

We are now ready to provide the definition of metabolic graph. For a complete description of a metabolic networks, uberedges are endowed with a sign to indicate their action as enhancer $(+)$ or inhibitor $(-)$. This subset of uberedges from a node to a hyperedge are depth 2 uberedges according to the general definition of ubergraph (see Appendix).

Definition 14. A metabolic graph is a weighted directed hypergraph endowed with signed depth-2 uberedges connecting nodes to hyperedges. More precisely, a metabolic graph is an ordered quintuplet $G = (V, \mathcal{H}, \mathcal{U}, \Psi_{\mathcal{H}}, \Psi_{\mathcal{U}})$ where $V = \{v_1, \ldots, v_n\}$ is the set of nodes, $\mathcal{H}$ is the set of directed hyperedges, $\Psi_{\mathcal{H}} = \{\Psi_h : h \in \mathcal{H}\}$ is the set of functions assigning weights to hyperedges, $\mathcal{U}$ is the set of e/i-uberedges and $\Psi_{\mathcal{U}} : \mathcal{U} \mapsto \{+, -\}$.

Definition 15. Given a metabolic graph G, a *path* is a sequence of distinct nodes $v_{i_1} \cdots v_{i_k}$, with $(v_{i_j} \in X, v_{i_{j+1}} \in Y)$ and $(X, Y) \in \mathcal{H}$ for $j = 1, \ldots, k-1$. A graph is *strongly connected* if there exists a path between every pair of nodes. A *strongly connected component* of a directed graph is a maximal strongly connected subgraph. A *terminal component* of a metabolic graph G is a strongly connected component

corresponding to a subset of nodes $V' \subset V$, such that for $v' \in V'$ and $v \in V \setminus V'$ there exists no hyperedge $h = (X, Y)$ with $v' \in X$ and $v \in Y$.

Definition 16. Given a metabolic graph G, *Intakes* (*Excretions*) are hyperedges (X, Y) such that $X = \{v_0\}$ ($v_{n+1} \in Y$). *Intake nodes* are nodes $v \in V$ such that there exits a hyperedge (X, Y) with $X = \{v_0\}$ and $v \in Y$. *Excretion nodes* are nodes $v \in V$ such that there exits a hyperedge (X, Y) with $v \in X$ and $v_{n+1} \in Y$. We indicate by I the set of intake nodes and by J the set of excretion nodes.

4.3.1 Central Carbon Metabolism

Let us go back to the central carbon metabolism network of MTB represented in Figure 4.2. This network can be described by a metabolic graph $G = (V, \mathcal{H}, \mathcal{U}, \Psi_{\mathcal{H}}, \Psi_{\mathcal{U}})$ as follows. The set V of nodes is given by:

$V = \{v_1, \ldots, v_{11}\} = \{\text{glucose}, \text{PEP}, \text{pyruvate}, \text{acetyl CoA}, \text{isocitrate}, \text{oxaloacetate}, \text{glyoxylate}, \text{malate}, \alpha - \text{ketoglutarate}, \text{succinate}, \text{Clofazimine}\}$.

The set of hyperedges $\mathcal{H}$ contains regular edges and three directed hyperedges:

$h_1 = (\{\text{acetyl CoA}, \text{oxaloacetate}\}, \{\text{isocitrate}\})$,

$h_2 = (\{\text{isocitrate}\}, \{\text{acetyl CoA}, \text{oxaloacetate}\})$,

and $h_3 = (\{\text{acetyl CoA}, \text{glyoxylate}\}, \{\text{malate}\})$.

The set of uberedges $\mathcal{U}$ has only one uberedge:

$u_1 = (\text{Clofazimine}, (\{\text{malate}\}, \{\text{oxaloacetate}\}))$.

For simplicity the functions $\Psi_{\mathcal{H}}, \Psi_{\mathcal{U}}$ are not listed, but can be deduced from the KEGG database [71].

This graph is more descriptive of important functions of drug action on the metabolic network, and it contains three hyperedges and a single uberedge. Each of the hyperedges combines three metabolites; h_1,h_3 have two reactants and one product whereas h_2 has two reactants and one product. The uberedge shown from Clofazimine to the malate-oxaloacetate reaction indicates the well known TB drug "Clofazimine" acts as

an inhibitor on a reaction which oxidizes malate.

4.4 Metabolic dynamics with inhibitors and enhancers

In the treatment of TB, drugs will act as enhancers and inhibitors to various edges in the network, and so will appear as uberedges. In this section we define how enhancers and inhibitors affect the dynamics as well as equilibria conditions. In order to include enhancers and inhibitors (which serve as the initial nodes for uberedges) we introduce a new assumption:

(H2) We assume that for every node v and hyperedge $h = (X, Y)$, with $v \in X$, the following holds. Let U_h be the set of nodes w such that there exists e/i-uberedge $(w, h) \in \mathcal{U}$, then we have:

$$S_{vh}(x) = \begin{cases} -\alpha_v \cdot \mathbf{F}_h(x) \cdot \mathbf{K}_h(x) & v \in X \\ \alpha_v \cdot \mathbf{F}_h(x) \cdot \mathbf{K}_h(x) & v \in Y \\ 1 & X = \{v_0\}, v \in Y, \\ 0 & \text{otherwise,} \end{cases} \tag{4.3}$$

where $\alpha_w = \Psi_h^{\text{in}}(w)$ if $w \in X$ and $\alpha_w = \Psi_h^{\text{out}}(w)$ if $w \in Y$ are the stoichiometric coefficients, $\mathbf{F}_h : \mathbb{R}^{d_{in}(h)} \to \mathbb{R}_+$ is given by

$$\mathbf{F}_h(x) = \min_{w \in X} \left\{ F_{w,h}(x_w) \cdot \frac{1}{\alpha_w} \right\}, \tag{4.4}$$

$F_{w,h} : \mathbb{R}_+ \to \mathbb{R}_+$ quantifies the potential flow of metabolite x_w due to reaction h, and

$$\mathbf{K}_h = \prod_{w \in U_h} K_{(w,h)}(x_w), \tag{4.5}$$

where $K_{(w,h)} : \mathbb{R}_+ \to \mathbb{R}_+$ quantifies the action of metabolite x_w on h, with the convention that $K_h = 1$ if $U_h = \emptyset$.

The stoichiometric coefficients $\alpha_w \in \mathbb{R}$ for the reaction corresponding to hyperedge $h$are normalized such that $\sum_{w\in X} \alpha_w = 1$ and $\sum_{w\in Y} \alpha_w = 1$. The functions $F_{w,h}$, $w \in Y$ are continuously differentiable. The functions $K_{(w,h)}$, $w \in U_h$, are continuously differentiable, monotonic with $K_{(w,h)}(0) = 1$. More precisely if $\Psi_U((w,h)) = +$ then $K_{(w,h)}$ is increasing (enhancer case), othwerwise $K_{(w,h)}$ is decreasing (inhibitor case).

Similarly to (H1), under assumption (H2) each function $F_{w,h}$ depends only on the metabolite x_w, but there is the additional factor $\mathbf{K}$ which corresponds to the action of one or more e/i-uberedges. This gives a nonlocal dependence, with respect to network topology, because the node(s) corresponding to an enhancer(s) or inhibitor(s) may be anywhere in the network not necessarily close to the edge it is affecting.

We are ready to state our first result:

Proposition 34. *Consider a system* (1.3) *satisfying (H2). Assume there exists an equilibrium $\bar{x} \in (\mathbb{R}_+)^n$ for a flux vector f such that $f_h > 0$ for every $h \in \mathcal{H}$. Then for every node $v \in V$ for which there exists a path from the intake nodes to v, there exists a path from v to the excretion nodes.*

Proof. Assume there exists an equilibrium $\bar{x} \in {\mathbb{R}_+}^n$ and, by contradiction, a node v for which there exists a path from w, an intake node, to v, but there exists no path from v to some excretion node. Since there is no path from v to excretion nodes, either v belongs to a terminal component, or there is a path from v to a terminal component with no excretion. Denote by $V_T \subset V$ the set of nodes of such a terminal component. Since there is a path from w, an intake node, to v and a (possibly trivial) path from v to V_T, then there is also a path from intake nodes to V_T. Denote by $v_0, v_1 = w, \ldots, v_{\ell-1}, v_\ell$ one such a path, such that $v_{\ell-1} \notin V_T$ and $v_\ell \in V_T$ (possibly

the path is a single hyperedge, in the case with $w \in V_T$).

It is easy to show that $x_{v_i} > 0$ for all $i = 1, \dots, \ell$, as follows. If $h = (X, Y)$ with $v_0 \in X, v_1 \in Y$, then by (H2) we have $S_{v_0,h}(\bar{x}) = 1$. On the other side, for every $h' = (X, Y)$ with $v_1 \in X, w' \in Y$, $x_{v_1} = 0$ implies $S_{v_1,h'}(x) = 0$. Consequentially $x_{v_1} = 0$ implies $\dot{x}_{v_1} \geq \alpha_{v_1} f_h \mathbf{F}_h(\bar{x}) > 0$ (where $h = (X, Y)$ with $v_0 \in X, v_1 \in Y$), contradicting $\bar{x}$ being an equilibrium. Having proved that $\bar{x}_{v_1} > 0$, we can proceed by induction: for $i = 1, \dots, \ell - 1$, $\bar{x}_{v_i} > 0$ implies $\bar{x}_{v_{i+1}} > 0$. The argument is the same as above, with a slight modification: looking at $h = (X, Y)$ with $v_i \in X, v_{i+1} \in Y$, $S_{v_{i+1},h}(\bar{x}) = \alpha_v \cdot \mathbf{F}_h(\bar{x}) \cdot \mathbf{K}_h(\bar{x}) > 0$ thanks to (H2) together with $\bar{x}_i > 0$, while above we were in the case of an intake.

Finally we have a terminal component with no excretion, and a hyperedge $h = (X, Y)$ with $v_{\ell-1} \in X, v_\ell \in Y$ with $v_\ell \in V_T$ and $v_{\ell-1} \notin V_T$, such that either $v_{\ell-1} = v_0$ or $\bar{x}_{v_{\ell-1}} > 0$. In either case, considering $\tilde{h} = (X, Y)$ with $v_{\ell-1} \in X, v_\ell \in Y$, by (H2) we have $S_{v_\ell,\tilde{h}}(\bar{x}) = \alpha_{v_\ell} \cdot \mathbf{F}_{\tilde{h}}(\bar{x}) \cdot \mathbf{K}_{\tilde{h}}(\bar{x}) > 0$. Now consider the variation of mass in the nodes of the component V_T: since there are no hyperedges leaving V_T, and there is at least the incoming hyperedge $\tilde{h}$, we have $\frac{\mathrm{d}}{\mathrm{d}t} \sum_{v \in V_T} x_v = \sum_{v \in V_T} \dot{x}_v \geq \alpha_{v_\ell} \cdot \mathbf{F}_{\tilde{h}}(\bar{x}) \cdot \mathbf{K}_{\tilde{h}}(\bar{x}) f_{\tilde{h}} > 0$, contradicting the fact that $\bar{x}$ is an equilibrium. □

For a system with fixed metabolites, the existence of feasible flows will depend on network structure. The Max-flow-min-cut Theorem [47] implies that a feasible flow exists if there is a path from intakes to excretions, and indeed this Theorem holds true also for metabolic graphs providing suitable definitions.

Definition 17. A flow on a metabolic graph $G = (V, \mathcal{H}, U, \Psi_{\mathcal{H}}, \Psi_U)$ is a function $g : \mathcal{H} \mapsto R_+$ such that g satisfies Kirchhoff's law for metabolic graphs, i.e. for every node v:

$$\sum_{h \in \Gamma^{out}(v)} \Psi_h^{out}(v) \cdot g(h) = \sum_{k \in \Gamma^{in}(v)} \Psi_k^{in}(v) \cdot g(k) \tag{4.6}$$

where $\Gamma^{out}(v) = \{(X, Y) \in \mathcal{H} : v \in X\}$, $\Gamma^{in}(v) = \{(W, Z) \in \mathcal{H} : v \in Z\}$.

Definition 18. For a given flow g on a metabolic graph G, the *amount of flow* from v_0 to v_{n+1} is $v(g) = \sum_{h\in\mathcal{H},v_0\in X} g(h) = \sum_{h\in\mathcal{H},v_{n+1}\in Y} \Psi_h^{out}(v_{n+1}) \cdot g(h)$.

Maximal Flow Problem. Consider a metabolic graph with a function $c : \mathcal{H} \mapsto R_+$ assigning to each edge a maximal capacity. The maximal flow problem is defined

$$\max(v(g)) \text{ such that } g(h) \leq c(h) \text{ for every } h \in \mathcal{H}. \tag{4.7}$$

We are now ready to define a cut.

Definition 19. Consider a metabolic graph $G = (V, \mathcal{H}, \mathcal{U}, \Psi_{\mathcal{H}}, \Psi_{\mathcal{U}})$. Given $S, T \subset V$, we define $\mathcal{H}(S,T) = \{(X,Y) \in \mathcal{H} : X \cap S \neq \emptyset \text{ and } Y \cap T \neq \emptyset\}$, as the set of edges connecting nodes of S to nodes of T.
Given a flow g on G, the total flow from nodes in S to nodes in T is defined by:

$$g(S,T) = \sum_{h=(X,Y)\in\mathcal{H}(S,T),i\in Y\cap T} \alpha_i h.$$

Definition 20. Consider a metabolic graph $G = (V, \mathcal{H}, \mathcal{U}, \Psi_{\mathcal{H}}, \Psi_{\mathcal{U}})$ with source v_0, and sink v_{n+1}. Let $S \subset V \cup v_0$ be a set such that $v_0 \in S$ and $v_{n+1} \notin S$ and define $T = (V \cup \{v_{n+1}\}) \setminus S$. Then the set of hyperedges $C_S = \mathcal{H}(S,T)$ is called a *cut separating v_0 from v_{n+1}*. The capacity of the cut is defined by:

$$c(C_S) = \sum_{h\in C_S} c(h). \tag{4.8}$$

Notice that the capacity of the branch of hyperedge h adjacent to a node v can be written as $c(h) \cdot \alpha_v$.

The max flow of a directed graph is not necessarily unique. For a graph with a directed cycle, the flow on the cycle can be increased (capacity permitting) without violating Kirchoff's law, or changing the total flow through the graph. For example,

the directed graph in Figure 4.3 (left) admits the maximal flow defined by setting the flow on edges (s,1),(1,2),(2,3) and (3,t) equal to 10 while the flow on (4,1),(3,4) equal to 0. However, the maximal flow can also be achieved setting (s,1),(3,4),(4,1) and (3,t) to 10 while the flow on (1,2),(2,3) to 20. Often when computing maximum flow on directed graph the flow through cycles is eliminated. In a metabolic graph with hyperedges, flow through cycles may be unavoidable. The metabolic graph in 4.3 (right) contains one hyperedge, with branches that are equally weighted, one of them leading to a sink, while the other leads to a cycle. This network admits a unique maximum flow. This is produced by setting (s,1),(3,4),(4,1) to 10 while the flow on (1,2),(2,(3,t)) to 20. Note that setting the flow (2,(3,t)) to 20 will cause 10 to flow to t and 10 to flow to 3 due to this hyperedge having equal weighting to its branches.

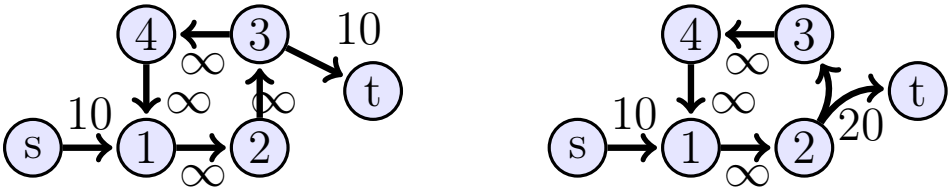


Figure 4.3: A directed graph (left) and a metabolic graph (right) with similar structure. Capacities are listed on edges and the hyperedge has a 1:1 ratio of flow for the two branches, i.e., 10 units flow across each branch. The maximum flow for each graph is 10, however the metabolic graph has a unique solution while the directed graph does not.

We are now ready to state the following:

Lemma 4. *Given a flow g from v_0 to v_{n+1}, and a cut C_S separating v_0 from v_{n+1}, it holds: $v(g) \leq c(C_S)$*

Proof. Fix a flow g and a cut C_S separating v_0 from v_{n+1}. The amount of flow $v(g)$ is equal to the flow entering the graph, equivalently the amount of flow entering S. We also have $v(g)$ will equal the transfer of flow from S to T,

$$v(g) = g(S,T) - g(T,S).$$

Since flows are positive we have

$$g(S,T) - g(T,S) \leq g(S,T) = \sum_{h=(X,Y)\in C_S, i\in Y\cap T} \alpha_i h \leq \sum_{h\in C_S} h = c(C_S).$$

□

Since Lemma 4 applies to every flow g and cut C_S, we can take the maximum over all flows and the minimum over all cuts to obtain

$$\max(v(g)) \leq min(c(S)). \tag{4.9}$$

The following Proposition shows that, as for simple graphs, an equality holds in (4.9).

Proposition 35. *(Max-flow-min-cut for Metabolic graphs) Formula* (4.9) *holds with equality sign, i.e.:*

$$\max(v(g)) = \min(C_S)$$

Proof. Take a flow g such that $v(g)$ is maximal. We construct a cut starting from the intake nodes by recursion. Define $S_0 = I$ to be all intake nodes of the metabolic graph. At each step ν, we define $S_{\nu+1}$ from S_ν as follows:

1. If $v \in S_\nu$ then $v \in S_{\nu+1}$,
2. if $v_1, v_2 \in S_\nu, w \in V$ and there exists a hyperedge $h = (X,Y)$ such that $v \in X, w \in Y$ and $g(h) < c(h)$ or $w \in X, v_2 \in Y$ and $g(h) > 0$,then $w \in S_{\nu+1}$,
3. if $v_1, v_2 \in S_\nu$ and there exists a hyperedge $h = (X,Y)$ such that $v_1 \in X, v_2 \in Y$, then for all $w \in Y$, $w \in S_{\nu+1}$

Clearly there will only be a finite number of steps until S_ν is stable, i.e. no more edges will be added after the $\bar{\nu} - th$ step. We define $S = S_{\bar{\nu}}$ and $T = V \setminus S$.
We claim that $v_{n+1} \notin S$. Assume, by contradiction that there exists a path $v_{i_1}, ..., v_{i_l}$

such that $v_{i_1} = v_0$, $v_{i_l} = v_{n+1}$. Then define for $k = 1, ..., l-1$,

$$\epsilon_k = \max \left\{ \sum_{h=(X,Y):v_{i_k} \in X, v_{i_{k+1}} \in Y} (c(h) - g(h)), \sum_{h=(X,Y):v_{i_{k+1}} \in X, v_{i_k} \in Y} g(h) \right\} > 0$$

and $\epsilon = \min_k \epsilon_k$. We now define a new flow $\tilde{g}$ to reach a contradiction as follows. If $\epsilon_k = \sum_{h=(X,Y):v_{i_k} \in X, v_{i_{k+1}} \in Y} (c(h) - g(h))$, then for each $h \in K_1$ we set

$$\tilde{g}(h) = g(h) + \epsilon.$$

Otherwise we set

$$\tilde{g}(h) = g(h) - \epsilon.$$

By construction $\tilde{g}$ is a flow, moreover $v(\tilde{g}) = v(g) + \epsilon$. Then we reach a contradiction for the maximality of $v(g)$.

Define the cut $C_S = \mathcal{H}(S, T)$. Then the flow from v_0 to v_{n+1} satisfies $v(g) = g(S, T) - g(T, S)$. Since g is maximal, we have $g(T, S) = 0$ and $g(S, T) = \sum_{h \in C_S} (g(h)) = \sum_{h \in C_S} c(h) = c(C_S)$. Therefore we have $v(g) = c(S)$. □

Definition 21. Given a metabolic graph with intakes, the vector $\bar{\phi}$ represents the intake flows to each node, i.e. $\bar{\phi}_i = f_{h(v_0, v_i)}$ if v_i is an intake node and $\bar{\phi}_i = 0$ otherwise.

Proposition 36. *Given a system satisfying (H2), fix an $x \in (\mathbb{R}_+)^n$ and intake flow vector $\bar{\phi}$ with strictly positive entries. There exists $f \in (\mathbb{R}_+)^m$ in the null space of $S(x)$ if and only if for each intake there exists a path to an excretion.*

Proof. Consider the maximum flow problem on the metabolic graph G, where intake edges h_I have capacity $\bar{\phi}$, and all other edges have infinite capacity. The feasible flows φ for this network are in one-to-one correspondence with the equilibrium fluxes $f \in \mathcal{N}(S(x)) \cap (\mathbb{R}_+)^m$ (where $\mathcal{N}(S(x))$ indicates the kernel of the matrix $S(x)$) such that $f_h \leq \bar{\phi}$ for all $h \in h_I$. The correspondence is simply given by $f_h = \varphi_h$ for all

$h \in h_I$ and $f_h = \varphi_h / F_h(X)$ for all $h = (X, Y) \in \mathcal{H} \setminus h_I$.

If for all $h \in h_I$ there is a path to an excretion, then the minimum cut is the collection of all edges $h \in h_I$. The maximum flow φ^* then satisfies $\varphi^*(h_I) = \bar{\phi}$, thus also ensuring the existence of an equilibrium flux f^* satisfying the same property.

If for some $h \in h_I$ there is no path to an excretion, then all feasible flows φ satisfy $\varphi_h = 0$, and hence all equilibrium fluxes f^* satisfy $f_h^* = 0$ which contradicts the assumption. □

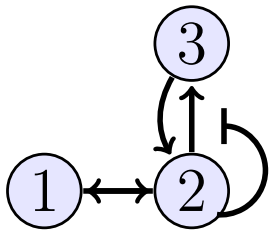


Figure 4.4: In this metabolic graph node 2 acts as an inhibitor to an edge leaving itself.

In [92] several Propositions guarantee boundedness of solutions and no periodic oscillation. Specifically, the main results are: (1) Trajectories are bounded if and only if there exists an equilibrium with positive entries; (2) If trajectories are bounded, then they approach an equilibrium set for $t \to \infty$. These conclusions are also true under assumption (H1) for simple graphs, but not under assumption (H2) for metabolic graphs.

We start providing a counterexample via a small system of three nodes and one inhibitor, see Figure4.4. We assume (H2) to hold with $F(x_v) = x_v$ for all nodes v,

$$K_2(x_2) = \begin{cases} \frac{1}{3}((x-6)^2 + 5) & x \leq 6 \\ \\ \frac{5}{3} & 6 < x, \end{cases}$$

and stoichiometric matrix given by

$$S = \begin{pmatrix} -x_1 & 0 & x_2 & 0 \\ x_1 & -x_2 \cdot K_2(x_2) & -x_2 & x_3 \\ 0 & x_2 \cdot K_2(x_2) & 0 & -x_3 \end{pmatrix},$$

with flux vector $f = [1, 3, 3, 2]$. The structure is visualized in Figure 4.4. This example will admit an oscillatory solution because $x_2 \cdot K_2(x_2)$ is not monotone increasing and has a negative slope for $\frac{1}{3}(12 - \sqrt{21}) < x_2 < \frac{1}{3}(12 + \sqrt{21})$. The example reduces to the example given in Remark 2 of [92], where further details can be found.

Under assumption (H1) we have several additional conditions which relate to existence and uniqueness of equilibria. These results again depend on the structure of the network as well as an additional constraint on the functions F_h. Adapted from [92, Theorems 4 and 5] these are: (1) There exists an equilibrium with positive entries for constant intakes if and only if for all $v \in V$ there is a path to X such that all edges in the path have $\lim_{x_v \to \infty} H_e(x_v) = +\infty$; (2) If there exists an equilibrium with positive entries, and all $v \in V$ connect to J, then the equilibrium is unique. Even if we are able to relate the existence of equilibria to network structure via Proposition 34 and 36), uniqueness of equilibria may fail without further assumption. Moreover, computation of equilibria is highly nontrivial. Indeed, for a simple graph with a single enhancer or inhibitor, we can write the dynamics as follows

$$\dot{x} = J_1(f)\eta(x) + J_2(f)k(x) + \phi \tag{4.10}$$

where J_1 is the Jacobian matrix without the edges affected by inhibitor/enhancer, J_2 is a matrix with the edges affected by inhibitor/enhancer, $\eta_i(x) = F_{v_i}(x_{v_i})$, $\phi_i = f_{(v_0,v_i)}$ if $(v_0, v_i) \in E$, and $\phi_i = 0$ otherwise, and $k_i(x) = K_{v_j}(x_{v_j}) \cdot F_{v_i}(x_{v_i})$ if v_j acts on an edge starting from v_i. If another node acts as an inhibitor, as it happens for the ad-

dition of drug to the system, the size of J_1, J_2 would be adjusted accordingly. Clearly (4.10) in general does not have a unique solution and the computation of solutions can not be done analytically.

4.4.1 Exploring the space of equilibria with inhibitors and enhancers

In this Section we further explore the problem of uniqueness and stability of equilibria for networks with enhancers and inhibitors. First, we provide an example of non-uniqueness and a condition to ensure uniqueness. Let us define two classes of systems:

Class A. Enhancers and inhibitors act in cascade, see Figure 4.5 top.

Class B. Enhancers and inhibitors act in parallel, see Figure 4.5 bottom.

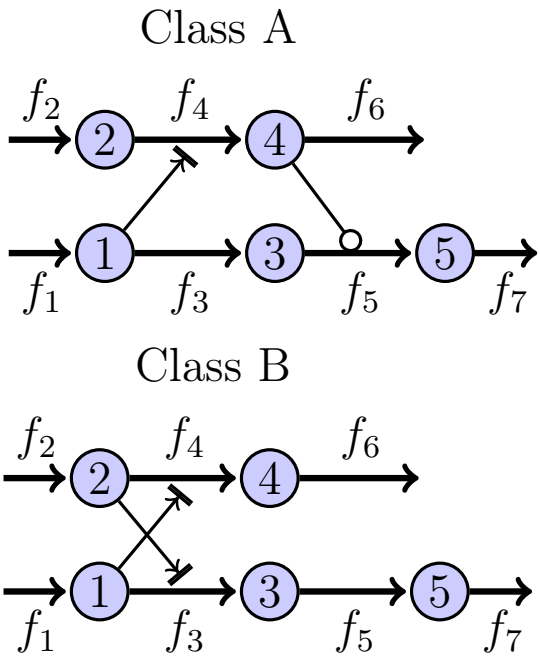


Figure 4.5: Two networks with inhibitors and enhancers. Class A (top): inhibitor/enhancers behave sequentially: node 1 affects downstream f_4 and so on. Class B (bottom), inhibitors/enhancers act in parallel.

To precisely define Class A and Class B networks let us first define an *uberpath* and

strictly upstream on a metabolic graph $G = (V, \mathcal{H}, \mathcal{U}, \Psi_{\mathcal{H}}, \Psi_{\mathcal{U}})$.

Definition 22. An *uber-path* is a sequence of nodes $(v_{k_1}, \ldots, v_{k_m})$ such that for every $i = 1, \ldots, m-1$ there is an hyperedge $(X, Y) \in \mathcal{H}$ for which one of the following holds: 1) $v_{k_i} \in X$, $v_{k_{i+1}} \in Y$; 2) $(v_{k_i}, (X, Y)) \in U$, $v_{k_{i+1}} \in X \cup Y$; 3) $(v_{k_{i+1}}, (X, Y)) \in \mathcal{U}$, $v_{k_i} \in X \cup Y$.

Definition 23. A node v_0 is *upstream* of node v_1 if there exists a hyperpath from v_0 to v_1. A node v_0 is *strictly upstream* of node v_1 if v_0 is upstream of v_1 and there is no uberpath from v_1 to v_0.

Definition 24. A network is called a *Class A* if: 1) there exists a sequence of enhancers or inhibitors $(v_{k_1}, \ldots, v_{k_m})$ such that v_{k_i} is strictly upstream of v_{k_j} for all $i < j$; 2) all enhancers (and inhibitors) are strictly upstream of nodes connected to hyperedges that they enhance (inhibit). Precisely, part 2. means if there exists e/i-uberedge $u = (v_j, h^*)$ with $h^* = (X, Y)$ then v_j must be strictly upstream of every node $v \in X \cup Y$.
Networks that are not Class A are called *Class B* networks.

The conditions of Class A allow the equilibrium state of $v_1 \in I$ to be determined based on the inflow and outflow for v_1. The magnitude of the edge affected by v_1 at equilibrium state can then be calculated. This will allow v_1 to be determined at equilibrium. Class A networks allow the equilibrium states of all enhancers or inhibitors to be calculated this way, which can determine all fluxes (even those affected by uberedges) in the system. Once all fluxes are determined, then the steady state of every node in the network can be calculated for this a unique steady state. This is shown in next:

Proposition 37. *For a Class A network with a given flux, if there exists an equilibrium with positive entries, and all $v \in V$ connect to J, then the equilibrium is unique.*

Proof. Let G be a Class A metabolic graph. Then there exists a sequence of n enhancers or inhibitors $(v_{k_1}, \ldots, v_{k_m}) \subset V$ with v_{k_i} is strictly upstream of $v_{k_{i+1}}$, and therefore all nodes upstream of v_{k_1} are unaffected by enhanced edges. It follows that x_{k_1} has a unique steady state determined by fluxes corresponding to incoming and outgoing edges. According to $(H2)$:

$$\dot{x}_{k_1} = \sum_{h \in \Gamma^{in}(v_{k_1})} \Psi_h^{\text{in}}(v_{k_1}) \mathbf{F}_h(x) f_h - \sum_{h \in \Gamma^{out}(v_{k_1})} \Psi_h^{\text{out}}(v_{k_1}) \mathbf{F}_h(x) f_h. \tag{4.11}$$

where we used the notation of Definition 17. If for all $(X, Y) \in \Gamma^{in}$ we have $X \subset I$ then for all $v_i \in X$ we can write

$\dot{x}_i = f_{j0} - \sum_{h \in \Gamma^{out}(v_{k_1})} \Psi_h^{\text{out}}(v_i) \mathbf{F}_h(x) f_h$. This determines the steady state for x_i, hence all steady state quantities in equation (4.11) are known and the unique steady state of x_{k_1} is determined. The same can be obtained by recursion if the nodes v_i are connected to I by some path with hyperedges not affected by enhancers or inhibitors, which is always the case because v_{k_1} is the first node acting as enhancer or inhibitor. Finally the value of x_{k_1} is determined.

Similarly, v_{k_2} may be the terminal node, or initial node for some edges that are not affected by inhibitors or enhancers except possibly v_{k_1}, for which the value of x_{k_1} has been determined. Denote the steady state value of x_{k_1} as $\bar{x}_{k_1}$ According to $(H2)$, either

$\dot{x}_{k_2} = \sum_{h \in \Gamma^{in}(v_{k_2})} \Psi_h^{\text{in}}(v_{k_2}) \mathbf{F}_h(x) f_h - \sum_{h \in \Gamma^{out}(v_{k_2})} \Psi_h^{\text{out}}(v_{k_2}) \mathbf{F}_h(x) f_h$ or

$\dot{x}_{k_2} = \sum_{h \in \Gamma^{in}(v_{k_2})} \Psi_h^{\text{in}}(v_{k_2}) \mathbf{F}_h(x) K(\bar{x}_{k_1}) f_h - \sum_{h \in \Gamma^{out}(v_{k_2})} \Psi_h^{\text{out}}(v_{k_2}) \mathbf{F}_h(x) K(\bar{x}_{k_1}) f_h$, with enhancer terms $K(\bar{x}_{k_1})$ inserted. Therefore the value of x_{k_2} is also determined.

By recursion, we can determine the value of all enhancers and inhibitors nodes, and therefore, all enhanced or inhibited fluxes are determined.Finally we can determine uniquely the value of all metabolites and fluxes. □

Class B networks are not expected to have a unique equilibrium. For instance the

network of Figure 4.5 bottom with $F_{i,h}(x_i) = \frac{x_i}{1+x_i}$ and $K_{i,h}(x_i) = \frac{1}{1+x_i}$ has two possible equilibria with values for x_2 given by $\frac{f_3\, x_1}{f_1\,(1+x_1)} - 1$ and $\frac{f_2\,(1+x_1)}{f_4 - f_2\,(1+x_1)}$. Choosing $f_{\{i=1,\ldots,7\}} = \{1, 1, 12, 2, 2, 2, 2\}$ the two equilibria have positive entries: $\tilde{x} = \{\frac{1}{2}, 3, 1, 1, 1\}$ and $\bar{x} = \{\frac{1}{3}, 2, 1, 1, 1\}$, see Fig 4.6.

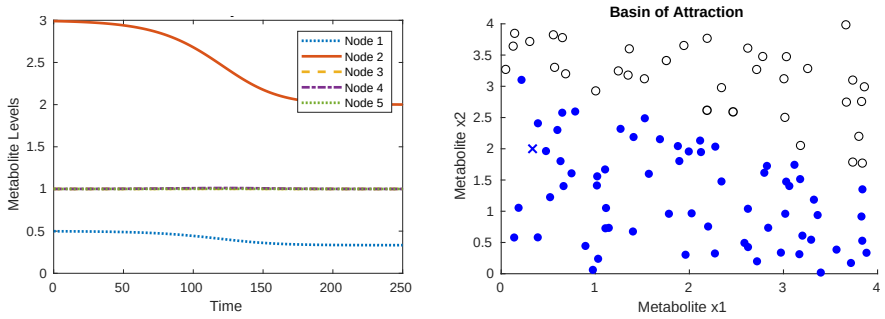


Figure 4.6: Simulations for the class B system in Fig. 4.5. Top: Starting near the unstable equilibrium $\tilde{x}$, the system tends to the stable one $\bar{x}$. Bottom: Basin of attraction to $\bar{x}$ in $x_1 - x_2$ space (marked with "X") with filled blue circles for convergent initial data and hollow black circles for divergent ones.

4.5 Modeling Tuberculosis with metabolic graphs

Here we focus on the synthesis of antibiotic network of MTB. Such a network is shown in Figure 4.7 including the action of a set of antibiotics commonly used in cure of tuberculosis. The action of drugs on edges is solely due to regulation of the enzyme for the corresponding reaction. We see that a group of three drugs regulates an enzyme for a reaction from node v_1 to node v_{10}, $f_{(v_1,v_{10})}$. In this case treatment with any single drug (Amikacin, Capreomycin, or Clofazimine) will effect a single gene called "Rv1475c." Amikacin and Capreomycin will upregulate the gene, and Clofazimine will down regulate the gene. For edge $f_{(v_3,v_2)}$, Clofazimine downregulates two genes, both acting as enzymes for the corresponding reaction. For the excretion from node v_2 out of the subnetwork, four drugs (Amikacin, Clofazimine, Ethionamide, and Isoniasid) all upregulate the gene Rv0211. The combination of downregulating $f_{(v_3,v_2)}$, and upregulating the excretion from v_2 will serve to lower the metabolite represented by

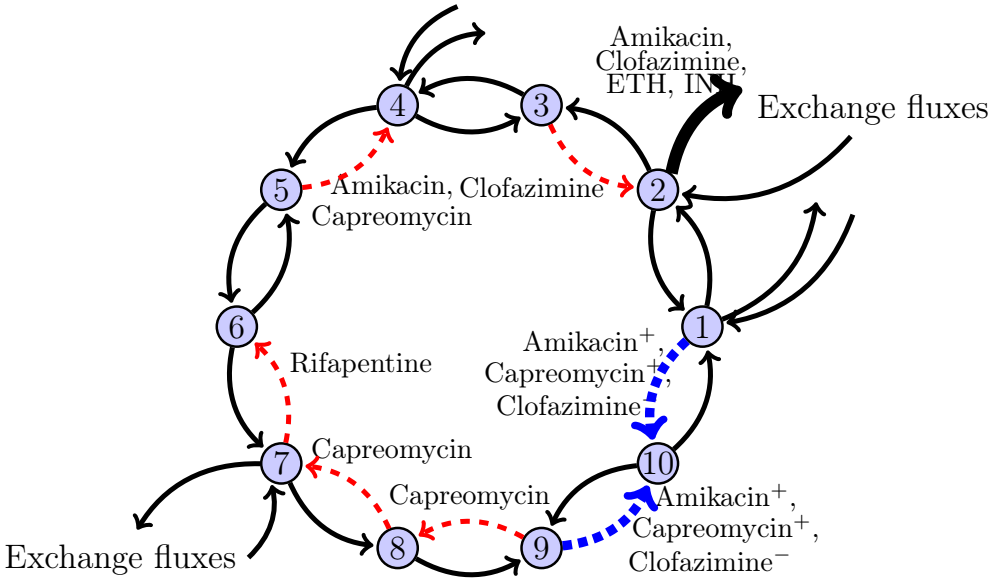


Figure 4.7: A subnetwork of the MTB synthesis of antibiotics. The drugs written show how they affect reactions (+ enhances the enzyme, - inhibits)

v_2, namely Oxalacetic acid (chemical formula: C4H4O5). Figure 4.8 compares two regimens with Rifapentine, showing more effectiveness of thee daily versus the weekly treatment on the biosynthesis of antibiotic network.

The dynamics of the biosynthesis of antibiotics network (Fig. 4.7) can be written as (1.3), where $S : \mathbb{R} \to M_{10\times 28}$ is a sparse matrix.

$$\begin{pmatrix}
-x_1 & x_{10} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -x_1 & C_1 & -x_1 & x_2 & 0 & 0 & 0 & 0 \\
0 & 0 & x_3 & -x_2 & -x_2 & C_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_1 & -x_2 & 0 & 0 & 0 & 0 \\
0 & 0 & -x_3 & x_2 & 0 & 0 & 0 & 0 & 0 & -x_3 & x_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_3 & -x_4 & 0 & 0 & 0 & 0 & x_5 & -x_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & C_3 & -x_4 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -x_5 & x_6 & -x_5 & x_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_7 & -x_6 & x_5 & -x_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & x_8 & -x_7 & 0 & 0 & 0 & 0 & -x_7 & x_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & C_4 & -x_7 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -x_8 & x_7 & x_9 & -x_8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -x_9 & x_8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -x_9 & x_{10} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
x_1 & -x_{10} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_9 & -x_{10} & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix} \tag{4.12}$$

We show a simulation of the MTB biosynthesis of antibiotics network (Figure 4.7) to illustrate what can happen. Figure 4.8 will converge to different periodic solutions based on the dosage chosen. The left hand side of Figure 4.8 shows the evolution with an hourly treatment of Rifapentine, and the right hand side shows the evolution with a weekly treatment. Note that most metabolites in the system converge to the

same equilibrium value, however the metabolite labeled “node 6” has oscillations that are centered around a very different value depending on dosage.

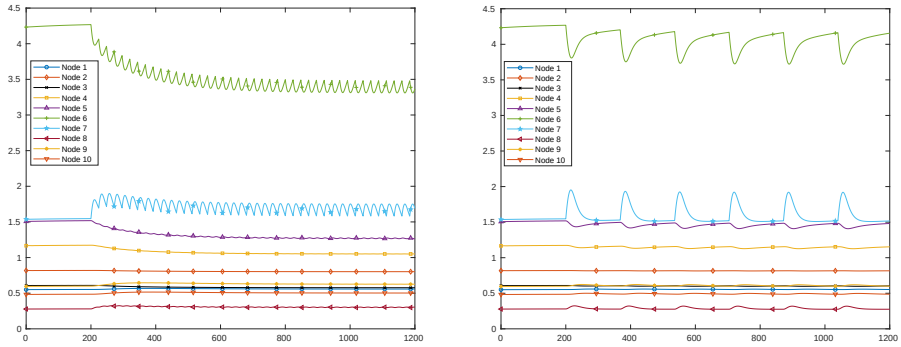


Figure 4.8: Evolution of MTB biosynthesis of antibiotics. The first 200 hours are without drug, then results with (left) daily one hour treatment with Rifapentine and (right) weekly treatment at a higher dose.

4.6 Appendix: Definitions for general metabolic ubergraphs

For completeness, this section constructs a more general version of a directed hypergraph, weighted directed hypergraph, and metabolic graph by building upon the classical notions of a graph and hypergraph. For the following definitions, V is a set of nodes, and $\mathcal{P}(\mathcal{S})$ is the power set of a set $\mathcal{S}$.

Definition 25. A *directed graph* is a pair $G = (V, E)$ where E is the set of ordered pairs such as (v_i, v_j) with $v_i, v_j \in V$ which indicates a directed edge, starting from the initial node v_i and pointing to the terminal node v_j.

Definition 26. A weighted directed hyperedge is a couple $\mathcal{H} \ni h = (X, Y)$ with $X \in \mathcal{P}(V)$ or $X = \{v_0\}$, $Y \in \mathcal{P}(V \cup \{v_{n+1}\})$, and corresponding weights $\Psi_h : h \mapsto (\Psi_h^{\text{out}}, \Psi_h^{\text{in}})$ where $\Psi_h^{\text{out}} : X \mapsto \mathbb{R}_+$ and $\Psi_h^{\text{in}} : Y \mapsto \mathbb{R}_+$.

Definition 27. A weighted directed hypergraph G is an ordered triplet $G = (V, \mathcal{H}, \Psi_{\mathcal{H}})$ where $\Psi_{\mathcal{H}}$ denotes the set of functions assigning weights to hyperedges, i.e. $\Psi_{\mathcal{H}} = \{\Psi_h : h \in \mathcal{H}\}$.

As shown in Section 4.3, weighted directed hypergraphs are the right mathematical object to describe biochemical reactions involving multiple compounds. Interestingly, while the theory of (not directed) hypergraphs seems well developed [18, 138], directed hypergraphs have been less explored in the literature.

We are now ready to introduce a general definition of ubergraphs. The concept was introduced in [68], but the used definition appears to allow some pathological examples, thus we provide an alternative definition below.
First, given a finite set of nodes $V = \{v_1, \ldots, v_n\}$ define recursively P_k as:

$$P_k = \mathcal{P}\left(\bigcup_{i=0}^{k-1} P_i\right) \setminus \{\emptyset\}, \quad P_0 = V \tag{4.13}$$

Definition 28. A depth-k ubergraph U is a $k+1$-tuple $G = (U_0 = V, U_1, \ldots, U_k)$ where $U_i \subseteq P_i$ is a finite set of uberedges and $U_i \subset \mathcal{P}\left(\bigcup_{i=0}^{k-1} U_i\right) \setminus \{\emptyset\}$. We call an element of U_i a depth-i uberedge.

Let us provide some details on such definition. First notice that a depth-1 Ubergraph is a classical hypergraph. Indeed in this case $P_1 = \mathcal{P}(V) \setminus \{\emptyset\}$ which is precisely the definition of the set of hyperedges.

In [68] the definition was slightly different allowing $\emptyset$ as an hyperedge. However, this give rise to the problem of the meaning in models of an empty heperedge. The problem becomes more dramatic for general depth k ubergraphs. Another difference is that we force U_i to be subset of $\mathcal{P}\left(\bigcup_{i=0}^{k-1} U_i\right) \setminus \{\emptyset\}$. If this is not the case, then we may have depth-k uberedges which connect uberedges of lower depths, without them being included in the ubergraph, see the following discussion for depth-2 ubergraphs. Depth-2 ubergraphs contains more general uberedges than hypergraphs. More precisely, $G = (V, U_1, U_2)$, where U_1 is a set of hyperedges and $U_2 \subset \mathcal{P}(V \cap U_1) \setminus \{\emptyset\}$. Therefore $u \in U_2$ is a set $\{v_{k_1}, \ldots, v_{k_m}, h_{i_1,}, \ldots, h_{i_p}\}$ where $v_{k_l} \in V$ and $h_{i_l} \in U_1$. A particular case would be an uberedge $\{v, h\}$, $v \in V$, $h \in U_2$, which could model an enhancer or inhibitor action of metabolite x_v on the reaction represented by the hyperedge h. Let us remark again the importance of restricting the condition $U_i \subset P_i$ to the more stringent we use. If we impose only $U_2 \subset P_2$ then we could allow uberges of the type $\{v, h\}$ with $h \notin U_1$, thus allowing metabolites to affect reactions which are not in the model.

Example 7. Given a set of three nodes $V = \{v_1, v_2, v_3\}$ we want to explore depth-2 ubergraphs. First $P_1 = \mathcal{P}(V) \setminus \{\emptyset\}$ thus

$$P_1 = \{\{v_1\}, \{v_2\}, \{v_3\}, \{v_1, v_2\}, \{v_1, v_3\}, \{v_2, v_3\}, \{v_1, v_2, v_3\}\}.$$

In other words P_1 contains all loops around nodes v_i, regular edges and the only

possible (strict) hyperedge $\{v_1, v_2, v_3\}$. U_1 is any subset of P_1. The cardinality of P_1 is $2^3 - 1 = 7$, since it is the power set of V excluding the empty set.

$P_2 = \mathcal{P}(V \cup P_1) \setminus \{\emptyset\}$ contains all ubereges $\{v_{k_1}, \ldots, v_{k_m}, h_{i_1,}, \ldots, h_{i_p}\}$, with $0 \leq m \leq 3$, $0 \leq p \leq 3$, $m + p \geq 1$, $v_{k_l} \in V$ and $h_{i_l} \in P_1$. P_2, being a power set excluding the empty set, has cardinality $2^{3+2^3-1} - 1 = 2^{10} - 1 = 1023$. Notice that there are objects not easily interpreted in P_2, for instance the uberedge $\{\{v_1\}\}$, that is the uberedge which loops over the loop over v_1.

Finally, the possible ubergraphs on three nodes are given by a triplet V, U_1, U_2. The possible choices of U_1 are among all subsets of P_1 thus we do have $2^7 = 128$. The possible choices for U_2, not taking into account U_1, are given by $2^{|P_2|} = 2^{1023}$ (where we used $|\cdot|$ to indicate the cardinality of a set). However, as explained above, the admissible U_2 may be chosen only as subsets of $\mathcal{P}(V \cup U_1) \setminus \{\emptyset\}$ thus the possible choices are $2^{2^{3+|U_1|}-1}$.

As shown above, the set of ubergraphs can be extremely complex and of high cardinality. Moreover, not all uberedges can be easily interpreted. Thus on one side we would like to restrict the definition to allows only objects with modeling meaning, but on the other side a further generalization is necessary to allow directed hper and uberedges, weights and signs. This is achieved by next definition.

Definition 29. A depth-k metabolic graph is an ubergraph with (nontrivial) directed weighted hyperedges and (nontrivial) directed signed uberedges of depth greater than or equal to 2. More precisely it is a $k+3$-tuple $G = (V, \mathcal{H}, \Psi_{\mathcal{H}}, U_2, \ldots, U_k, \Psi_U)$ such that the following holds: V is a finite set; $\mathcal{H}$ is the set of directed hyperedges (X, Y), $X, Y \subset V$, $X, Y \neq \emptyset$; $\Psi_{\mathcal{H}} = \{\Psi_h^{out}, \Psi_h^{in}\}_{h \in H}$, $\Psi_h^{out} : X \to \mathbb{R}_+$, $\Psi_h^{in} : Y \to \mathbb{R}_+$, is the set of weights; for $j = 2, \ldots, k$, $U_j = (X, Y)$ with $X, Y \subset \mathcal{P}\left(\bigcup_{i=0}^{j-1} U_i\right) \setminus \{\emptyset\}$, $X, Y \neq \emptyset$; $\Psi_U = \{\Psi^j\}_{2 \leq j \leq k}$ with $\Psi^j : U_j \to \{-1, +1\}$.

Example 8. A depth-2 metabolic graph includes directed weighted hyperedges and signed uberedges of the type (X, Y) with X, Y containing nodes and hyperedges.

Such a type of uberedge may represent the enhancer or inhibitor action of a set of metabolites and reactions (those in X) over another set of metabolites and reactions (those in Y).

A depth-3 metabolic graph may also include the action of a depth-2 uberedge over another set of depth-2 ubderedges. This type of edges may represent the effect of a part of a metabolic network over another one by molecular mechanisms which are not known.

Chapter 5

Testing four-drug combinations of tuberculosis treatment using microarray data and simulations via LIFE.

5.1 Introduction

The bacterium Mycobacterium tuberculosis (brifly MTB) infects a large portion of the population. The World Health Organization (briefly WHO) estimates that in 1999 there were 8.4 million new cases, up from 8 million in 1997. More recently, the WHO also produced a list of prioritization of pathogens to guide drug discovery which focuses on the risks the population faces from drug resistent bacterial infections with an emphasis on TB [105].

Patients with TB are usually prescribed multiple antibiotics simultaneously. The standard treatment for tuberculosis (briefly TB) commonly combines up to four drugs. There are many antibiotics to choose from, and new drugs are currently being de-

veloped. Determining the optimal set of dosages and drugs is costly. To test all possibilities would require too much time and resources. Therefore, techniques to estimate the potential of a drug combination with minimal time and effort are valuable in order to indicate which combinations are most justified for further investigation. Such techniques as diagonal measurement of n-way drug interactions (briefly DiaMOND) [28] which serves to expedite the information given by a checkerboard assay, Response Surface Methodology (briefly RSM) [76] which is a set of mathematical and statistical tools that can be used to determine functional relationships between drugs, and Central Composite Design (Briefly CCD) [8] which is used to extrapolate chemical properties from a small number of precisely chosen experiments.

MTB has a refined gene regulatory network that allows for changes in the state of its metabolism as a consequence of environmental or internal chemical conditions [51, 109, 10]. This dynamic metabolic network causes challenges in modeling MTB metabolism which leads to less predictable drug effects, impeding the development of new TB treatment. Another aspect of TB infection that inhibits attempts to treat infection is the formation of a caseum, a spheroid structure of necrotic cells that drugs only partially penetrate [123, 127]. This work aims to extend models for drug interactions by capturing the effect on more complex chemical reactions described by hyperedges (reactions between more than two metabolites). Our main results are 1.ranking potential TB treatment according to microarray data of individual drugs on MTB gene expression, and 2.to validate our techniques for modeling chemical reactions as hyperedges and drug interactions.

The remainder of this paper is organized as follows: we characterize the microarray data set that we use to perform analyses, and explain how we filter out significant drug effects from the data. Approaches of quantifying the drug action on a network are introduced, and 4-drug combinations are ranked accordingly. Next we explain how chemical reactions between more than two metabolites are implemented. Then,

the target reactions of the metabolic network per drug are shown. Moreover, the dynamics of these reactions in the presence of drug is defined. Using pharmacokinetic data to guide our choice of simulation parameters, example simulations are presented and discussed.

5.2 Methods

5.2.1 Microarray Datasets

Microarray data showing the gene expression levels of MTB (strain H37Rv) genes under different drug treatments was collected from NCBI's Gene Expression Omnibus (GEO) [43]. This data was contributed from the work [16] and is a transcriptional profiling of MTB performed using 430 whole-genome microarrays to measure the effects of 75 different drugs, drug combinations, or different growth conditions at various times relative to a sample of logarithmically growing MTB. This data was downloaded from GEO accession GSE1642.

The microarray gene expression data found in [16] represents the response of MTB genes to treatment of a single drug after 6 hours. The value given in the dataset represents the log base 2 transform of the gene expression values. For each trial there are 4320 individual wells which contain sequences encoding genes or hypothetical proteins.

Our focus is on 13 of the 75 reported drugs. These 13 drugs were chosen based on begin part of PK studies of interest as well as being part of clinical trials. []. Additionally each of the MTB microarray samples chosen to analyze were taken after 6 hours of drug exposure. Eleven of these have PK studies which are available, specifically, Amikacin, Capreomycin, Clofazimine, Ethambutol (EMB), Ethionamide (ETH), Isoniazid (INH), Levofloxacin, Pyrazinamide (PZA), Pretomanid, Rifapentine, and Rifampicin. The remaining two drugs, Streptomycin and Ofloxacin were

chosen because of their frequent use in clinical studies. For each drug there are at least two trials performed. Many of the drugs also have trials at 2 separate doses. Complete information about the drug trials and dosages can be seen in Table 5.1.

Drug	MIC	Dose 1	# of trials	Dose 2	# of trials	Dose 3	# of trials
Amikacin	1ug/mL	5ug/mL	2	10ug/mL	1		
Capreomycin	50ug/ml	5ug/mL	2	10ug/mL	2		
Clofazimine	1.25ug/ml	10ug/mL	1	13ug/mL	2		
EMB	0.6ug/ml	10ug/mL	2				
ETH	0.5ug/ml	12ug/mL	2	40ug/mL	2		
INH	0.02ug/ml	0.2ug/mL	1	0.4ug/mL	2		
Levofloxacin	1ug/ml	10ug/mL	2				
PZA	10ug/ml	0.12mg/mL	2	1.2mg/mL	2		
Pretomanid	0.4ug/ml	0.2ug/mL	2	0.4ug/mL	4	2ug/mL	2
Rifapentine	N.D.	0.1ug/mL	2	0.5ug/mL	2		
Rifampicin	0.4ug/ml	0.2ug/mL	2				
Streptomycin	1ug/ml	2ug/mL	2	5ug/mL	2		
Ofloxacin	1ug/ml	5ug/mL	2	10ug/mL	2		
Capreomycin	2mg/mL	4mg/mL	1	8mg/mL	1		
ETH	1mg/mL	2mg/mL	1	4mg/mL	1		
INH	1.5ug/mL	3ug/mL	2	12ug/mL	1	24ug/mL	1
Moxifloxacin	0.4ug/mL	0.8ug/mL	1	1.6ug/mL	1	3.2ug/mL	1
Pretomanid	0.15ug/mL	0.3ug/mL	1	0.6ug/mL	1	1.2ug/mL	1
Rifampicin	0.1ug/mL	0.2ug/mL	1	0.4ug/mL	1		
Streptomycin	1mg/mL	2mg/mL	2	4mg/mL	2	8mg/mL	1

Table 5.1: Table detailing dosage information and number of trials for various drug treatments. The rows in white correspond with 6 hour trials from [16] while the shaded rows in cyan correspond with 16 hour trials from [91]. The MIC given is the minimum inhibition concentration reported in the paper [16]or [91].

Another source of microarray data was found in [91], downloaded from GEO accession GSE71200. In this data bacteria were expossed for 16 hours to concentrations of each drug, at varying multiples of the drug's minimum inhibitory concentration. This dataset considered 9 drugs, 6 of which were also considered in [16]. These 6 drugs (Capreomycin, ETH, INH, Pretomanid, Rifampicin, Streptomycin) in addition to Moxifloxacin were analyzed. Although there is a large overlap in genes analyzed, the two datasets are not identical even when the same drug is considered. For each drug trial in [91], there are 5250 wells, as oppposed to only 4320 in [16]. Because of the differences in process of the two datasets, they are not directly comparable. Therefore, we show our results of these datasets independent of each other throughout

this work.

5.2.2 Filtering of Data

To analyze the consistency of a given drug's effect on a gene in the microarray data, the variability between different trials of the same drug were measured using the L_1 norm. Taking the L_1 norm between intra-drug trials confirmed that the microarray data contains a lot of variability. Several filtering methods were constructed in order to identify genes that had both a consistent and significant response to drug trials which includes the Significance Analysys of Microarray (briefly SAM) technique for significance [136]. The SAM technique seeks to further limit one's false discover rate for significant genes.

In order to determine significance, a threshold of ± 1 was chosen. This threshold corresponds with a gene having two times (or one half) its expression when untreated. One method to consider whether a gene has significant response to a given drug is if there was a consensus in every trial that the gene response was greater than the threshold. Trials for which there was no data were ignored when considering the consensus of trials for a drug, however if the trials lacking data for a gene were greater than 50% of the trials with that drug, the gene was not considered significant despite the remaining trials having a consensus above ± 1. For the data from [16], out of the 4320 microarray wells, there were 1165 that had a consensus of trials greater than ± 1 for at least one drug.

Significant genes could alternately be determined by the average response to a drug. Similar to the consensus fold change, a threshold of ± 1 was used such that if a the gene responses to trials of one drug had a mean response greater than the threshold it is considered significant.

5.2.3 Significance Analysis of Microarrays

Another method to determine the consistencey and significance of a given drug's effect is a method called Significance Analysis of Microarrays (briefly SAM) [136]. This method uses the variation of a single gene between multiple trials to correct a significance estimate. Thus, the method lowers the false discovery rate while maintaining a comparable level of genes identified as significant to the fold change method.

A brief look at the number of significant genes found under each method is shown in Table 5.2. Differences in the methods are readily apparent, in general the Consensus Significance is most restrictive with SAM having the highest number of significant genes. The notable exceptions to this rule are drugs which had only two trials; Rifampicin, EMB, and Levofloxacin. We chose to use SAM method to determine significant genes, except for drugs which only had two trials, for which consensus would be used. Among the SAM and Consensus significant genes there were several cases for which the average mean fold change was near zero, despite being classified as significant from the other methods. This is particularly concerning when using the Consensus method, as it suggests that all of the trials were outside the ± 1, however, some trials were < -1 and others $> +1$. For the SAM method it may suggest that although the gene had greater than expected expression, the overall effect was still small. Because of this only genes which had a mean effect outside the threshold in and were considered significant in another method were classified significant. The gene sets used for each drug are shown in red in Table 5.2. For the remaining of the paper when the term significant genes is used, it will refer to this set unless otherwise stated.

Figure 5.1 shows the significant genes for these 13 drugs. For the data from [91], there were 918 out of 5250 wells that were significant for at least one drug. Figure 5.2 shows the significant genes for these seven drugs.

Drug	Consensus	Mean	SAM	Consensus & Mean	SAM & Mean
PZA	43	160	647	43	47
EMB	48	63	6	46	0
Rifampicin	659	959	176	650	75
ETH	48	94	606	48	35
Amikacin	241	481	916	240	168
INH	65	105	625	64	36
Capreomycin	224	603	2236	224	361
Levofloxacin	68	109	2	68	0
Rifapentine	24	157	372	23	26
Clofazimine	251	388	858	251	133
Pretomanid	95	712	2485	95	499
Ofloxacin	72	129	1313	72	69
Streptomycin	23	187	1134	23	80

Table 5.2: A summary of the number of significant genes identified by various methods is shown for a given drug. The highlighted cells indicate which method was chosen for use in simulations.

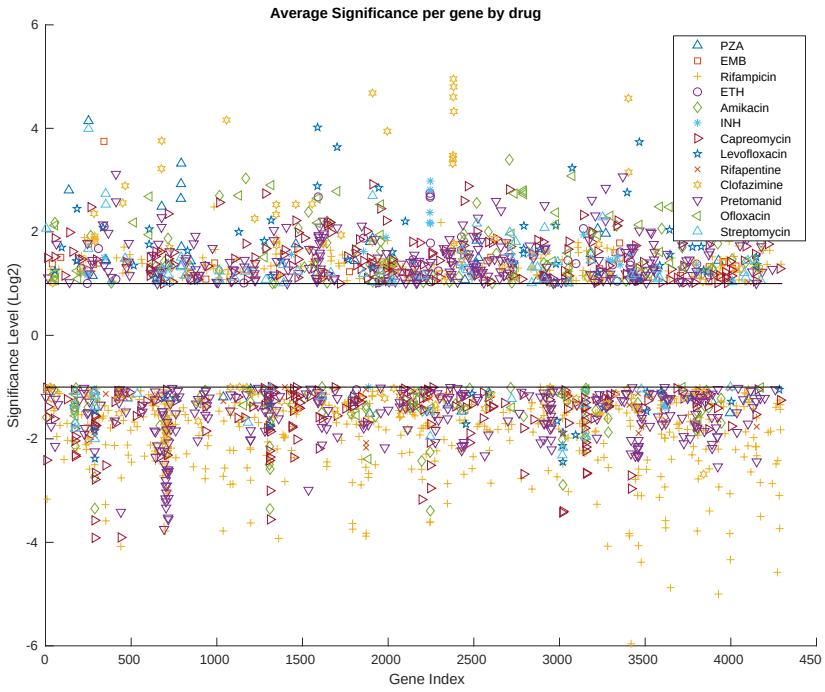


Figure 5.1: Plot of average value of significant gene expressions for the 13 drugs from [16]. The colors/shapes indicate different drugs.

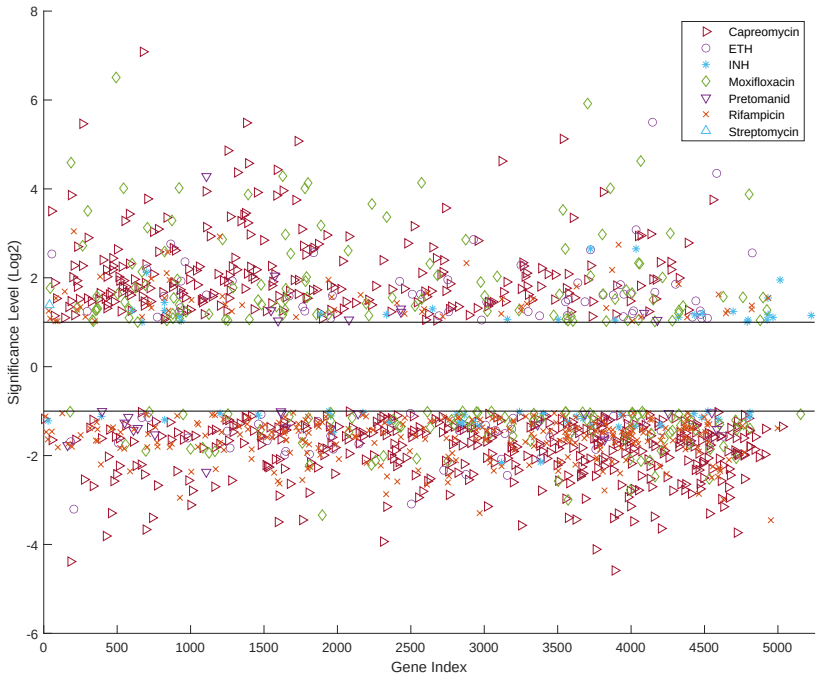


Figure 5.2: Plot of average value of significant gene expressions for the 7 drugs in the data from [91]. The colors/shapes indicate different drugs.

The simulations that we show later require a determination of which genes will be actually affected by treatment, and we use the fourth method in the table: "mean fold change > 1 & SAM." EMB, Rifampicin, and levofloxacin have two microarray experiments(because there are only two data points per drug per gene, SAM identifies very few genes as significant).

The microarray data we analyze has a different number of microarray experiments for each drug, shown in Table 5.1. The boshoff data includes three trials for Amikacin (two low dose experiments, and one high), INH, and Clofazimine (two high dose experiments, and one low).

All other drugs have at least two trials for each of two doses, additionally Pretomanid has three doses (two trials for the low dose, four for medium, and two for high).

5.2.4 Pathway Analysis

Using Kyoto Encyclopedia of Genes and Genomes (KEGG) [73, 70, 72] allows the linking between the genes and MTB metabolic pathways. The metabolic pathways in KEGG are networks which can be represented by directed graphs, where the nodes represent metabolites and the edges represent metabolic reactions. This was used to identify which significant genes directly affect enzymes involved in reactions in MTB metabolic pathways.

There are many genes which are not directly associated with any pathway are likely not involved with enzyme creation or protein translation, but other processes such as gene regulation. Genes may also be associated to a particular pathway, but the association has yet to be determined, or is unknown to KEGG. Another case is that of RNA sequences that code hypothetical proteins in MTB: of the 1165 RNA sequences in the 6 hour drug tests only 1063 correspond with known MTB genes. Similarly only 792 of the 918 RNA sequences of the 16 hour drug tests were associated with known genes. Of these genes, only 398 (273) of them from [16] (from [91]) data were associated with a pathway in KEGG, these genes mapped to pathways is shown in Figure 5.3 and 5.4. Additional work, perhaps analyzing the gene regulatory network, is required to more precisely determine the function of genes not involved with pathways.

Analyzing significantly affected genes and their corresponding pathways revealed which pathways were targets regarding treatment. These pathways could be governing mechanisms restricted by the drug, as well as pathways detailing the response of MTB to the drugs.

5.2.5 Simulations

Linear-In-Flux-Expression is an extention of flux balance analysis which allows non-linearities in metabolites while preserving linearity with respect to the flux. Dynamics

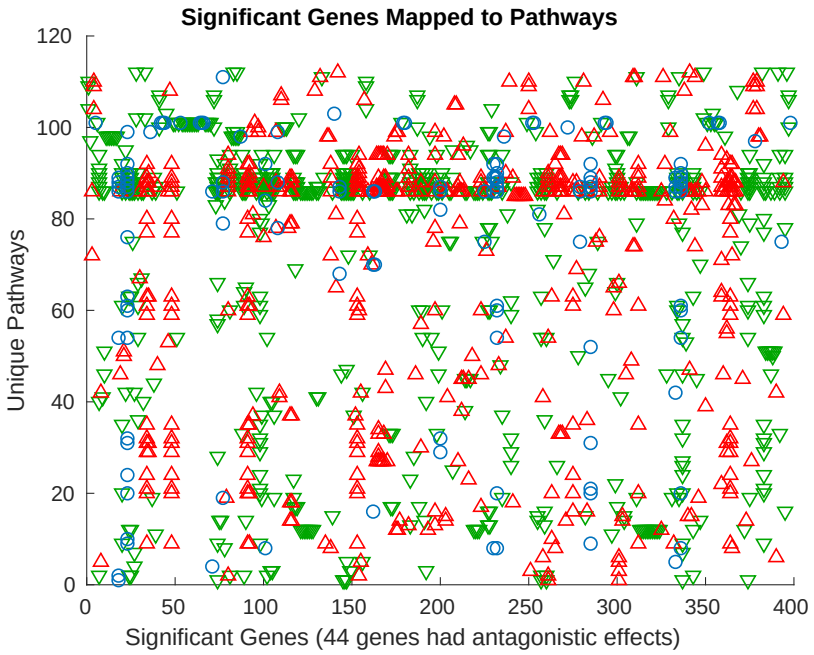


Figure 5.3: This image shows the mapping of the 398 genes to the metabolic pathways. Downregulated genes are shown in green downward facing triangles, upregulated genes are shown in red upward facing triangles and genes that were upregulated by some drugs and downregulated by others are shown as blue circles.

for a typical flux balance analysis system are defined as:

$$\dot{x} = S(f)\dot{x}. \tag{5.1}$$

where $x \in \mathbb{R}^n$ is the vector of metabolite levels, $f \in \mathbb{R}^m$ is the vector of fluxes and S is an $n \times n$ matrix called the stoichiometric matrix. Eq (5.2.5) is linear in x. Commonly the dynamics of a biochemical system are non linear with respect to x, such as Michaelis-Menten type kinetics, but the linearity with respect to f is essentially part of the definition of *flux*.

LIFE, and other methods such as flux balance analysis rely on representing the metabolic network as a directed graph, where edges represent biochemical reaction. Recently in [98] features were added to LIFE systems in order for the model to be more representative of notable biological features.

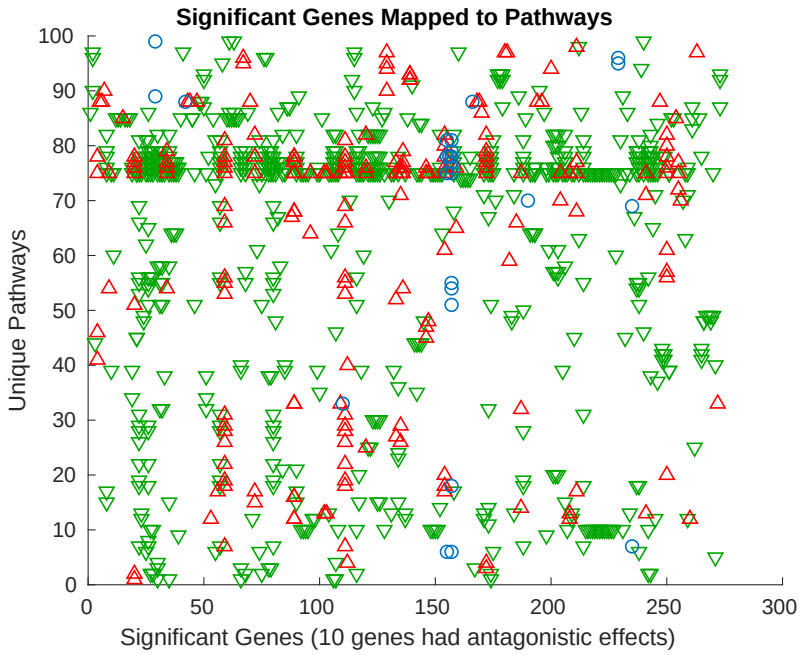


Figure 5.4: This image shows the mapping of the 273 genes to the metabolic pathways. Downregulated genes are shown in red downward facing triangles, upregulated genes are shown in green upward facing triangles and genes that were upregulated by some drugs and downregulated by others are shown as blue circles.

1. Most networks include inflows and outflows (also called intakes and excretions in LIFE methodology) to the external environment or to other networks. Virtual nodes to represent such flow must be included or, alternatively, one must include directed edges with a node only on one end.

2. Some biochemical reactions necessarily involve more than two metabolites, e.g. when two or more compounds interact to form a set of other compounds. Therefore edges with multiple entering and exiting nodes must be included.

3. The action of enzymes, genes and drugs often times affect a specific reaction acting as enhancer or inhibitors. Such actions can be represented by edges joining a node to another edge.

Features 2 and 3 are represented in Figure 5.5.

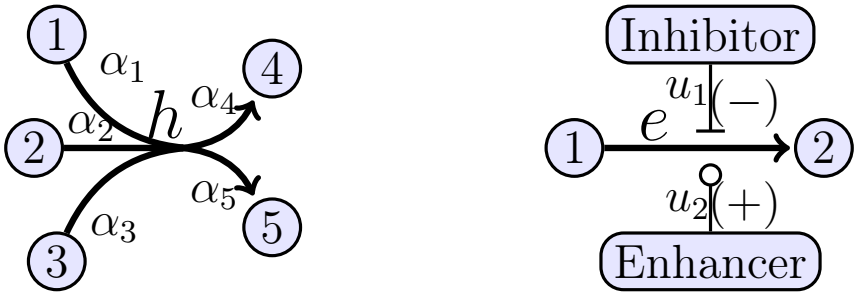


Figure 5.5: The dipicted features were added to LIFE dynamics in order to include important features os biochemical reactions. Left: A hyperedge h which connects three reactants to two products in a metabolic reaction. Each reactant and product has a weight corresponding to the stoichiometry of the reaction. Right: An enhancer(inhibitor) molecule promotes(inhibits) a chemical reaction. The edges u_1, u_2 are called "uberedges" and connect a metabolite or drug to an edge e.

5.2.6 Drug Dosage

Dosage information from a number of papers was used [95, 135, 131, 111, 46, 139, 104, 9, 146, 54]. These studies listed the halflife of the drugs in the bloodstream, as well as the maximum amount found after a dose was given. When two studies gave differing values, we used the average of the two values. Several of these maximum concentrations were less than the dosages given in [16]. In each of these cases the maximum concentration was still larger than the minimum inhibition concentration reported in [16], we also note that for these drugs the dosage in [16] was much larger than the minimum inhibition concentration. Because of this we assume that if MTB was treated with the concentration of drug in the bloodstream it would have a similar or greater effect than the concentrations used by [16]. Another issue with the dosages found in the bloodstream and the dosages used in [16] is that it is not always clear how much drug is able to penetrate into the MTB granulomas. Many models have been created to address this [83, 29, 33, 118, 130, 90, 139], however, in our simulations it is assumed that the entire concentration in the bloodstream will reach MTB. We recognize that the amount of drug in the bloodstream and the amount which reaches

MTB are different, but leave this for further exploration.

5.2.7 Central Carbon Metabolism in *Mycobacterium tuberculosis*

We utilize an example metabolic network from [128] to test our implementation of hyperedges, and drug effects and compare with four drug rankings from section 5.3.2.

Recall the Metabolic network shown in Figure 4.2. This is the central carbon metabolism of MTB derived from an infection in mouse. The network shown is a metabolic network of MTB in a state of growing bacilli. This network depicts carbon from lipid and sugar catabolism used by MTB for generating energy and biosynthetic precurors required for growth. This network is of particular interest due the switching that occurs whether MTB is in growing or non-growing state. Genes required for using fatty acids as opposed to glucose are upregulated when MTB is in the non-growing state, and the knowledge of how drugs affect this pathway could be beneficial for halting MTB growth.

5.3 Results

5.3.1 Drug Combinations

Similarity Scores

It is important to consider not only the effect of a single drug, but how they work in combination with each other. In cases where two drugs target the same genes, using them in combination would be redundant []. In order to eliminate redundant gene combinations the common significant genes between drugs and quantitative effect on gene expression were analyzed. The common significant genes shows that drugs have similar targets, although the effect may be very different. A table of overlap

percentage is shown in Table 5.3. The total regulation effect gives a score on how similarly the two drugs effect genes. This effect is achieved through the calculation of a normalized L_1 score between the significant genes of both drugs.

	PZA	EMB	Rifampicin	ETH	Amikacin	INH	Capreomycin	Levofloxacin	Rifapentine	Clofazimine	Pretomanid	Ofloxacin	Streptomycin
PZA	1.00	0.02	0.17	0	0.29	0	0.24	0.05	0	0.22	0.07	0.05	0.05
EMB	0.02	1.00	0.15	0.23	0.21	0.19	0.13	0	0.02	0.08	0.06	0.02	0.04
Rifampicin	0.01	0.01	1.00	0.03	0.15	0.05	0.12	0.04	0.04	0.09	0.06	0.03	0.02
ETH	0	0.23	0.46	1.00	0.33	0.79	0.25	0.04	0.02	0.21	0.13	0.08	0.02
Amikacin	0.05	0.04	0.41	0.07	1.00	0.10	0.53	0.05	0.03	0.15	0.09	0.05	0.06
INH	0	0.14	0.51	0.58	0.37	1.00	0.20	0.03	0.03	0.14	0.14	0.09	0.02
Capreomycin	0.05	0.03	0.34	0.05	0.57	0.06	1.00	0.09	0.05	0.11	0.09	0.04	0.06
Levofloxacin	0.03	0	0.41	0.03	0.18	0.03	0.29	1.00	0.10	0.19	0.06	0.54	0.07
Rifapentine	0	0.04	0.96	0.04	0.33	0.08	0.46	0.29	1.00	0.17	0.08	0.04	0.29
Clofazimine	0.04	0.02	0.25	0.04	0.14	0.04	0.10	0.05	0.02	1.00	0.17	0.04	0.02
Pretomanid	0.03	0.03	0.38	0.06	0.23	0.10	0.20	0.04	0.02	0.46	1.00	0.02	0.02
Ofloxacin	0.03	0.01	0.28	0.06	0.17	0.08	0.11	0.51	0.01	0.14	0.03	1.00	0.04
Streptomycin	0.09	0.09	0.45	0.05	0.64	0.05	0.64	0.23	0.32	0.23	0.09	0.14	1.00

Table 5.3: Table showing the overlaping percentage of different drugs. Each row shows the percentage of significant genes of that drug with the drugs of the column. The yellow cells highlight more than 50% overlap, while the red cells highlight more than 90% overlap. The green cells highlight the diagonal.

The number of common significant genes were plotted against a normalized L_1 distance between the genes significant for either drug. A high L_1 distance indicates that the two drugs affect genes differently, while a low L_1 distance indicates similar effects. If the L_1 distance is low, and the number of genes in common is high, this may indicate that the drugs have a very similar effect. When considering drug regimens it is important to consider how drugs will work in combination, and drugs that are too similar should not be paired together.

Admissible Combinations

Our criteria for admissible combinations is as follows: when the normalized L_1 distance between two drugs was less than 0.3 and more than 50% of the significant

genes overlapped, drugs were considered to be similar drugs and avoided in drug combination. Upon analysis of the drug data from [16], is was found that two drug combinations, Levofloxacin with Ofloxacin and ETH with INH, met this criteria. Another drug combination Rifapentine with Rifampicin, does not meet the criteria because it does not have a low L_1 distance, but over 95% of the genes affected by Rifapentine were also affected by Rifampicin. This result is partly due to the fact that Rifampicin has many more significant genes than Rifapentine, making it unrealistic to expect a low L_1 distance. Because of the large overlap between these drugs, Rifapentine and Rifampicin were also considered to be similar drugs. A near miss in this criteria that could be considered in the future are the combination of Rifapentine with Levofloxacin, which had an L_1 distance of 0.2628, but share less than 30% of their genes. Another near miss was Amikacin with Capreomycin, which share more than 50% of significant genes but had an L_1 distance of 0.3342.

Table 5.3 provides coverage information, however for a more comprehensive pairwise comparison between drugs, we observe the number of genes significantly affected by the drugs as well as the L_1-score of each drug's affect. We plotted the drugs on axes based on both pieces of information. The quadrants partitioning the plots classify pairs of drugs. For example, in the plot labeled "**ETH**" one can see that ETH appears as a plotted point near (50,0). This reflects that ETH has 50 significant genes in common with itself, and that the L_1 distance $d(\text{ETH}, \text{ETH})$ is 0. It was indicated that ETH and INH would not be prescribed together because they are very similar, and this was corroborated by our similarity analysis; In the plot for **ETH** and **INH** observe that the INH and ETH respectively are in the lower right, indicating that they are similar. This is also clearly visible in Fig 5.6. Because INH and ETH are in the lower right quadrant of the plot, these two drugs affect the same genes significantly and in the same way (inhibition or promotion). In addition to disallowing drugs that were too similar, the drugs Amikacin and Capreomycin were removed be-

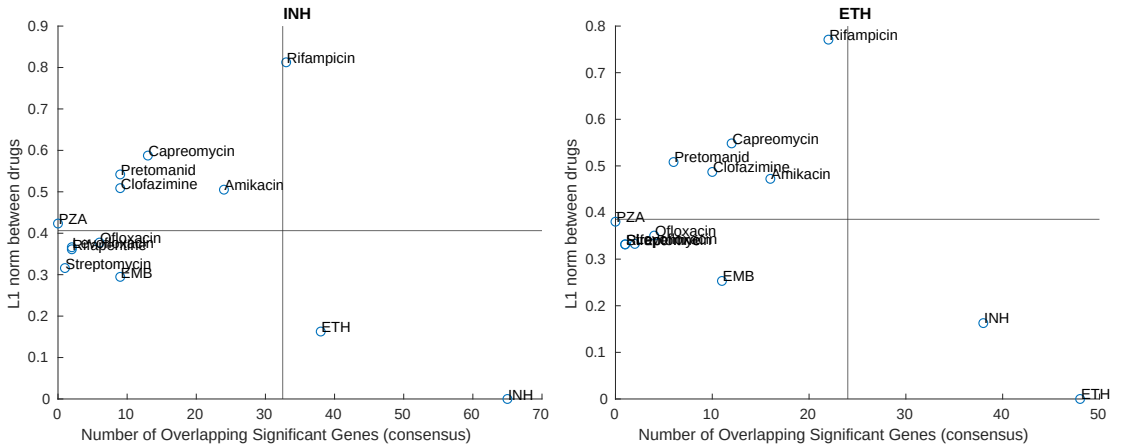


Figure 5.6: This plot shows the number of common significant genes vs the difference in L_1 score of the significant microarray data. Left shows the similarity of other drugs to INH, while the right shows the similarity to ETH.

cause their primary treatment mode is injection. Table 5.4 summarizes the allowable drug combinations.

This analysis was also performed for the dataset in [91]. There were two interesting cases when considering similarity scores here. The first was was the pairing ETH with INH. The L_1 distance across significant genes was 0.2720, and over 50% of the significant genes of INH are also significant genes of ETH, however only 42% of the significant genes of ETH were also significant genes of INH. This is due to the fact that there are more significant genes for ETH. Despite the discrepancy, ETH with INH was deemed to be an inadmissible combination. The other case of note is that of Streptomycin, which only had one gene which had consensus significance. Because of the very low number of significant genes, the L_1 distance and overlap with Streptomycin doesn't make for a good comparison.

5.3.2 Drug Ranking Results

After determining inadmissible drug pairings, all remaining 4-drug combinations were ranked according to effect they had on MTB metabolic networks. This was done by

	PZA	EMB	Rifampicin	ETH	Amikacin	INH	Capreomycin	Levofloxacin	Rifapentine	Clofazimine	Pretomanid	Ofloxacin	Streptomycin
PZA													
EMB													
Rifampicin													
ETH													
Amikacin													
INH													
Capreomycin													
Levofloxacin													
Rifapentine													
Clofazimine													
Pretomanid													
Ofloxacin													
Streptomycin													

Table 5.4: This table shows the admissible drug pairs. Cells that are green represent an admissible combination, while red cells represent drugs that cannot appear in the same combination treatment.

first establishing the top ten most affected MTB pathways when every drug was considered. Using the top ten pathways considers only the pathways deemed most important for a drug to limit. For each drug combination the total L_1-score of the affected genes were calculated for each pathway, then the ranking of the combination was based on the sum of effect on the top ten pathways.

After the combination scores are calculated, the four drug regimen is ranked. Generating the bar graphs in Figure 5.7 we noticed that there was a minimal amount of antagonism between the drugs that is not shown. Higher L_1-scores favored combinations with less antagonism, but also favored combinations with drugs that covered many genes in significant pathways. We visually depict the total L_1-score of combinations, and indicate the breakdown of contributions from various drugs with segmented bar graphs. Some drugs did show antagonism, where one drug in the 4-drug combination affects a gene either positively or negatively, but a different drug in the

combination had the opposite effect.

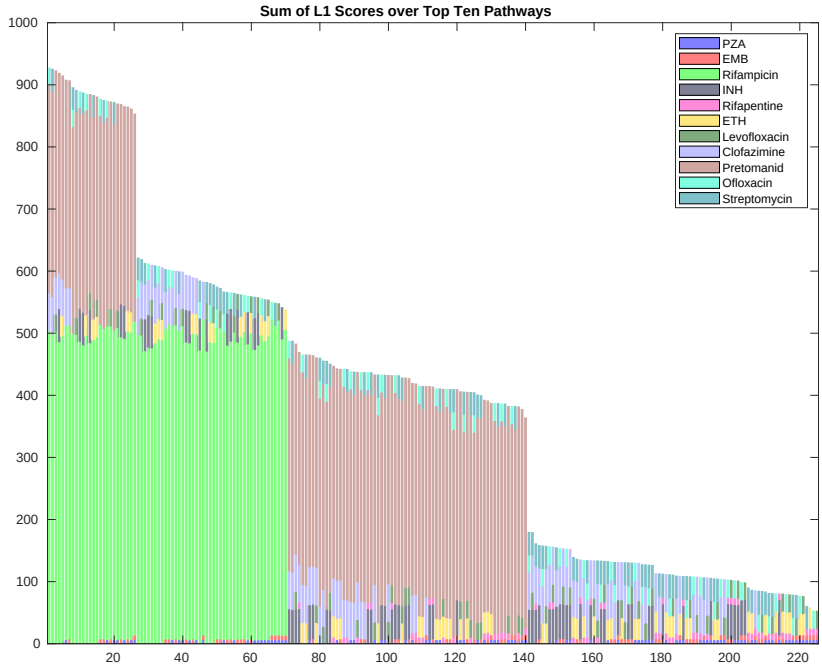


Figure 5.7: 225 four-drug combinations were scored regarding the genes significantly effected by drugs across the top ten significant metabolic pathways. This bar graph shows the L_1 score of each drug combination, and the contributions to the score from each of the four drugs. Injectables such as Amikacin and Capreomycin were excluded.

To achieve this breakdown of contributions by each drug in the presence of antagonistic effects, we show the bar graph at the correct total hight (L_1 score) and with antagonism not explicitly shown, but instead all of the positive (negative) contributions for upregulation (downregulation) are rescaled to give the accurate total height of the bars while maintaining a good estimate of contribution from each drug. For example, the L_1-scores in Figure 5.7, on average are 98% as large as they would be when ignoring the antagonistic effects. The most dramatic case of this (PZA, Ethambutol, Rifampicin, Isoniazid) is the new score is 91% of the score when ignoring the antagonistic effects. Hence this breakdown given in the figures of contribution by each drug is an appropriate estimate.

In Figure 5.7, the L_1-scores of all significant genes in the top 10 pathways are

summed to provide the total score, and this score is partitioned to observe the L_1-score contributions by each drug in the combination.

When filtering the microarray data by significant genes, a trend appeared that rifampicin has the largest effect in combinations. This is consistent with the clinical study of [14] which showed that many combinations with Rifampicin had culture negativity percentage of 85% after two months, while cultures without Rifampicin rarely had above 50%. However we note that in [16], rifampicin had only two experiments instead of three or four as with other drugs in the microarray data. Because of this the Consensus significance was used for Rifampicin, which had a higher value than for any other drug. When analyzing consensus genes from [91], rifampicin again had the largest effect in combinations. This leads to the conclusion that rifampicin is an important drug in treating TB, despite the discrepancy between consensus significance and the SAM method. In the supplemental material of [14], they listed the Beta regression coefficient for several individual drugs. Rifampicin had the strongest effect, with a coefficient of 0.23. The next strongest effect was PZA, with a coefficient of 0.1. Our rankings did show as large an effect for PZA as might be expected. Part of this is due to the minimum fold change of ± 1. As seen in Table 5.2 there are 647 genes of PZA shown to be significant by the SAM method, but only 47 of these had a mean fold change of ± 1.

5.3.3 Simulating drug combinations on central carbon metabolism

Six simulations are presented using the Carbon Metabolism network (mtu01200) from KEGG. Four of these simulations were using data from the [16] dataset, while used data from [91]. Using only genes which directly correspond to reactions in the Carbon Metabolism network, the L_1-scores of the 225 admissible drug combinations (15 for [91]) were calculated. The combinations were ranked based on the L_1-score for this pathway and combinations were chosen to simulate such that results reflect a wide

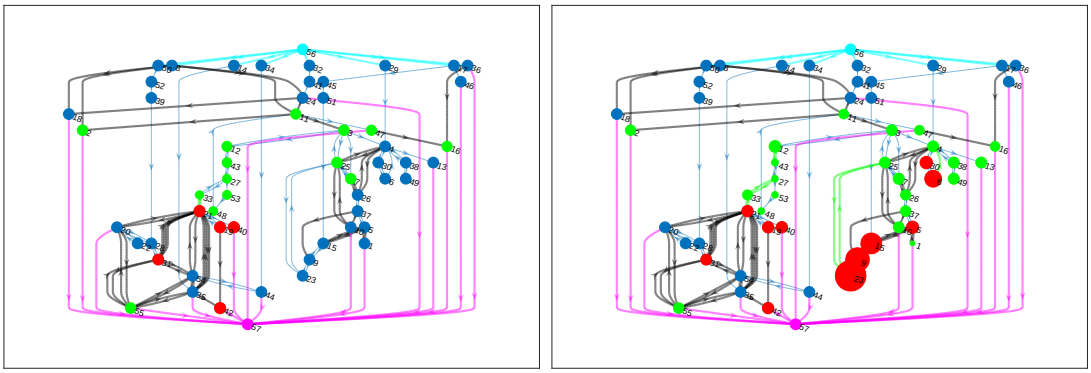


Figure 5.8: Carbon metabolism networks highlighting the metabolites that were affected by the drug combinations Left: INH, Rifapentine, Levofloxacin, Clofazimine and Right: Rifampicin, ETH, Levofloxacin, Pretomanid. The red nodes represent metabolites which increased in mass, while green nodes decreased. Cyan lines represent intakes to the system while magenta represent excretions. Black edges represent hyperedges. Green edges represent enhacned reactions. For the figure on the right, the metabolite that increased the most was Fumarate, while the metabolite that decreased the most was Carbon dioxide.

range of L_1-scores. These combinations are: EMB, Rifapentine, ETH, Lefloxacin with a score of 6.70, INH, Rifapentine, Levofloxacin, Clofazimine with a score of 15.75, PZA, EMB, Rifampicin, Pretomanid with a score of 33.90, and Rifampicin, ETH, Levofloxacin, Pretomanid with a score of 46.80. From the [91] data, the combinations ETH, INH, Moxifloxacin, Pretomanid and ETH, INH, Moxifloxacin, Rifampicin were chosen. The system was run for 200 hours in order to reach equilibrium, then the drug treatment began. At this stage the fluxes were chosen to be uniform. Figure 5.8 and 5.9 show the simulation results of several drug combinations. The metabolites that had significant changes are highlighted in green (decreased) or red (increased metabolite levels).

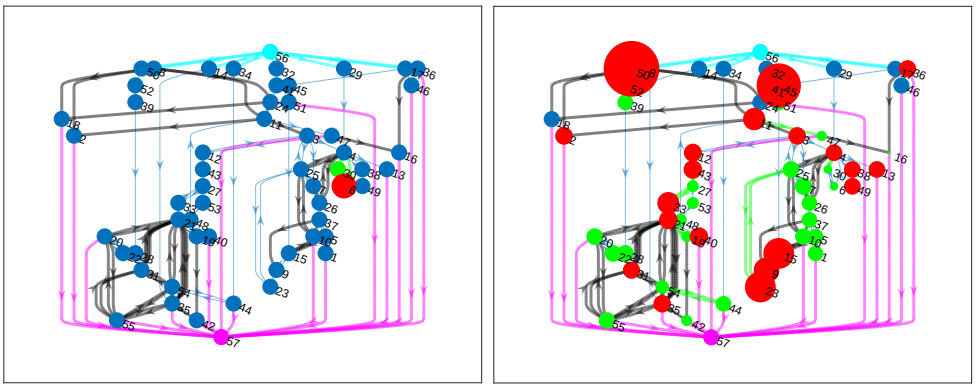


Figure 5.9: Carbon metabolism networks highlighting the metabolites that were affected by the drug combinations Left: ETH INH Moxifloxacin Pretomanid and Right: ETH INH Moxifloxacin Rifampicin. The red nodes represent metabolites which increased in mass, while green nodes decreased. Cyan lines represent intakes to the system while magenta represent excretions. Black edges represent hyperedges. Green edges represent enhacned reactions. For the figure on the left, only two metabolites were affected, Acetic acid increased in mass and Acetyl phosphate decreased. For the figure on the right there were many changes. The most significant are the increases of D-Methylmalonyl-CoA and beta-D-Glucose 6-phosphate.

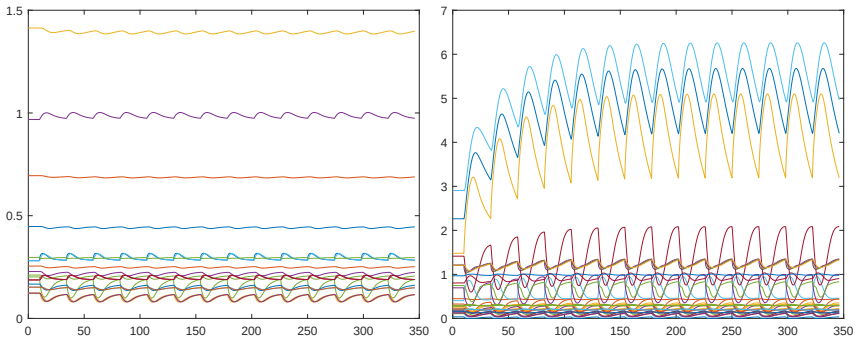


Figure 5.10: 10 hours of equilibrium followed by 336 hours of daily drug treatment. Trajectories of metabolites in central carbon metabolism. Left shows the combination INH, Rifapentine, Levofloxacin, Clofazimine and Right shows the combination Rifampicin, ETH, Levofloxacin, Pretomanid. Due to the system having many metabolites, only the metabolites that had significant change are shown for each simulation.

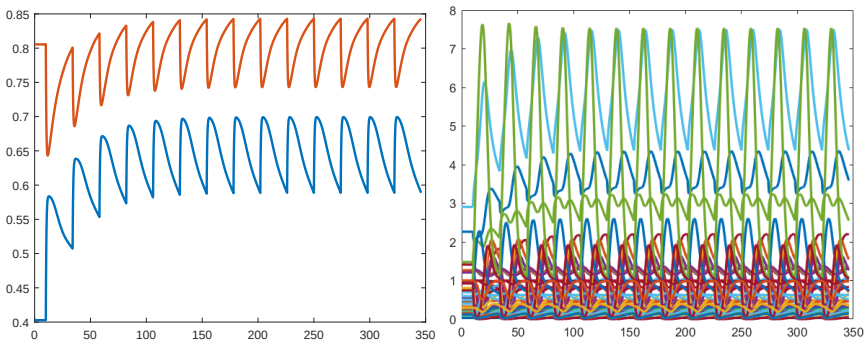


Figure 5.11: Simulations using data from [91]. Simulations had 10 hours of equilibrium followed by 336 hours of daily drug treatment. Trajectories of metabolites in central carbon metabolism. Left shows the combination ETH INH Moxifloxacin Pretomanid and Right shows the combination ETH INH Moxifloxacin Rifampicin. Due to the system having many metabolites, only the metabolites that had significant change are shown for each simulation.

5.4 Discussion

Significant results of the microarray data was the foundation for this analysis. Therefore, a crucial step was to carefully choose our rules for a significant gene expression result. The intersection of the set of genes determined to be significant via a minimum fold change and the set of genes determined by SAM provided the best of both cases, i.e. we could be sure that we see a measurable, significant effect of a drug and also limit the false discovery rate of significant genes. Ideally the MTB gene expression measurements in response to multiple drugs would be obtained using the same array platform and experimental conditions, but this is difficult to obtain as most studies focus on a single drug or a small group of drugs. We tested a novel way to determine admissible drug combinations and rank the effectiveness of these combinations. We then improved our model of metabolism by including two new features 1. a more sophisticated framework for modeling drug action on enzymes and 2. a method of modeling complex biochemical reactions involving more than two metabolites that maintains dependencies among reactant variables. We incorporated this model in simulations in order to compare top ranked drug combinations with their effects on

metabolite trajectories.

Calculating the L_1-scores for each drug as it pertained to microarray data allowed us to calculate L_1 distance between drug effects. An initial result of ranking drugs by the L_1-scores across significant genes in microarray data was predicting drugs that cannot be used in combination treatment. For a pair of drugs, a low L_1 distance, and a high number of genes significant for both drugs (a high overlap) would suggest that the drugs have behavior that is too similar. We recovered known pairs of drugs considered inadmissible with each other in this way as a preliminary validation of the L_1-score. These results are shown in Table 5.3, and Table 5.4.

From figure 5.7 we see the dramatic effect of Rifampicin and Pretomanid. All 4-drug combinations omitting both of these drugs have a total effect on gene expression that is smaller than Rifampicin or Pretomanid alone. This highlights the importance of these drugs to treatment, and suggested that we should test this with simulation. On the other hand, combinations with a lower score that show dramatic results in a clinical setting or simulation suggests importance of the pathways that they affect.

Metabolic graphs combined with LIFE methodology model drug treatment that interact as inhibitors or enhancers of biochemical reactions, moreover these biochemical reactions often should be depicted as hyperedges in a metabolic network as opposed to simple edges. The additional modeling features of LIFE methodology account for 1. more sophisticated dynamics for drug interaction with reaction enzymes and 2. reactions involving three or more metabolites and the dependence of these metabolites on each other for a reaction to proceed. This framework allowed the last step of our analysis, which was to compare a drug combination's L_1-score with simulated treatment on a network (shown in Table 5.5).

Figure 5.8 and Figure 5.9 show exactly which nodes in the central carbon pathway were affected most by drug treatment. Figure 5.10 and Figure 5.11 show corresponding metabolite trajectories. The figure on the right of Figure 5.8 shows that Fumarate

Drug Combination	Sig. Gene L_1	Simulation L_1
Rifampicin, ETH, Levofloxacin, Pretomanid	46.7965	9.2416
PZA, EMB, Rifampicin, Pretomanid	33.9035	6.6246
INH, Rifapentine, Levofloxacin, Clofazimine	15.7467	0.4005
EMB, Rifapentine, ETH, Lefloxacin	6.7010	1.4407

Table 5.5: The L_1 score based on a drug combination affect on significant genes and system simulations. The simulation L_1 is based on the difference of metabolites from the system equilibrium when no drug is applied. Each combination is based on data from [16].

increased the most and Carbon dioxide decreased the most. We show that microarray data and simulation can suggest the most important pathways with regard to a combination treatment, and the most important biomarkers therein. For example if it is discovered that a patient has low Fumarate levels, then the combination {Rifampicin, ETH, Levofloxacin, Pretomanid} may be the superior treatment.

Based on microarray data from the Sherman lab, the L_1-score of drug combinations had to be simulated for some network in order to compare the score with metabolic effects of the drug combination. We chose central carbon metabolism for its size and simplicity, as it has a relatively small number of hyperedges compared to most metabolic pathways. We showed drug combinations of {ETH, INH, Moxifloxacin, and Pretomanit} and compared this to {ETH, INH, Moxifloxacin, and Rifampicin} as those drugs interacted with many edges of the network, and were able to illustrate the effect of including Rifampicin in treatment. The L_1-scores we had computed highlighted the dramatic impact of Rifampicin on gene expression, and we wanted to explore its role in combination therapy.

Chapter 6

A Two-step Model for Circadian Entrainment

[6] [113] [119] [87]

6.1 Introduction to Circadian Rhythm

Various organisms have biological rhythms that cycle daily, monthly, or at other frequencies, these rhythms assist in predicting future stressors [120]. Biological clocks that cycle near 24 hours are designated circadian rhythms. One hallmark of biological clocks is that they will entrain to an external signal, called a zeitgeber. Entrainment provides a source of clock resetting, allowing the organism to maintain synchronization of its internal rhythm to the external environment. The mechanisms which regulate natural rhythm processes, including clock resetting, vary among species. However, the most significant zeitgeber for circadian rhythms is the sun.

Two notable aspects of circadian rhythm are the endogenous period and the phase of entrainment. The endogenous period, or free-running period, is the length of time it takes a biological rhythm to cycle in the absence of a zeitgeber [38]. The phase refers to the difference between the peaks of the natural cycle and the zeitgeber cycle.

Upon introduction to a new environment, the phase between the biological clock and zeitgeber will be in constant flux. After entrainment, the phase difference becomes stable, and its value is the phase of entrainment [7].

Many studies focus on the endogenous period of an organism. Genetic analysis has concentrated on genes that affect the endogenous period [88, 145]. There have been fewer studies that focus on finding genes that alter the phase of entrainment. There is a common assumption that there is a simple relationship between the phase of entrainment and the endogenous period, namely that the period mismatch and phase of entrainment are negatively correlated. Variations of both period and phase of entrainment exist in natural populations, suggesting natural selection could select for either trait [30, 101].

Several circadian rhythm disorders are related to the phase of entrainment. Familial advanced sleep phase syndrome (FASPS) is associated with waking up and going to sleep at earlier times than usual, while delayed sleep phase disorder (DSPD) marks difficulty falling asleep at an appropriate time [82, 143]. These conditions motivate a study of the phase of entrainment.

Previous work illustrates different views on the entrainment process, as well as how to model the phase relationship between the circadian clock and zeitgeber clock. Many mathematical models reproduce entrainment and simulate phase changes [1, 80, 132, 134]. However, many models only produce data that follow a simple rule; a long endogenous period leads to phase delay, while a short endogenous period leads to phase advance [1, 55]. Many studies show relationships between the endogenous period and the phase that support this assumption [35, 62, 67, 124, 141]. An additional paper concludes that the endogenous period can characterize phase phenotypes [63]. Finally, some research related to FASPS reports the human phenotype of an advanced phase is associated with a shorter circadian period [143]. Despite its common usage, there exists experimental data that does not follow this usual assumption. The same

paper ([143]) that suggested a short period is associated with FASPS also states that different mechanisms may regulate the period and phase, implying other period phase relationships may be possible. Another group investigating FASPS found a variant for which subjects have a standard circadian period, but still, develop FASPS [82]. While the elderly often have an advanced phase, this is not associated with a shortening of the circadian period [36, 37]. In [58], the period mismatch is not a predictor for the phase of entrainment. A more general model that incorporates this data is essential to analyze the phase of entrainment.

In [86], they developed a Two-Step Entrainment model. The main goal of this model is to describe a wide range of period to phase relationships. The framework of this model is to use three clocks, a zeitgeber, and SCN and an endogenous clock to model entrainment. The SCN clock entrains to the zeitgeber, while the endogenous entraining to the SCN thus it is two-step entrainment. This model can generate in-silico data that does not follow the common assumption and can fit a more extensive range of results. Here a two-step model is necessary, as single-step models (such as [1]) have not yet generated this type of data.

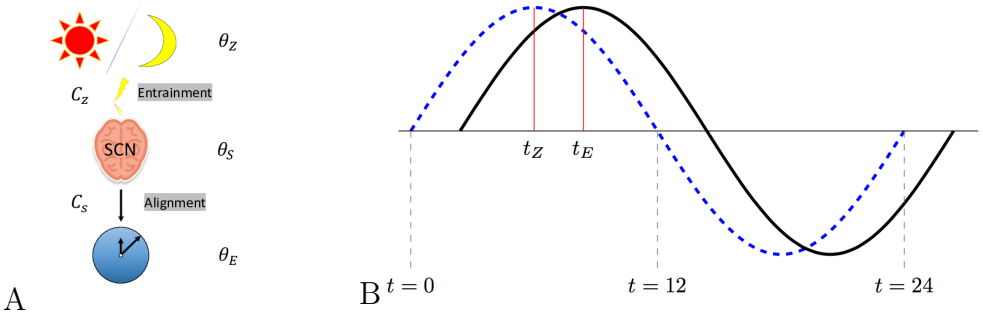


Figure 6.1: (A) Schematic illustrating a two-step entrainment process between the zeitgeber and peripheral clock. In step 1, the SCN clock is entrained to the zeitgeber during the entrainment window with strength c_z. In step 2, the peripheral clock is aligned to the SCN clock constantly with strength c_s. (B) Evolution of phase angles in one period through the plots of sine curves. The dotted line represents the zeitgeber clock, and the solid line represents the peripheral clock. The time when phase angles of the zeitgeber and peripheral clock are equals to $\frac{\pi}{2}$ are label as t_Z and t_E, respectively. The phase difference is measured as $\phi = t_Z - t_E$. A positive value of ϕ indicates an advanced phase, a negative value of ϕ indicates a delayed phase.

This work uses phase data from [58] to fit the model. Throughout their research, they measured core body temperature and plasma melatonin rhythms to determine phase differences. During their study, subjects free-cycle in constant-dark conditions, then they were entrained by being exposed to light in the evening. While being entrained, they measured the subjects' core body temperature and plasma melatonin rhythms to determine phase differences [58]. They provide phase data throughout the experiment, not only the start and end, allowing fitting the entire trajectory of the phase difference. This work analyzes the potential of the two-step entrainment model. This includes the effects each parameter has on entrainment, the time it takes to entrain [56], the range for which entrainment is possible, and the reachable phase of entrainment values.

6.2 Mathematical Model

The two-step circadian model is based on the Kuramoto model [81] but has three oscillators. These are the phase angle of the zeitgeber, the phase angle of the SCN, and the phase angle of the peripheral clock. The SCN clock entrains to the zeitgeber at a constant rate, yet the peripheral clock only entrains to the SCN during a small window. Some studies suggest that entrainment doesn't occur throughout the day but during times such as dawn and dusk [110]. The two-step model uses an entrainment window of 1 hour near dusk. The entrainment window can also be adjusted to suit what is appropriate for different organisms [69]. [86] applied the two-step model to a Neurospora system, whereas this work uses it to model human data.

The following system of ODEs governs the dynamics of the three oscillators:

$$\begin{cases} \frac{d\theta_Z}{dt} &= \frac{2\pi}{T} \\ \frac{d\theta_S}{dt} &= \frac{2\pi}{\tau} + \Psi(\theta_Z)(\theta_Z(t) - \theta_S(t)) \\ \frac{d\theta_E}{dt} &= \frac{2\pi}{\tau} + c_s(\theta_S(t - t_0) - \theta_E(t)) \end{cases} \tag{6.1}$$

with:

$$\Psi(\theta_Z) = \begin{cases} c_z & |\theta_Z - \pi| < \epsilon \\ 0 & \text{otherwise.} \end{cases} \tag{6.2}$$

The variable θ_Z represents the angle of the zeitgeber clock, the variable θ_S represents the angle of the SCN clock, and the variable θ_E represents the angle of the endogenous clock. T represents the period of the zeitgeber, and τ represents the period of the SCN and endogenous clocks. c_z represents the entrainment strength between the SCN and the zeitgeber. The entrainment strength measures the intensity of the light as well as the light sensitivity of the individual. c_s represents the alignment strength between endogenous and SCN. The alignment strength corresponds with how the SCN relates to the endogenous. ϵ is the radius of the entrainment window. $\Psi(\theta_Z)$ is the entrainment function, which will be zero except when the zeitgeber is near dusk. Finally, t_0 is the time offset between the endogenous and SCN clocks.

Because of the entrainment window, the system cannot be solved analytically. Finding a partial piecewise solution is possible, and this solution can still be advantageous. The following equations show the piecewise solution, with five additional constants

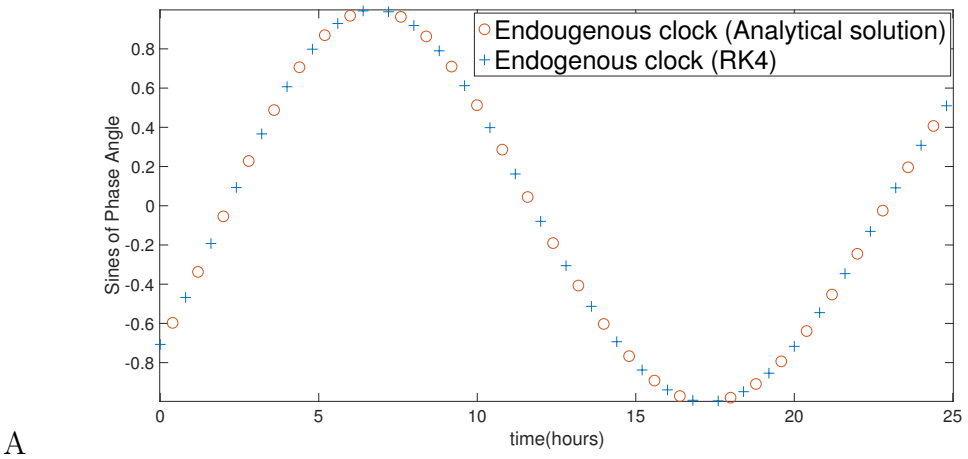
$$\theta_Z = \frac{2\pi}{T}t + k_1, \tag{6.3}$$

$$\theta_S = \begin{cases} \frac{1}{c_z}\frac{2\pi}{\tau} - \frac{1}{c_z}\frac{2\pi}{T} + \theta_z + k_2 e^{-c_z t} & , |\theta_Z \pmod{2\pi} - \pi| < \epsilon \\ \frac{2\pi}{\tau}t + k_3 & , \text{otherwise} \end{cases} \tag{6.4}$$

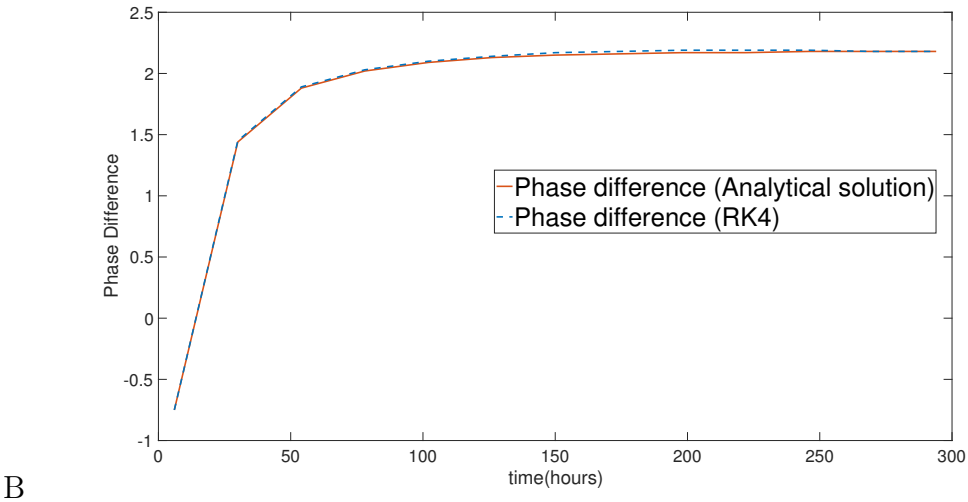
$$\theta_E = \begin{cases} \theta_s(t) + \frac{1}{c_s}(\frac{2\pi}{\tau} - \frac{2\pi}{T}) + \frac{c_z k_2 e^{-c_z t}}{c_s - c_z} + k_4 e^{-c_s t} & , |\theta_Z \pmod{2\pi} - \pi| < \epsilon \\ \theta_s(t) + k_5 e^{-c_s t} & , \text{otherwise.} \end{cases} \tag{6.5}$$

The initial values of k_1, k_3, k_5 have biological meaning with k_1 equal to the initial phase angle of the zeitgeber (denoted by θ_{Z0}), k_3 equal to the initial phase angle of the SCN clock (denoted by θ_{S0}), and $k_5 + k_3$ equal to the initial phase angle of the peripheral clock (denoted by θ_{E0}).

We simulated the system 6.3,6.4,6.5 using Matlab (2018a). To ensure continuity of the piecewise solution, the constants k2-k5 must be re-solved whenever $|\theta_Z \pmod{2\pi} - \pi| = \epsilon$. In other words, whenever the entrainment window is entered or exited, the k2-k5 must be resolved. Using this method of simulation is more than ten times faster than the Runge-Kutta scheme of system 6.1 with similar accuracy. Upon adding additional parameters finding a partial solution may not be possible, in which case a Runge-Kutta scheme would be necessary. Figure 6.2 shows a comparison of the endogenous clock evolution and phase difference for each method. Experimentally, the daily progression of the phase is calculated by taking the phase difference between the zeitgeber and the endogenous clock. Our simulations calculate the phase difference by measuring the phase between the peaks of the endogenous clock and the zeitgeber, i.e., the time the zeitgeber peaks minus the time the endogenous clock peaks. This means that a negative phase of entrainment corresponds to a phase delay, and a positive phase of entrainment corresponds to a phase advance [89].



A



B

Figure 6.2: Simulations are performed by two different approaches with the same parameters. A. The comparison between the phase angle of the endogenous clocks simulated by two approaches. The differences is less than 0.01h. B. The comparison between the phase differences simulated by two approaches. The phase differences which are measured at noon of each day between the zeitgeber and the peripheral clock. The parameters used are: $T = 24$, $\tau = 23$, $c_z = 0.5$, $c_s = 0.1$, $\epsilon = \frac{\pi}{24}$ (30 minutes) , $t_0 = 0$, $\theta_{Z0} = 0$, $\theta_S = \frac{\pi}{12}$, $\theta_E = -\frac{\pi}{4}$.

6.3 Results

6.3.1 Parameter Space

Before fitting the model to real-world data, it is essential to understand the range of the output. It is also critical to confirm that the two-step entrainment model produces

results that both follow and do not follow the common assumption. To create a range of reachable phases of entrainment values, the parameters, c_z and τ, which have the most substantial impact were varied. We used a comprehensive domain of both τ and c_z to capture modeling capabilities. The domain of τ's consisted of thirty-one values, varying from 22.5-25.5 hours at 6-minute intervals, while c_z's consisted of twenty-two values, ranging from 0.15-1.00 hours at intervals of 0.04. We fixed the other parameters at $c_s = 0.26, T = 24$hours, $\epsilon = \frac{\pi}{24}$, $\theta_{Z0} = 0, \theta_{S0} = 0$, and $\theta_{E0} = 0$ the ran simulations ran for 1100 hours. Figure 6.3 panel A shows the results of these simulations. Each of the 682 tests entrained within 600 hours. Although there is a large range of phases reached, 92% of the trials had a phase of entrainment within [-6h to 4h].

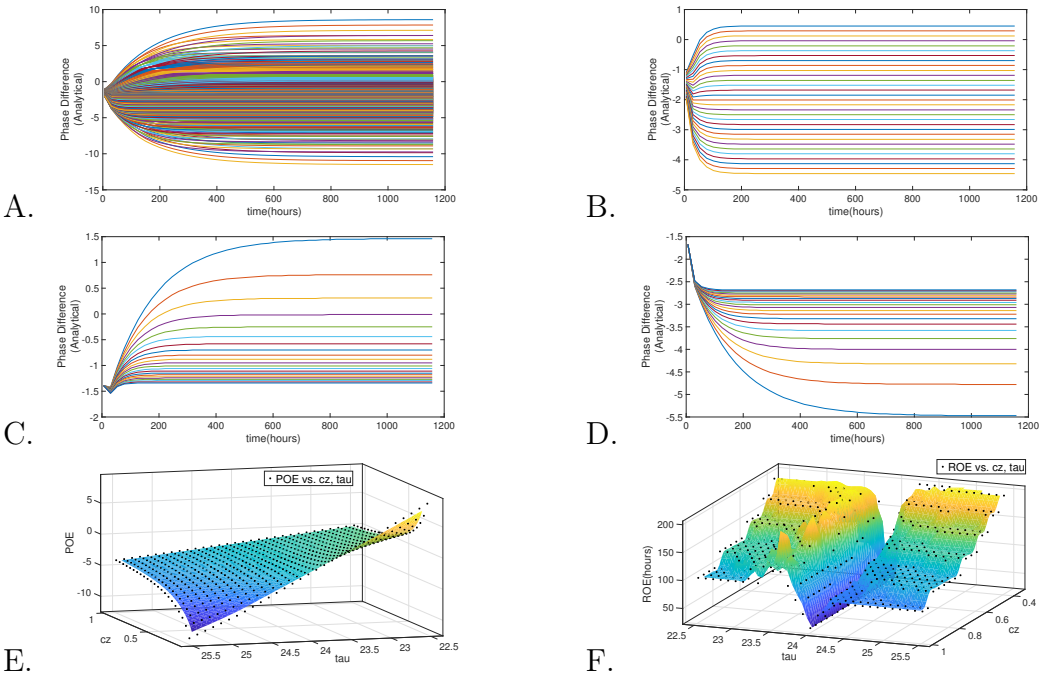


Figure 6.3: Simulations present the parameter space when varying τ and c_z in a wide range. (A) The trajectories of phase differences of all subjects (682 pairs of τ, c_z) are shown. (B) Fix c_z at 0.76, the trajectories of phase differences change with varying τ from 22.5 to 25.5. (C) Fix τ at 23.5, the trajectories of phase differences change with varying c_z from 0.15 to 1. (D) Fix τ at 24.5, the trajectories of phase differences change with varying c_z from 0.15 to 1. (E) 3D linear interpolate surface represents the simulated Ψ vs. $-c_z, \tau$ parameter space (all 682 data points are plotted.) (F) 3D linear interpolate surface represents the simulated ROE vs. $-c_z, \tau$ parameter space. In E (F), blue color corresponds to small values of POE (ROE), yellow color corresponds to lager values of POE (ROE).

To determine when the model would produce results that did not follow the common assumption, τ and c_z were varied individually. Panel B of Figure 6.3 shows that when c_z is fixed at a high value that no variation of τ produced results which did not follow the common assumption. We used τ of 23.5 and 24.5 to test whether τ values near 24 hours could produce results not following the common assumption, as before c_z varied from 0.15-1.00. Panel D that when $\tau = 24.5$ did not produce any results which don't follow the common assumption, however, for $\tau = 23.5$, small values of c_z do produce results not following the common assumption. This shows that the model can achieve results which do not follow the common assumption, although

these results are still not a frequent occurrence.

Figure 6.3 panel E shows a 3-dimensional plot of the simulations performed varying c_z and τ. This plot shows how the phase of entrainment changes for each variable and shows the manifold they exist on. Similarly, Figure 6.3 panel F shows the surface relating the rate of entrainment to the variables of c_z, τ.

6.3.2 Range of Entrainment

The previous section was primarily to understand how the phase of entrainment would change with respect to c_z and τ, and what entrainment values are possible. This section deals with the range of entrainment, or what parameter combinations can overcome $\tau - T$ period mismatches [55]. In [15], they used an Arnold Tongue to display when entrainment strength could overcome period mismatch. To do this, they varied the entrainment strength and zeitgeber period while keeping the endogenous period constant. We create a comparable structure by varying the entrainment strength, c_z, entrainment window, ϵ, and zeitgeber period, T while keeping the endogenous period, τ constant. Figure 6.4 shows the results of these variations. The 3-D Arnold Tongue represents the range of entrainment for the SCN. To interpret the figure, any data point that lies within the bounds represents conditions that will overcome the period mismatch and entrain to the zeitgeber.

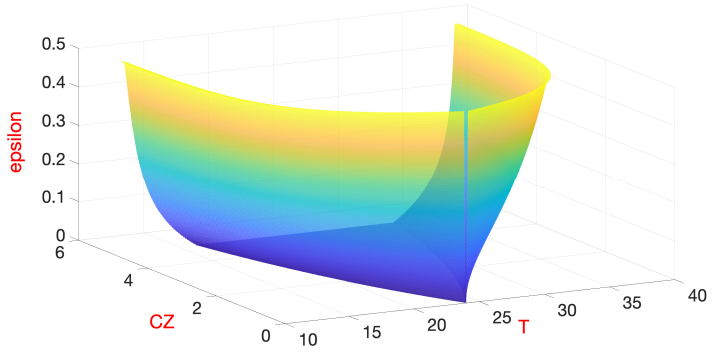


Figure 6.4: A 3D Arnold Tongue represents the range of entrainment for the SCN clock. The SCN clock is entrained to the zeitgeber when the data point locates above the surface (inside). The SCN clock is always entrained when T= τ.

6.3.3 Parameter Affect on Entrainment

After the parameter space was determined, it is critical to understand how each parameter influences the final entrainment value. The parameters c_z, c_s,ϵ, and t_0 were varied individually to find ranges that generate a reasonable phase of entrainment. Three key indicators defined the differentiation among parameter outcomes. These were 1. whether or not the system entrained, 2. the final entrainment value, and 3. the rate of entrainment. Figure 6.5 shows the results of varying each parameter.

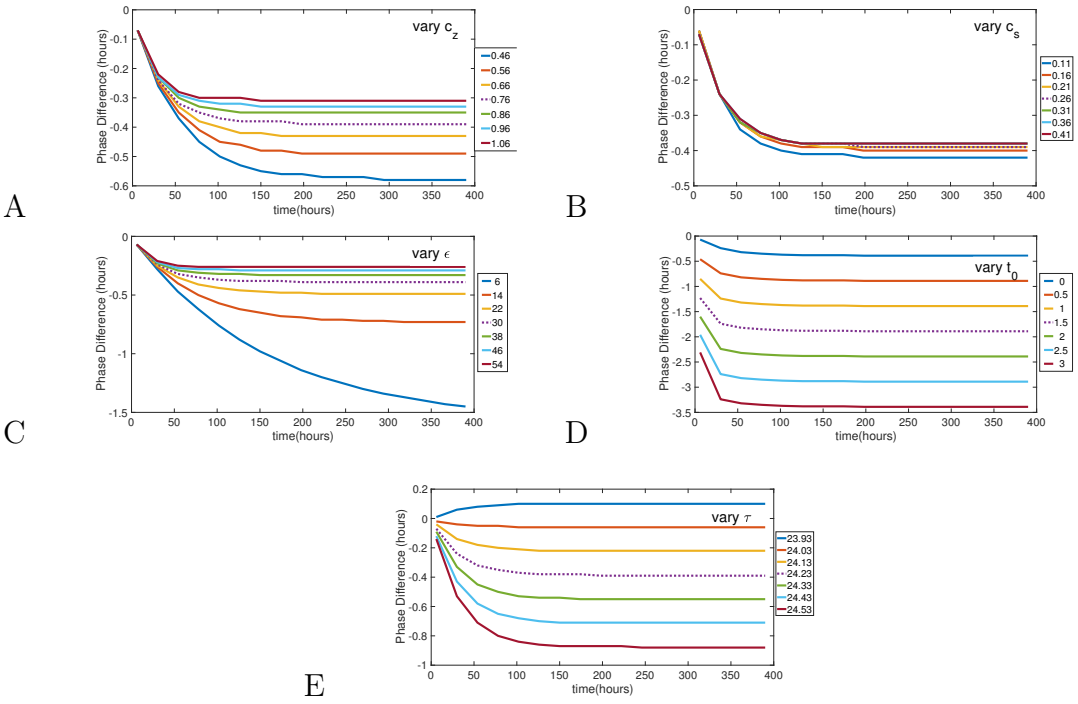


Figure 6.5: Single parameter variations for the “average human" based on the 12 subjects. In each plot only one parameter varies while the others remain constant. The legend represents the varying parameter values. The dashed line in each plot represents the baseline parameter value which is held constant in all other plots except for when it is the varied parameter. (A), (B), (C), (D), and (E) show how trajectories change when only vary c_z, c_s, ϵ, t_0, and τ, respectively.

The panels of Figure 6.5 show variations of each parameter. c_z and e have similar effects on the phase of entrainment. This similarity is due to the comparable functions of these parameters; increasing c_z increases the intensity of light, while e increases the amount of light used in entrainment. Panel A shows the trials of c_z from 0.46 to 1.06 at 0.1 increments. Panel C shows trials of ϵ varying from 6 minutes to 54 minutes in 8-minute increments. The parameter c_s has a much smaller impact on the system, with both the rate and phase of entrainment changing only slightly. Panel B shows this effect. Panel D shows the variations of t_0 from 0.50 hours to 3.5 hours by 0.5 hours increments. Varying t_0 shifts the trajectory downward but doesn't affect the rate of entrainment. Finally, panel E shows that as τ varies from 23.93 hours to 24.53 hours by 0.1 hour, both the rate of entrainment and phase of entrainment are

significantly changed.

An important observation from this variation is that the parameters are monotonic. When a parameter changes, the phase of entrainment will change in a predictable direction, although the magnitude of this change depends on the current parameter value. Another observation is that each parameter has different effects on the phase of entrainment and the rate of entrainment. These effects are necessary to consider when fitting data.

6.3.4 Trajectory Fitting

The two-step entrainment has a total of three oscillating clocks. Although the SCN clock is necessary for the two-step entrainment process, trajectory fitting doesn't use the output from the SCN clock. The primary interest of trajectory fitting is to fit the phase between the endogenous clock compared to the zeitgeber clock, particularly at entrainment. To this end, the trajectory fitting weighs data near the entrainment to be more significant than points near the commencement of the experiment. Giving more weight to data near the end of the trial means the fitting favors phase of entrainment over the rate of entrainment. In most circumstances, both the phase of entrainment and the rate of entrainment can have a suitable fit.

We used data from [58] in which they studied entraining patients to longer than 24-hour days. They used a zeitgeber period $T = \tau + 1$ h and entrained 12 patients in 3 different light conditions, dim light 25 lux, room light of 100 lux, and a modulated light exposure (MLE). Their experiment lasted 30 days, and they measured core body temperature and plasma melatonin to determine circadian periods and phases. They considered a subject entrained when the 95% confidence interval of their period included the zeitgeber period [58]. They found that all four MLE and room light patients entrained, while only one of the dim light patients entrained [58]. We use their data to test the two-step entrainment model's fitting ability.

To simplify the parameter space, only two variables, CZ and t0, were used for fitting. Using these parameters in combination, both the rate of entrainment and phase of entrainment can be modified. The initial phase angles of each oscillator are set equal to the initial phase difference of each subject. The variable cs was fixed at 0.26, while epsilon was set at $\frac{\pi}{T}$, which corresponds to a one-hour window for all subjects.

Finding the best-fit values for c_z, t_0, and θ_{E0} was done using a golden-section search. All patients in the 100 lux and MLE experiments entrained, but only subject 2195, shown in Figure 6.6 panel A, entrained in the 25 lux. The optimal values of c_z, t_0, and τ, as well as other parameter values for each patient are found in Table 6.1. The fitting results show that subjects entrained in higher light intensities have larger c_z values.

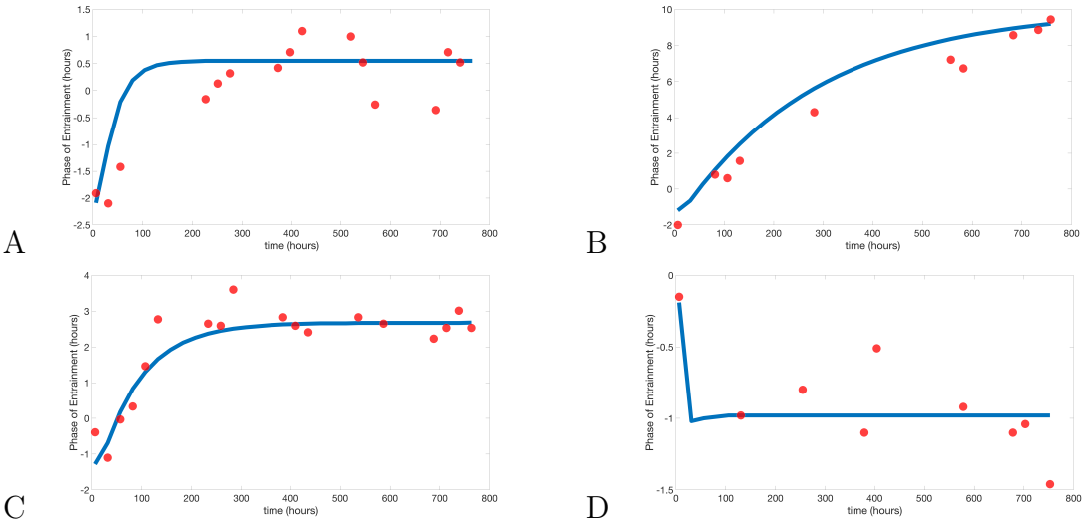


Figure 6.6: Best fit curves from the TSE model to fit phase tracjectories of four subjects from [58]. Panels A and B represent subjects 2195 and 2313 in 25 lux conditions. Panel C represents subject 2123 in 100 lux conditions, and panle D represent subject 2196 in MLE conditions.

Light Conditions	Subject	T	τ	c_z	t_0	θ_{E0}	θ_{S0}	Ψ	cost
25lux	2195	24.47	23.47	0.74	1.11	0.30	-0.50	0.55	1.48
	2209	24.58	23.58	0.14	1.25	-0.50	-0.60	6.05	3.06
	22T1	24.75	23.75	0.03	0.88	0.13	-0.10	20.68	4.19
	2313	25.05	24.05	0.09	1.50	0.05	-0.10	9.20	2.61
100lux	2123	25.24	24.24	0.30	0.94	0.30	-0.30	2.68	0.77
	2072	25.33	24.33	0.44	1.21	0.30	0.00	1.35	1.42
	2109	25.30	24.30	0.55	0.60	0.30	-0.50	1.51	1.47
	22F2	25.24	24.24	0.57	0.82	0.27	-0.85	1.23	1.14
MLE	2082	25.49	24.35	0.74	0.79	0.30	0.00	1.11	0.41
	2111	25.48	24.48	0.96	1.49	0.30	-0.60	-0.12	0.52
	2196	24.87	23.87	0.86	2.46	0.30	0.30	-0.98	0.70
	2210	25.25	24.25	0.75	0.94	0.30	-0.60	0.71	1.25

Table 6.1: Table shows Ψ and the optimized parameter values of c_z, t_0, θ_{E0} for all 12 subjects.

Chapter 7

Ecological Networks Reconstructed from Paleomiddens

7.1 Introduction

The interactions of organisms is at the root of ecology. A crucial step to fully understanding a community is to understand the interactions of its species [84]. A significant question in ecology today is how communities will react to changes, including changes in climate, introduction or loss of species, and urbanization. The robustness of a community can be difficult to determine, and identification of key or at risk species is not a simple task [57, 112].

Networks are a tool that help us describe the interactions of species [84, 115, 114]. Species are represented as nodes in the network while their interactions are directed edges. A networks approach is advantageous as it makes relationships between species explicit, as opposed to other methods which treat each species as independent [116]. Often in ecology the links between species cannot be precisely determined and must be inferred through other methods. Some methods could include live or video observation of consumer-resource interactions, predation experiments in cage studies, prey

baits, and molecular analysis of gut content [13]. A networks approach is used to examine how entire communities might respond to change. Network dynamic models can help determine the stability of the community, and centrality metrics can highlight which species might be integral. Lotka–Volterra models can be used to predict the equilibrium populations or which species may be heading toward extinction. A limitation of modern networks is there is a limited temporal window. This scale provides few changes in dynamics to analyze.

Paleoecological networks are promising for analyzing dynamics because of the extended temporal window. This allows us to see changes in the community makeup, such as addition or removal of species and links in the network. More importantly, the ways communities have adapted in the past may help predict how ecological systems may respond to environmental changes. Paleoecology presents difficulties as there are additional biases not present when studying modern communities. There are two primary sources of bias that can affect paleoecologic data. These are 1. Taphonomic processes, which affect an organism's likelihood of being preserved, and 2. Time Averaging, where organisms from different time periods may be found preserved together. These biases cause some uncertainty about the recovered data.

There are many unknowns when dealing with paleoecology. The links between species are not recoverable, the population sizes may not be known, and there may even be species that are absent as they were not preserved. Due to the unknowns there are many methods of constructing paleoecological networks. The method used will depend on the goals of the project, as well as the data available. In [144] they integrated the depictions of animals from archeological sites to use in reconstructing egyptian large mammal networks. Another study uses a number of ten categories to specify links between ancient species [42]. These categories are (i) taxonomic uniformitarianism, (ii) functional morphology, (iii) gut contents, (iv) damage patterns, (v) stratigraphic co-occurrence, (vi) body size, (vii) coprolites, (viii) host relationship,

(xi) chemical and isotopic signatures and (x) ichnological evidence. This approach requires a very good resolution of the data, which may not always be present. [122] uses metanetworks to sort the specific species into similar bins and record likely links between the meta-species bins. Species specific networks can then be created by using random draws to create links between species which exist in connecting bins. This method is used primarily to determine the robustness of a network to cascading extinction effects. This work aims to use a method similar to [42] to recreate networks based on information from rodent middens.

This work uses arthropod data, whose remains were found in rodent middens preserved in the Atacama desert [34]. Rodent middens are a combination of plant matter, arthropod remains, and fecal remains of the rodent bonded by the urine. Middens are often found in dry and arid climates and provide excellent preservation of material. Despite the excellent preservation, there are several drawbacks to paleomidden data [34]. First, a midden can be created in a few months, or over many years, possibly even centuries. Second, creating a database of middens requires the collection of many middens from the same area. Different middens, even from the same area, may not have been created contemporarily, this can result in the data having irregular and possibly significant temporal gaps. Despite middens being found in many locations around the world, and having good preservation properties for arthropods, there has not yet been methods specifically focused on network reconstruction for middens. Here we presents first steps for a systematic approach to recreate paleonetworks using rodent middens.

7.2 Paleomidden Data

The paleomiddens were collected from two sites in the Atacama Desert, Lomas de Tilocalar and Vegas de Tilocalar. Middens were collected from cavities in the rocks

and cliffs in these locations. A total of 46 total middens were collected and analyzed, 14 from Vegas de Tilocalar and 32 from Lomas de Tilocalar.

There are three principal divisions of the data collected in [34]. These are the Modern Dead (MD), Modern Living (ML), and Paleomiddens. Modern Dead were dead arthropods collected from death assemblages collected close to paleomiddens as well as from modern middens. The exact time of creation of modern middens is unknown, however they are likely to have been created in the last few decades. In the dataset Modern Dead assemblages are assumed to be from 50 years ago. Modern Living is a combination of local live data and regional live data. The local live data was collected through pitfall and sticky traps from within 15 meters of found midden. Regional live is from between .5km 3km away from closest fossil. Associated with each paleomidden are its latitude and longitude as well as its carbon-14 date. The most recent paleomidden formed around 400 years ago, while the most ancient appeared close to 40,900 years ago. Species data was collected from 42 different locations. They collected paleomiddens at 26 locations, modern dead from 12 locations, and information about modern living at 4 locations. There are a total of 85 different morphospecies, however, only 43 of these appear in the paleomiddens. Additionally, there are only 26 species that appear in both the paleomidden data and either modern living or modern dead. Table 7.1 summarizes how many species appear in various data partitions.

	# of Species	MD	ML
Total	85	36	46
Paleo Middens	43	24	13
Modern Dead (MD)	36	36	14
Modern Living (ML)	46	14	46

Table 7.1: This table shows the number of species found in the different subsections of the data. The columns MD (ML) show the number of species which are contained in MD (ML) and the data partition described in the row.

Spatial distribution of species and potential links

The places where a species is found are significant in determining its potential interactions. Two species may interact even when not found at the same location. However, the chance of interaction lowers as the distance between locations increases. The sizes of the two midden collection sites were very different. The 14 middens collected at the Vegas de Tilocalar site were collected in a 30 square kilometer area with a length of about 12 kilometers. The Lomas de Tilocalar site was much larger, with middens being collected from an area of nearly 300 square kilometer. Within the Lomas de Tilocalar there were two main clusters of middens. The first had 18 middens and has an area of 70 square kilometers, the second with 14 middens was more than 10 kilometers away from the first and has an area of 40 square kilometers.

The maximum distance between any two paleomiddens is 33 kilometers. Figure 7.1 shows the locations of all the paleomiddens. The x and y-axis of the figure show the longitude and latitude, respectively. The horizontal red line shows a distance of 10 kilometers. The colors represnt how long ago the middens were formed, from the figure there are no clear correlations with when the middens are formed and where they are found. Figure 7.2 shows the locations of modern dead and modern living collection.

Whether two species interact may also depend on the distributional range of each species. Here we look at two species and how often their ranges overlap when considering several potential range values. We analyzed the potential range overlap between Physogaster spp, a detrivore species, and Bothriuridae spp, a carnivore species, which may interact. Only considering paleomidden appearances, Physogaster spp had 17 appearances, while Bothriuridae had 22. With a given range of zero, we examine how often they appeared at the same locations. 11 of the 17 Physogaster appeared at a site with Bothriuridae, while 10 of the 22 Bothriurid appeared at a site with Physogaster. This gives a ratio of $\frac{11+10}{17+22}$ or 53.8% of occurrences have overlapping range. When we

	0m range	500m range	2500m range
Ancient Middens	53.8%	69.2%	94.9%
Full Data	45.1%	82.4%	97.8%
Modern Dead	58.8%	70.6%	85.3%
Modern Living	NA	NA	NA

Table 7.2: Spatial Comparison of Physogaster spp. and Bothriuridae spp. and how often they appear within 0 km, 1 km and 5km of each other. Bothriuridae spp. does not appear in the modern living set, which makes analysis not available.

extend the range of each species to 500 meters, this increases to 69.2%. Finally, when we increased the range to 2500 meters, it is 94.9%. Table 7.2 shows this comparison when all data is considered as well as when only modern dead information is considered. It is clear that the range of a species will increase its potential to react with nearby species.

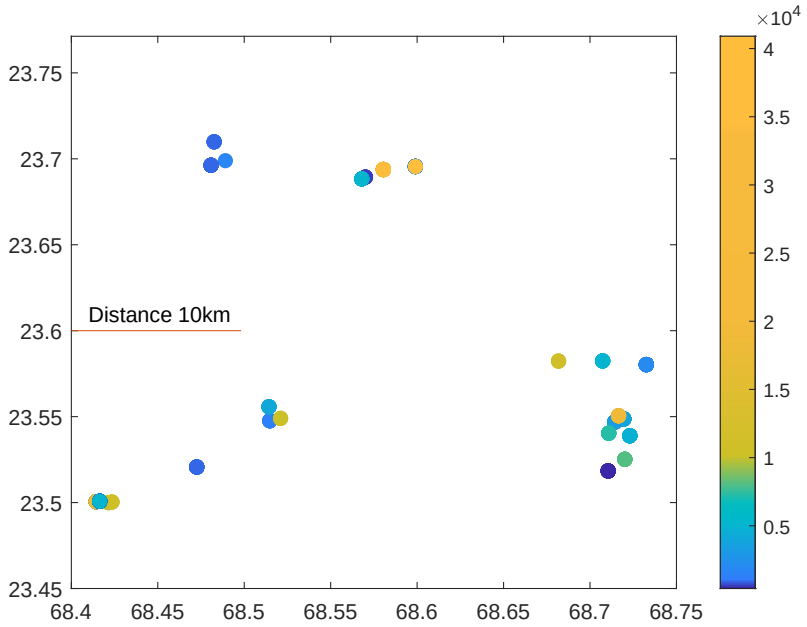


Figure 7.1: Each point represents a location where paleomiddens were collected. The x axis represents longitude while the y axis represents latitude of collected midden. The color of the point represents how long ago the midden was formed, with cooler colors representing recent middens while warmer colors represent older middens.

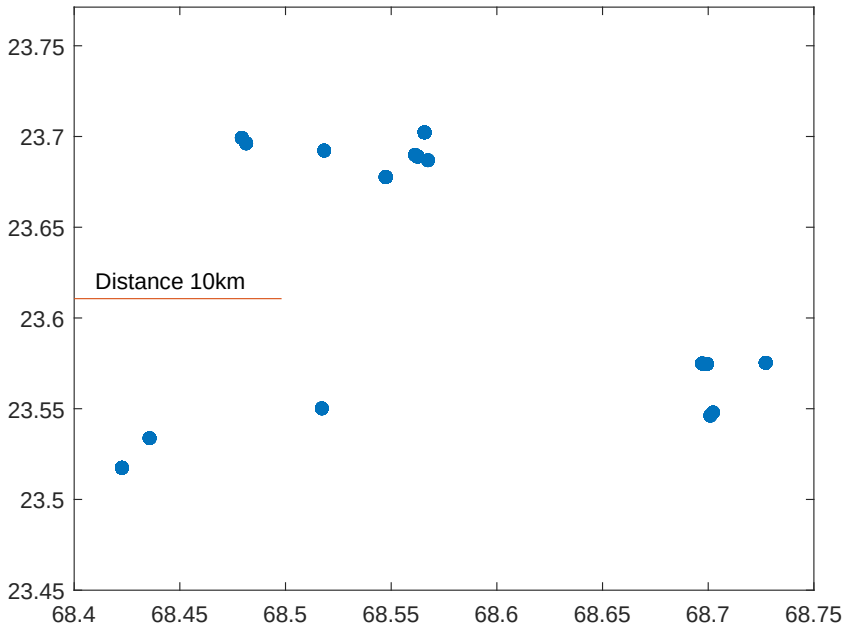


Figure 7.2: Each point represents a location where data for Modern Living or Modern Dead was collected. The x axis represents longitude while the y axis represents latitude of collected midden.

Temporal resolution and species interactions

Another factor in whether species will interact is if they are found at the same time. Due to taphonomic processes, it is possible that two species existed at the same time but it is reflected in the record, or that two species appear to have coexisted in the record when in reality they did not. The paleomiddens collected were created between 400 and 40,000 years ago, with the majority of them being more recent. Table 7.3 shows the number of middens more recent than several different dates. Similar to before we investigate the appearances of two species in the record with respect to time.

Years Ago	# of Midddens
20,000-40,900	3
10,000-20,000	6
5,000-10,000	9
1,000-5,000	19
400-1,000	9

Table 7.3: The dates of the middens found. There are a total of 46 middens ranging from 400 years ago to 40,900 years ago. More middens are found closer to the present, with only 9 being older than 10,000 years ago.

7.3 Approach

The process of inferring species reactions given the last of direct observation can be difficult. While this is an issue in modern ecology as well, it is even more difficult in paleoecology due to the taphonomic and time averaging biases which may be present. There are a number of methods that have been used to recreate ancient networks [144, 42].

One method, proposed by [40], uses ten categories to infer species interactions. The ten categories used are (i) taxonomic uniformitarianism, (ii) functional morphology, (iii) gut contents, (iv) damage patterns, (v) stratigraphic co-occurrence, (vi) body size, (vii) coprolites, (viii) host relationship, (xi) chemical and isotopic signatures and (x) ichnological evidence. Each category is either "yes" or "no", and then link probabilities based on the number of categories present for each species interaction. Using this method they have links of three classes; certain, uncertain, and unlikely. When a potential link has evidence from at least 3 categories, it is classified as certain, if it has evidence from 2 categories it is uncertain, and if it has evidence from only 1 category it is unlikely. Unfortunately the extracted paleomidden data has not been analyzed to the degree that all ten of the categories from [40] can be used. Here the aim is to look specifically at three categories, these are taxonomic uniformitarianism, stratigraphic co-occurrence, and body size. Rather than using "yes" and "no" for the

categories, we consider that each category could have an associated metric, and the overall probability of a link is a weighted sum of the categories.

Spatial Metric:

When we consider spatial information alone, a weighted sum can be used to assess how likely species were to interact with each other. The species (A) will interact with species (B) based on the nearest (B) found to each (A). Here information about the range of species can assist in determining the interaction probability. We use an expected interaction rate that two species have an interaction probability of $\frac{1}{2}$ when they are α meters apart. The probability that (A) interacts with (B) is,

$$int(A,B) = \frac{1}{n} \cdot \left(\frac{1}{\frac{d(A_1,B)}{\alpha}+1} + \frac{1}{\frac{d(A_2,B)}{\alpha}+1} + \frac{1}{\frac{d(A_3,B)}{\alpha}+1} + \cdots + \frac{1}{\frac{d(A_n,B)}{\alpha}+1}\right) \quad (7.1)$$

Where $d(A_i, B)$ is the distance in meters from that location of A_i to the nearest B. The addition of 1 in the denominator assures there will be no division by zero should two species be in the same location. The division by α guarantees the interaction probability of $\frac{1}{2}$ when they are α meters apart. This distance can be adjusted depending on the expected range of the interacting species.

Temporal Metric:

Similar to the spatial data, each sample has an associated date. The dates can be used in a comparable manner to determine species interaction probability. Like the spatial information, we can use a weighted average of the time between two species. The primary difference is that the expected interaction rate is now such that species have an interaction probability of $\frac{1}{2}$ when they are β years apart. The probability

that (A) interacts with (B) is then,

$$int(A,B) = \frac{1}{n} \cdot (\frac{1}{\frac{t(A_1,B)}{\beta}+1} + \frac{1}{\frac{t(A_2,B)}{\beta}+1} + \frac{1}{\frac{t(A_3,B)}{\beta}+1} + \cdots + \frac{1}{\frac{t(A_n,B)}{\beta}+1} \quad (7.2)$$

Spatial-Temporal Metric:

To use both the spatial and temporal distances simultaneously, we created a metric to combine this information. The combined interaction probability is written:

$$int(A,B) = \frac{1}{n} \cdot (\frac{1}{d_{st}(A_1,B)+1} + \cdots + \frac{1}{d_{st}(A_n,B)+1}) \quad (7.3)$$

where $d_{st}(A_i, B)$ defines the closest A_i to B considering a weighted sum of both the distance and time between them. More precisely, $d_{st}(A_i, B) = \min_{b \in B}(\frac{d(A_i,b)}{\alpha} + \frac{t(A_i,b)}{\beta})$, where $d(A_i, b)$ is the distance between A_i and b and $t(A_i, b)$ is the time between A_i and b. Tables 7.4 and 7.5 show how the network changes as α and β are adjusted. Based on this analysis the networks reconstructed in the Results used values of α and β set to 1000 and 500 respectively.

Predator/prey:

We assume that predators will consume prey and possibly other predators, but that prey will not consume other species. The data from paleomiddens contains 43 different species. Five of the species are predators, two species are parasites, and the remaining species are classified as prey. Since the parasites are not parasites for arthropods, they are also classified as prey in the networks.

Body Size:

The size of a predator affects what size of prey it may eat. Prey that are too small may contain too little energy to be worth capturing, while prey that are too large

Meters	# Links	Species #	L/S	L/S^2
1	48	28	1.714	0.061
1000	53	32	1.656	0.052
2000	55	33	1.667	0.051
3000	56	33	1.697	0.051
4000	57	33	1.727	0.052
5000	59	34	1.735	0.051
10000	68	38	1.789	0.047
20000	76	39	1.949	0.050
40000	80	39	2.051	0.053
∞	101	43	2.349	0.055

Table 7.4: Variation of links and species as the spatial window α is varied. When $\alpha = \infty$ only the temporal window has any affect on the network. The temporal window β is fixed at 1000.

may be too difficult to capture. In a study by [19], 80% of predators were larger than their prey. They also found that the body mass ratio between predators and prey increases with as the size of the predator increases, the ratio was also higher for vertebrates than for invertebrates. Additionally for invertebrates they found no habitat specific changes in body mass ratio. Body size can be used as a metric to determine links between predators and their potential prey. The work [19] found that the log_{10} ratio between terrestrial invertebrate predators and their prey to be 0.6 ± 0.03. The exact mechanics for a how body size will affect link probabilities have not yet been determined and will be implemented in future work.

Years	# Links	Species #	L/S	L/S^2
1	43	24	1.792	0.075
500	48	28	1.714	0.061
1000	53	32	1.656	0.052
1500	55	32	1.719	0.054
2000	56	33	1.697	0.051
2500	57	34	1.676	0.049
3000	58	35	1.657	0.047
4000	62	37	1.676	0.045
5000	64	37	1.730	0.047
10000	75	40	1.875	0.047
20000	83	41	2.024	0.049
40000	85	41	2.073	0.051
∞	99	43	2.302	0.054

Table 7.5: Variation of links and species as the temporal window β is varied. When $\beta = \infty$ only the spatial window has any affect on the network. The spatial window α is fixed at 1000.

Reconstruction of Networks:

Networks were created and analyzed at 1,000-year intervals from 40,000 years ago to the newest paleomidden (400 years ago). Although the oldest paleomiddens are more than 40,000 years old, there is only a small amount of data preceding 20,000 years ago. The time range for each species was determined to be their earliest appearance in the paleomiddens to the most recent appearance in the paleomiddens, or in modern communities. One consequence of this time range is that species will not leave and

then return.

There are a number of important metrics in network ecology. Some of these include Clustering Coefficient, Degree Distribution, Average Distance between nodes, and Connectance [39]. This work focuses on connectance, which is the number links per species squared [44, 11]. Other network properties change systematically with connectance [140], and some studies suggest that stability of a network is correlated with connectance [41, 49, 78]. We chose to compare connectance and number of total links for the recreated networks at each 1000 year mark. Other properties are left to further study.

7.4 Results

We compared the networks consisting of only the most significant links. Figure 7.4 shows all of the significant links and the species they connect. This network is time averaged, and shows all species and links that had high probability at any period of time. Figure 7.3 shows all of the potential links, with blue links being unlikely, yellow being uncertain, and red being the most significant.

Table 7.6 lists the number of species, including number of predator and prey, the links in the network, links per species and connectance of the network from 40,000-400 years ago. Upon comparison of connectance and links per species, several patterns emerged. The connectance of the networks fit into three main groupings, more than 10,000 years ago, between 7,000 and 10,000 years ago, and 7,000 to 400 years ago. Upon further investigation, this was found to be due to three main species. These three species of prey do not appear closely in space or time to any predators, and so do not have any links that were the highest significance category. When these species become part of the network, the number of species increases, but the number of links remains the same, which decreases connectance.

Years Ago	Species	Predators	Prey	Links	L/S	Connectance
400	26	22	4	25	0.96	0.037
1000	29	24	5	31	1.07	0.037
4000	31	26	5	35	1.13	0.036
6000	30	25	5	33	1.1	0.037
7000	29	24	5	32	1.1	0.038
8000	27	22	5	32	1.19	0.044
9000	26	21	5	31	1.19	0.046
10000	22	18	4	27	1.23	0.056
13000	21	17	4	27	1.29	0.061
18000	11	10	1	8	0.73	0.066
35000	8	7	1	7	0.88	0.109
36000	4	4	0	0	0	0
40000	1	1	0	0	0	0

Table 7.6: Table showing the connectance of the most significant links every 1000 years. Years that are not present had no change from the previous (older) year.

The connectance found in these networks is lower than the connectance of reconstructed networks in [42], however it is larger than the Messel lake and forest networks of [40]. The work of [39, 42] suggests that modern networks have connectance around 0.10-0.15. This may suggest that our links need refining. Several possible refinements include more accurately indicating the range of specific species, as well as considering a larger number of links significant, such as all links with values > 0.5. Another necessary refinement is to introduce the body size of the species. Unlike the previous refinement, adding body size is expected to result in fewer links, as small species will not be able to consume species that are too large.

7.5 Conclusion

This work presents an approach to recreate networks from rodent paleomiddens. The most important novelty of this method is the use of a spatial-temporal metric which allows a probability of interaction based on how close two species are in time and in space. Differently than [40] this metric allows values other than just “yes” or “no”.

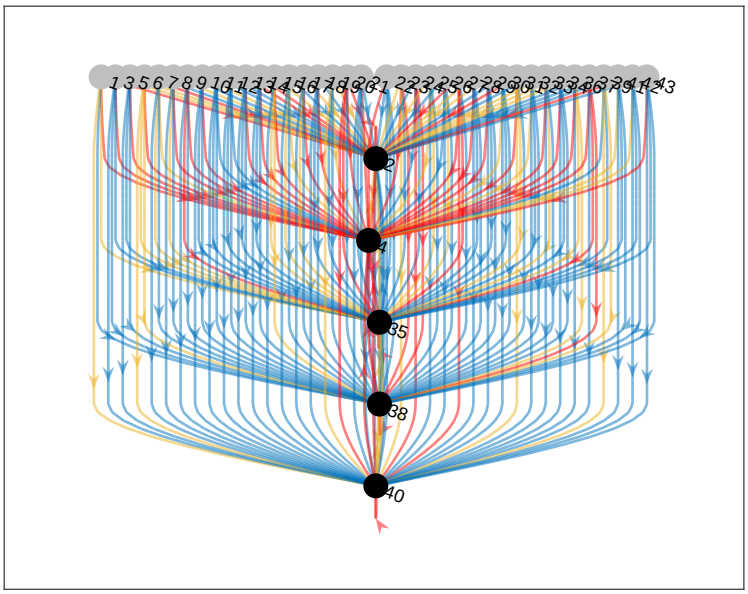


Figure 7.3: Network showing all possible links between species. Black nodes are predators, light nodes are prey. Blue edges are unlikely links ($< \frac{1}{3}$), Yellow edges are uncertain links ($\frac{1}{3} < \frac{2}{3}$), and Red edges are certain links ($> \frac{2}{3}$).

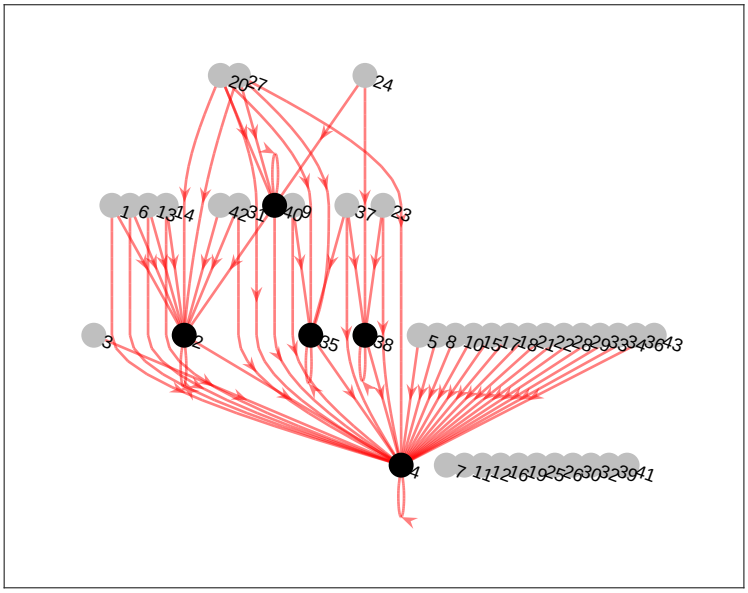


Figure 7.4: Network showing only the most certain links ($> \frac{2}{3}$). Black nodes are predators, light nodes are prey.

In the current work the expected ranges of species were assumed to be the same, however, this can be adjusted to give varying range depending on the expected range of each species. This metric could also be combined with other potential measures (such as body size, damage patterns, etc.) to give a more precise expected interaction probability. The connectance of the recreated networks was compared with networks from [42, 40] and were similar connectances found in these works, however, not in the range of modern network connectance suggested in [39]. This may be due to the fact we chose only the links which were most significant to be included in our recreated networks. Tuning of what percentage determines a "significant" link could bring the connectance into the desired range, however, this analysis has not been performed. Further comparison in additional network metrics, such as centrality and average path length, is needed to better determine how realistic these recreations are. This method provides a tool for midden network reconstruction that can also be applied to other midden systems in other locations.

Bibliography

[1] U. Abraham, A. E. Granada, P. O. Westermark, M. Heine, A. Kramer, and H. Herzel, *Coupling governs entrainment range of circadian clocks*, Molecular systems biology, 6 (2010).

[2] R. Allen, T. R. Rieger, and C. J. Musante, *Efficient generation and selection of virtual populations in quantitative systems pharmacology models*, CPT: pharmacometrics & systems pharmacology, 5 (2016), pp. 140–146.

[3] S. Allerheiligen, *Next-generation model-based drug discovery and development: quantitative and systems pharmacology.*, Clinical Pharmacology & Therapeutics, 88 (2010).

[4] U. Alon, *An Introduction to Systems Biology*, Chapman and Hall/CRC, 2006.

[5] Z. An, N. J. Merrill, S. T. McQuade, and B. Piccoli, *Equilibria and control of metabolic networks with enhancers and inhibitors*, Mathematics in Engineering, 1 (2019), p. 648.

[6] J. Aschoff, *Freerunning and entrained circadian rhythms*, in Biological rhythms, Springer, 1981, pp. 81–93.

[7] J. Aschoff and H. Pohl, *Phase relations between a circadian rhythm and its zeitgeber within the range of entrainment*, Naturwissenschaften, 65 (1978), pp. 80–84.

[8] A. Asghar, A. Raman, A. Aziz, and W. M. A. W. Daud, *A comparison of central composite design and taguchi method for optimizing fenton process*, The Scientific World Journal, 2014 (2014).

[9] B. Auclair, D. E. Nix, R. D. Adam, G. T. James, and C. A. Peloquin, *Pharmacokinetics of ethionamide administered under fasting conditions or with orange juice, food, or antacids*, Antimicrobial agents and chemotherapy, 45 (2001), pp. 810–814.

[10] G. Balázsi, A. P. Heath, L. Shi, and M. L. Gennaro, *The temporal response of the mycobacterium tuberculosis gene regulatory network during growth arrest*, Molecular systems biology, 4 (2008).

[11] A. P. Beckerman, O. L. Petchey, and P. H. Warren, *Foraging biology predicts food web complexity*, Proceedings of the National Academy of Sciences, 103 (2006), pp. 13745–13749.

[12] N. Biggs, N. L. Biggs, and E. N. Biggs, *Algebraic graph theory*, vol. 67, Cambridge university press, 1993.

[13] K. Birkhofer, H. Bylund, P. Dalin, O. Ferlian, V. Gagic, P. A. Hambäck, M. Klapwijk, L. Mestre, E. Roubinet, M. Schroeder, et al., *Methods to identify the prey of invertebrate predators in terrestrial field studies*, Ecology and evolution, 7 (2017), pp. 1942–1953.

[14] L. J. Bonnett, G. Ken-Dror, G. C. Koh, and G. R. Davies, *Comparing the efficacy of drug regimens for pulmonary tuberculosis: Meta-analysis of endpoints in early-phase clinical trials*, Clinical Infectious Diseases, 65 (2017), pp. 46–54.

[15] G. Bordyugov, U. Abraham, A. Granada, P. Rose, K. Imkeller, A. Kramer, and H. Herzel, *Tuning the phase of circadian entrainment*, Journal of The Royal Society Interface, 12 (2015), p. 20150282.

[16] H. I. Boshoff, T. G. Myers, B. R. Copp, M. R. McNeil, M. A. Wilson, and C. E. Barry, *The transcriptional responses of mycobacterium tuberculosis to inhibitors of metabolism novel insights into drug mechanisms of action*, Journal of Biological Chemistry, 279 (2004), pp. 40174–40184.

[17] A. Bressan and B. Piccoli, *Introduction to mathematical control theory, 2007*, AIMS series on applied mathematics, Philadelphia, (2007).

[18] A. Bretto, *Hypergraph theory: An introduction*, Mathematical Engineering. Cham: Springer, (2013).

[19] U. Brose, T. Jonsson, E. L. Berlow, P. Warren, C. Banasek-Richter, L.-F. Bersier, J. L. Blanchard, T. Brey, S. R. Carpenter, M.-F. C. Blandenier, et al., *Consumer–resource body-size relationships in natural food webs*, Ecology, 87 (2006), pp. 2411–2417.

[20] F. Bullo, *Lectures on network systems*, Online at http://motion. me. ucsb. edu/book-lns, with contributions by J. Cortes, F. Dorfler and S. Martinez, (2016).

[21] F. Bullo, *Lectures on network systems*, Kindle Direct Publishing, 2019. Edition 1.

[22] A. Cappuccio, F. Castiglione, B. Piccoli, and V. Tozzi, *Evaluation of HIV-1 and CD4+ T cell dynamic parameters in patients treated with genotypic resistance testing-guided HAART*, Current HIV research, 6 (2008), pp. 363–369.

[23] F. Castiglione and B. Piccoli, *Optimal control in a model of dendritic cell transfection cancer immunotherapy*, Bulletin of Mathematical Biology, 68 (2006), pp. 255–274.

[24] ——, *Cancer immunotherapy, mathematical modeling and optimal control*, Journal of theoretical Biology, 247 (2007), pp. 723–732.

[25] J. S. Caughman and J. Veerman, *Kernels of directed graph laplacians*, the electronic journal of combinatorics, 13 (2006), p. 39.

[26] N. A. Cilfone, C. R. Perry, D. E. Kirschner, and J. J. Linderman, *Multi-scale modeling predicts a balance of tumor necrosis factor-α and interleukin-10 controls the granuloma environment during mycobacterium tuberculosis infection*, PloS one, 8 (2013), p. e68680.

[27] E. Cinlar, *Introduction to stochastic processes*, Courier Corporation, 2013.

[28] M. Cokol, N. Kuru, E. Bicak, J. Larkins-Ford, and B. B. Aldridge, *Efficient measurement and factorization of high-order drug interactions in mycobacterium tuberculosis*, Science advances, 3 (2017), p. e1701881.

[29] H. Cordes, C. Thiel, H. E. Aschmann, V. Baier, L. M. Blank, and L. Kuepfer, *A physiologically based pharmacokinetic model of isoniazid and its application in individualizing tuberculosis chemotherapy*, Antimicrobial agents and chemotherapy, 60 (2016), pp. 6134–6145.

[30] C. Darrah, B. L. Taylor, K. D. Edwards, P. E. Brown, A. Hall, and H. G. McWatters, *Analysis of phase of luciferase expression reveals novel circadian quantitative trait loci in arabidopsis*, Plant physiology, 140 (2006), pp. 1464–1474.

[31] P. De Leenheer, *The zero deficiency theorem*, in Notes for the Biomath Seminar I–MAP6487, Fall, vol. 9, 2009.

[32] J. Delahaye, C. R. Plumlee, C. Plaisier, N. Baliga, D. R. Sherman, and K. B. Urdahl, *Alveolar and monocyte-derived macrophages differentially engage antibacterial programs during adaptive immunity to mycobacterium tuberculosis.*, 2018.

[33] D. Deshpande, S. Srivastava, C. Meek, R. Leff, G. S. Hall, and T. Gumbo, *Moxifloxacin pharmacokinetics/pharmacodynamics and optimal dose and susceptibility breakpoint identification for treatment of disseminated mycobacterium avium infection*, Antimicrobial agents and chemotherapy, 54 (2010), pp. 2534–2539.

[34] O. Dezerald, C. Latorre, J. L. Betancourt, G. A. B. Vera, and A. L. González, *Ecological fidelity and spatiotemporal resolution of arthropod death assemblages from rodent middens in the central atacama desert (northern chile)*, Quaternary Science Reviews, 210 (2019), pp. 15–25.

[35] J. F. Duffy, S. W. Cain, A.-M. Chang, A. J. Phillips, M. Y. Münch, C. Gronfier, J. K. Wyatt, D.-J. Dijk, K. P. Wright, and C. A. Czeisler, *Sex difference in the near-24-hour intrinsic period of the human circadian timing system*, Proceedings of the National Academy of Sciences, 108 (2011), pp. 15602–15608.

[36] J. F. Duffy and C. A. Czeisler, *Age-related change in the relationship between circadian period, circadian phase, and diurnal preference in humans*, Neuroscience letters, 318 (2002), pp. 117–120.

[37] J. F. Duffy, D.-J. Dijk, E. B. Klerman, and C. A. Czeisler, *Later endogenous circadian temperature nadir relative to an earlier wake time in older people*, American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 275 (1998), pp. R1478–R1487.

[38] J. C. Dunlap and J. J. Loros, *Making time: conservation of biological clocks from fungi to animals*, Microbiology spectrum, 5 (2017).

[39] J. A. Dunne, *The network structure of food webs*, Ecological networks: linking structure to dynamics in food webs, (2006), pp. 27–86.

[40] J. A. Dunne, C. C. Labandeira, and R. J. Williams, *Highly resolved early eocene food webs show development of modern trophic structure after the end-cretaceous extinction*, Proceedings of the Royal Society B: Biological Sciences, 281 (2014), p. 20133280.

[41] J. A. Dunne, R. J. Williams, and N. D. Martinez, *Network structure and biodiversity loss in food webs: robustness increases with connectance*, Ecology letters, 5 (2002), pp. 558–567.

[42] J. A. Dunne, R. J. Williams, N. D. Martinez, R. A. Wood, and D. H. Erwin, *Compilation and network analyses of cambrian food webs*, PLoS biology, 6 (2008).

[43] R. Edgar, M. Domrachev, and A. E. Lash, *Gene expression omnibus: Ncbi gene expression and hybridization array data repository*, Nucleic acids research, 30 (2002), pp. 207–210.

[44] A. Eklöf and B. Ebenman, *Species loss and secondary extinctions in simple and complex model communities*, Journal of animal ecology, 75 (2006), pp. 239–246.

[45] M. Feinberg and F. J. Horn, *Dynamics of open chemical systems and the algebraic structure of the underlying reaction network*, Chemical Engineering Science, 29 (1974), pp. 775–787.

[46] G. A. for TB Drug Development, *Handbook of anti-tuberculosis agents*, Tuberculosis (Edinb), 88 (2008), pp. 85–6.

[47] L. R. Ford and D. R. Fulkerson, *Maximal flow through a network*, Canadian journal of Mathematics, 8 (1956), pp. 399–404.

[48] C. Friedrich, *A model qualification method for mechanistic physiological qsp models to support model-informed drug development*, CPT: pharmacometrics & systems pharmacology, 5 (2016), pp. 43–53.

[49] G. F. Fussmann and G. Heber, *Food web complexity and chaotic population dynamics*, Ecology Letters, 5 (2002), pp. 394–401.

[50] K. Gadkar, N. Budha, A. Baruch, J. Davis, P. Fielder, and S. Ramanujan, *A mechanistic systems pharmacology model for prediction of ldl cholesterol lowering by pcsk9 antagonism in human dyslipidemic populations*, CPT: pharmacometrics & systems pharmacology, 3 (2014), pp. 1–9.

[51] J. E. Galagan, K. Minch, M. Peterson, A. Lyubetskaya, E. Azizi, L. Sweet, A. Gomes, T. Rustad, G. Dolganov, I. Glotova, et al., *The mycobacterium tuberculosis regulatory network and hypoxia*, Nature, 499 (2013), p. 178.

[52] D. Gale, H. W. Kuhn, and A. W. Tucker, *Linear programming and the theory of games*, Activity analysis of production and allocation, 13 (1951), pp. 317–335.

[53] E. P. Gianchandani, A. K. Chavali, and J. A. Papin, *The application of flux balance analysis in systems biology*, Wiley Interdisciplinary Reviews: Systems Biology and Medicine, 2 (2010), pp. 372–382.

[54] A. M. Ginsberg, M. W. Laurenzi, D. J. Rouse, K. D. Whitney, and M. K. Spigelman, *Safety, tolerability, and pharmacokinetics of pa-824 in healthy subjects*, Antimicrobial agents and chemotherapy, 53 (2009), pp. 3720–3725.

[55] A. E. Granada, G. Bordyugov, A. Kramer, and H. Herzel, *Human chronotypes from a theoretical perspective*, PLoS One, 8 (2013), p. e59464.

[56] A. E. Granada and H. Herzel, *How to achieve fast entrainment? the timescale to synchronization*, PLoS One, 4 (2009), p. e7057.

[57] D. Gravel, T. Poisot, C. Albouy, L. Velez, and D. Mouillot, *Inferring food web structure from predator–prey body size relationships*, Methods in Ecology and Evolution, 4 (2013), pp. 1083–1090.

[58] C. Gronfier, K. P. Wright, R. E. Kronauer, and C. A. Czeisler, *Entrainment of the human circadian pacemaker to longer-than-24-h days*, Proceedings of the National Academy of Sciences, 104 (2007), pp. 9081–9086.

[59] S. M. Grundy, N. J. Stone, A. L. Bailey, C. Beam, K. K. Birtcher, R. S. Blumenthal, L. T. Braun, S. de Ferranti, J. Faiella-Tommasino, D. E. Forman, et al., *2018 aha/acc/aacvpr/aapa/abc/acpm/ada/ags/apha/aspc/nla/pcna guideline on the management of blood cholesterol: a report of the american college of cardiology/american heart association task force on clinical practice guidelines*, Journal of the American College of Cardiology, 73 (2019), pp. e285–e350.

[60] J. Gunawardena, *A linear framework for time-scale separation in nonlinear biochemical systems*, PloS one, 7 (2012), p. e36321.

[61] G. T. Heineman, G. Pollice, and S. Selkow, *Algorithms in a nutshell, chapter 8: network flow algorithms*, 2009.

[62] C. Heintzen and Y. Liu, *The neurospora crassa circadian clock*, Advances in genetics, 58 (2007), pp. 25–66.

[63] A. Hida, Y. Ohsawa, S. Kitamura, K. Nakazaki, N. Ayabe, Y. Motomura, K. Matsui, M. Kobayashi, A. Usui, Y. Inoue, et al., *Evaluation of circadian phenotypes utilizing fibroblasts from patients with circadian rhythm sleep disorders*, Translational psychiatry, 7 (2017), p. e1106.

[64] I. Hosseini and F. Mac Gabhann, *Mechanistic models predict efficacy of ccr5-deficient stem cell transplants in hiv patient populations*, CPT: pharmacometrics & systems pharmacology, 5 (2016), pp. 82–90.

[65] J. A. Jacquez and C. P. Simon, *Qualitative theory of compartmental systems*, Siam Review, 35 (1993), pp. 43–79.

[66] K. A. Johnson and R. S. Goody, *The original michaelis constant: translation of the 1913 michaelis–menten paper*, Biochemistry, 50 (2011), pp. 8264–8269.

[67] C. R. Jones, S. S. Campbell, S. E. Zone, F. Cooper, A. DeSano, P. J. Murphy, B. Jones, L. Czajkowski, and L. J. Ptček, *Familial advanced sleep-phase syndrome: A short-period circadian rhythm variant in humans*, Nature medicine, 5 (1999), pp. 1062–1065.

[68] C. Joslyn and K. Nowak, *Ubergraphs: A definition of a recursive hypergraph structure*, arXiv preprint arXiv:1704.05547, (2017).

[69] P. C. Kandalepas, J. W. Mitchell, and M. U. Gillette, *Melatonin signal transduction pathways require e-box-mediated transcription of per1 and per2 to reset the scn clock at dusk*, PloS one, 11 (2016), p. e0157824.

[70] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima, *KEGG: new perspectives on genomes, pathways, diseases and drugs*, Nucleic Acids Research, 45 (2016), pp. D353–D361.

[71] M. Kanehisa and S. Goto, *Kegg: kyoto encyclopedia of genes and genomes*, Nucleic acids research, 28 (2000), pp. 27–30.

[72] ——, *KEGG: Kyoto Encyclopedia of Genes and Genomes*, Nucleic Acids Research, 28 (2000), pp. 27–30.

[73] M. Kanehisa, Y. Sato, M. Furumichi, K. Morishima, and M. Tanabe, *New approach for understanding genome variations in KEGG*, Nucleic Acids Research, 47 (2018), pp. D590–D595.

[74] K. J. Kauffman, P. Prakash, and J. S. Edwards, *Advances in flux balance analysis*, Current opinion in biotechnology, 14 (2003), pp. 491–496.

[75] S. L. Kendall, M. Withers, C. N. Soffair, N. J. Moreland, S. Gurcha, B. Sidders, R. Frita, A. Ten Bokum, G. S. Besra, J. S. Lott, et al., *A highly conserved transcriptional repressor controls a large regulon involved in lipid degradation in mycobacterium smegmatis and mycobacterium tuberculosis*, Molecular microbiology, 65 (2007), pp. 684–699.

[76] A. I. Khuri and S. Mukhopadhyay, *Response surface methodology*, Wiley Interdisciplinary Reviews: Computational Statistics, 2 (2010), pp. 128–149.

[77] D. J. Klinke and S. D. Finley, *Timescale analysis of rule-based biochemical reaction networks*, Biotechnology progress, 28 (2012), pp. 33–44.

[78] M. Kondoh, *Foraging adaptation and the relationship between food-web complexity and stability*, Science, 299 (2003), pp. 1388–1391.

[79] S. Kou, B. J. Cherayil, W. Min, B. P. English, and X. S. Xie, *Single-molecule michaelis- menten equations*, 2005.

[80] R. E. Kronauer, C. A. Czeisler, S. F. Pilato, M. C. Moore-Ede, and E. D. Weitzman, *Mathematical model of the human circadian system with two interacting oscillators*, American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 242 (1982), pp. R3–R17.

[81] Y. Kuramoto, *Cooperative dynamics of oscillator communitya study based on lattice of rings*, Progress of Theoretical Physics Supplement, 79 (1984), pp. 223–240.

[82] P. Kurien, P.-K. Hsu, J. Leon, D. Wu, T. McMahon, G. Shi, Y. Xu, A. Lipzen, L. A. Pennacchio, C. R. Jones, et al., *Timeless mutation alters phase responsiveness and causes advanced sleep phase*, Proceedings of the National Academy of Sciences, 116 (2019), pp. 12045–12053.

[83] L. Lalande, L. Bourguignon, P. Maire, and S. Goutelle, *Mathematical modeling and systems pharmacology of tuberculosis: Isoniazid as a case study*, Journal of theoretical biology, 399 (2016), pp. 43–52.

[84] M. K. Lau, S. R. Borrett, B. Baiser, N. J. Gotelli, and A. M. Ellison, *Ecological network metrics: opportunities for synthesis*, Ecosphere, 8 (2017), p. e01900.

[85] J. M. Lee, E. P. Gianchandani, and J. A. Papin, *Flux balance analysis in the era of metabolomics*, Briefings in bioinformatics, 7 (2006), pp. 140–150.

[86] K. Lee, P. Shiva Kumar, S. McQuade, J. Y. Lee, S. Park, Z. An, and B. Piccoli, *Experimental and mathematical analyses relating circadian period and phase of entrainment in neurospora crassa*, Journal of biological rhythms, 32 (2017), pp. 550–559.

[87] A. J. Lewy, B. P. Hasler, J. S. Emens, and R. L. Sack, *Pretreatment circadian period in free-running blind people may predict the phase angle of entrainment to melatonin*, Neuroscience letters, 313 (2001), pp. 158–160.

[88] J. J. Loros and J. C. Dunlap, *Genetic and molecular analysis of circadian rhythms in n eurospora*, Annual review of physiology, 63 (2001), pp. 757–794.

[89] P. L. Lowrey, K. Shimomura, M. P. Antoch, S. Yamazaki, P. D. Zemenides, M. R. Ralph, M. Menaker, and J. S. Takahashi, *Positional syntenic cloning and functional characterization of the mammalian circadian mutation tau*, Science, 288 (2000), pp. 483–491.

[90] M. A. Lyons, *Modeling and simulation of pretomanid pharmacokinetics in pulmonary tuberculosis patients*, Antimicrobial agents and chemotherapy, 62 (2018), pp. e02359–17.

[91] S. Ma, K. J. Minch, T. R. Rustad, S. Hobbs, S.-L. Zhou, D. R. Sherman, and N. D. Price, *Integrated modeling of gene regulatory and metabolic networks in mycobacterium tuberculosis*, PLoS computational biology, 11 (2015), p. e1004543.

[92] H. Maeda, S. Kodama, and Y. Ohta, *Asymptotic behavior of nonlinear compartmental systems: nonoscillation and stability*, IEEE Transactions on Circuits and Systems, 25 (1978), pp. 372–378.

[93] V. M. Malhotra, M. P. Kumar, and S. N. Maheshwari, *An o (| v| 3) algorithm for finding maximum flows in networks*, Information Processing Letters, 7 (1978), pp. 277–278.

[94] M. T. Mc Auley, D. J. Wilkinson, J. J. Jones, and T. B. Kirkwood, *A whole-body mathematical model of cholesterol metabolism and its age-associated dysregulation*, BMC systems biology, 6 (2012), p. 1.

[95] H. McIlleron, P. Wash, A. Burger, J. Norman, P. I. Folb, and P. Smith, *Determinants of rifampin, isoniazid, pyrazinamide, and ethambutol pharmacokinetics in a cohort of tuberculosis patients*, Antimicrobial agents and chemotherapy, 50 (2006), pp. 1170–1177.

[96] S. T. McQuade, R. E. Abrams, J. S. Barrett, B. Piccoli, and K. Azer, *Linear-in-flux-expressions methodology: Toward a robust mathematical framework for quantitative systems pharmacology simulators*, Gene regulation and systems biology, 11 (2017), p. 1177625017711414.

[97] S. T. McQuade, Z. An, N. J. Merrill, R. E. Abrams, K. Azer, and B. Piccoli, *Equilibria for large metabolic systems and the life approach*, in 2018 Annual American Control Conference (ACC), IEEE, 2018, pp. 2005–2010.

[98] S. T. McQuade, N. J. Merrill, and B. Piccoli, *Metabolic graphs, life method and the modeling of drug action on mycobacterium tuberculosis*, 2020.

[99] N. J. Merrill, Z. An, S. T. McQuade, F. Garin, K. Azer, R. Abrams, and B. Piccoli, *Stability of metabolic networks via linear-in-flux-expressions*, Networks & Heterogeneous Media, 14 (2019), pp. 101–130.

[100] C. D. Meyer, *Matrix analysis and applied linear algebra*, vol. 71, Siam, 2000.

[101] T. P. Michael, P. A. Salome, J. Y. Hannah, T. R. Spencer, E. L. Sharp, M. A. McPeek, J. M. Alonso, J. R. Ecker, and C. R. McClung, *Enhanced fitness conferred by naturally occurring variation in the circadian clock*, Science, 302 (2003), pp. 1049–1053.

[102] I. Mirzaev and J. Gunawardena, *Laplacian dynamics on general graphs*, Bulletin of mathematical biology, 75 (2013), pp. 2118–2149.

[103] H. Nakajima, *Tuberculosis: a global emergency*, World Health, 46 (1993), p. 3.

[104] D. E. Nix, R. D. Adam, B. Auclair, T. S. Krueger, P. G. Godo, and C. A. Peloquin, *Pharmacokinetics and relative bioavailability of clofazimine in relation to food, orange juice and antacid*, Tuberculosis, 84 (2004), pp. 365–373.

[105] W. H. Organization et al., *Prioritization of pathogens to guide discovery, research and development of new antibiotics for drug-resistant bacterial infections, including tuberculosis*, tech. rep., World Health Organization, 2017.

[106] J. D. Orth, I. Thiele, and B. O. Palsson, *What is flux balance analysis?*, Nature biotechnology, 28 (2010), p. 245.

[107] B. Palsson, *Systems biology*, Cambridge university press, 2015.

[108] V. I. Pérez-Nueno, *Using quantitative systems pharmacology for novel drug discovery*, Expert opinion on drug discovery, 10 (2015), pp. 1315–1331.

[109] E. J. Peterson, D. J. Reiss, S. Turkarslan, K. J. Minch, T. Rustad, C. L. Plaisier, W. J. Longabaugh, D. R. Sherman, and N. S. Baliga, *A high-resolution network model for global gene regulation in mycobacterium tuberculosis*, Nucleic acids research, 42 (2014), pp. 11291–11303.

[110] A. J. Phillips, P. Vidafar, A. C. Burns, E. M. McGlashan, C. Anderson, S. M. Rajaratnam, S. W. Lockley, and S. W. Cain, *High sensitivity and interindividual variability in the response of the human circadian system to evening light*, Proceedings of the National Academy of Sciences, 116 (2019), pp. 12019–12024.

[111] K. E. Pickerill, J. A. Paladino, and J. J. Schentag, *Comparison of the fluoroquinolones based on pharmacokinetic and pharmacodynamic parameters*, Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy, 20 (2000), pp. 417–428.

[112] M. M. Pires, *Rewilding ecological communities and rewiring ecological networks*, Perspectives in ecology and conservation, 15 (2017), pp. 257–265.

[113] C. S. Pittendrigh, W. T. Kyner, and T. Takamura, *The amplitude of circadian oscillations: temperature dependence, latitudinal clines, and the photoperiodic time measurement*, Journal of Biological Rhythms, 6 (1991), pp. 299–313.

[114] T. Poisot, A. R. Cirtwill, K. Cazelles, D. Gravel, M.-J. Fortin, and D. B. Stouffer, *The structure of probabilistic networks*, Methods in Ecology and Evolution, 7 (2016), pp. 303–312.

[115] T. Poisot, D. B. Stouffer, and D. Gravel, *Beyond species: why ecological interaction networks vary through space and time*, Oikos, 124 (2015), pp. 243–251.

[116] T. Poisot, D. B. Stouffer, and S. Kéfi, *Describe, understand and predict: why do we need networks in ecology?*, Functional Ecology, 30 (2016), pp. 1878–1882.

[117] K. Raman and N. Chandra, *Flux balance analysis of biological systems: applications and challenges*, Briefings in bioinformatics, 10 (2009), pp. 435–449.

[118] B. Reisfeld, C. Metzler, M. Lyons, A. Mayeno, E. Brooks, and M. Degroote, *A physiologically based pharmacokinetic model for capreomycin*, Antimicrobial agents and chemotherapy, 56 (2012), pp. 926–934.

[119] T. Roenneberg, S. Daan, and M. Merrow, *The art of entrainment*, Journal of Biological Rhythms, 18 (2003), pp. 183–194.

[120] T. Roenneberg and M. Merrow, *The circadian clock and human health*, Current biology, 26 (2016), pp. R432–R443.

[121] M. Rogers, P. Lyster, and R. Okita, *Nih support for the emergence of quantitative and systems pharmacology*, CPT Pharmacometrics Syst Pharmacol, 2 (2013), p. e37.

[122] P. D. Roopnarine, G. Dietl, and K. Flessa, *Ecological modeling of paleocommunity food webs*, (2009).

[123] J. P. Sarathy, L. E. Via, D. Weiner, L. Blanc, H. Boshoff, E. A. Eugenin, C. E. Barry, and V. A. Dartois, *Extreme drug tolerance of mycobacterium tuberculosis in caseum*, Antimicrobial agents and chemotherapy, 62 (2018), pp. e02266–17.

[124] D. Saunders, S. Gillanders, and R. Lewis, *Light-pulse phase response curves for the locomotor activity rhythm in period mutants of drosophila melanogaster*, Journal of Insect Physiology, 40 (1994), pp. 957–968.

[125] C. H. Schilling, D. Letscher, and B. O. Palsson, *Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective*, Journal of theoretical biology, 203 (2000), pp. 229–248.

[126] B. J. Schmidt, F. P. Casey, T. Paterson, and J. R. Chan, *Alternate virtual populations elucidate the type i interferon signature predictive of the response to rituximab in rheumatoid arthritis*, BMC bioinformatics, 14 (2013), p. 1.

[127] J. P. Scott-Browne, S. Shafiani, G. Tucker-Heard, K. Ishida-Tsubota, J. D. Fontenot, A. Y. Rudensky, M. J. Bevan, and K. B. Urdahl, *Expansion and function of foxp3-expressing t regulatory cells during tuberculosis*, The Journal of experimental medicine, 204 (2007), pp. 2159–2169.

[128] L. Shi, C. D. Sohaskey, C. Pfeiffer, P. Datta, M. Parks, J. McFadden, R. J. North, and M. L. Gennaro, *Carbon flux rerouting during mycobacterium tuberculosis growth arrest*, Molecular microbiology, 78 (2010), pp. 1199–1215.

[129] P. K. Sorger, S. R. Allerheiligen, D. R. Abernethy, R. B. Altman, K. L. Brouwer, A. Califano, D. Z. D'Argenio, R. Iyengar, W. J. Jusko, R. Lalonde, et al., *Quantitative and systems pharmacology in the post-genomic era: new approaches to discovering drugs and understanding therapeutic mechanisms*, in An NIH white paper by the QSP workshop group, NIH Bethesda, 2011, pp. 1–48.

[130] S. Srivastava, C. Modongo, C. W. S. Dona, J. G. Pasipanodya, D. Deshpande, and T. Gumbo, *Amikacin optimal exposure targets in the hollow-fiber system model of tuberculosis*, Antimicrobial agents and chemotherapy, 60 (2016), pp. 5922–5927.

[131] H. Stass and D. Kubitza, *Pharmacokinetics and elimination of moxifloxacin after oral and intravenous administration in man*, Journal of Antimicrobial Chemotherapy, 43 (1999), pp. 83–90.

[132] J. Stelling, E. D. Gilles, and F. J. Doyle, *Robustness properties of circadian clock architectures*, Proceedings of the National Academy of Sciences, 101 (2004), pp. 13210–13215.

[133] G. R. Stewart, V. A. Snewin, G. Walzl, T. Hussell, P. Tormay, P. O'Gaora, M. Goyal, J. Betts, I. N. Brown, and D. B. Young, *Overexpression of heat-shock proteins reduces survival of mycobacterium tuberculosis in the chronic phase of infection*, Nature medicine, 7 (2001), p. 732.

[134] T.-L. To, M. A. Henson, E. D. Herzog, and F. J. Doyle III, *A molecular model for intercellular synchronization in the mammalian circadian clock*, Biophysical journal, 92 (2007), pp. 3792–3803.

[135] A. Tostmann, C. M. Mtabho, H. H. Semvua, J. van den Boogaard, G. S. Kibiki, M. J. Boeree, and R. E. Aarnoutse, *Pharmacokinetics of first-line tuberculosis drugs in tanzanian patients*, Antimicrobial agents and chemotherapy, 57 (2013), pp. 3208–3213.

[136] V. G. Tusher, R. Tibshirani, and G. Chu, *Significance analysis of microarrays applied to the ionizing radiation response*, Proceedings of the National Academy of Sciences, 98 (2001), pp. 5116–5121.

[137] A. Van der Schaft, S. Rao, and B. Jayawardhana, *A network dynamics approach to chemical reaction networks*, International Journal of Control, 89 (2016), pp. 731–745.

[138] V. I. Voloshin, *Introduction to graph and hypergraph theory*, Nova Science Publ., 2009.

[139] B. P. White, B. Lomaestro, and M. P. Pai, *Optimizing the initial amikacin dosage in adults*, Antimicrobial agents and chemotherapy, 59 (2015), pp. 7094–7096.

[140] R. J. Williams, E. L. Berlow, J. A. Dunne, A.-L. Barabási, and N. D. Martinez, *Two degrees of separation in complex food webs*, Proceedings of the National Academy of Sciences, 99 (2002), pp. 12913–12916.

[141] K. P. Wright, R. J. Hughes, R. E. Kronauer, D.-J. Dijk, and C. A. Czeisler, *Intrinsic near-24-h pacemaker period determines limits of circadian entrainment to a weak synchronizer in humans*, Proceedings of the National Academy of Sciences, 98 (2001), pp. 14027–14032.

[142] W. Xia and M. Cao, *Analysis and applications of spectral properties of grounded laplacian matrices for directed networks*, Automatica, 80 (2017), pp. 10–16.

[143] Y. Xu, Q. S. Padiath, R. E. Shapiro, C. R. Jones, S. C. Wu, N. Saigoh, K. Saigoh, L. J. Ptáček, and Y.-H. Fu, *Functional consequences of a ckiδ*

mutation causing familial advanced sleep phase syndrome, Nature, 434 (2005), pp. 640–644.

[144] J. D. Yeakel, M. M. Pires, L. Rudolf, N. J. Dominy, P. L. Koch, P. R. Guimarães, and T. Gross, *Collapse of an ecological network in ancient egypt*, Proceedings of the National Academy of Sciences, 111 (2014), pp. 14472–14477.

[145] W. A. Zehring, D. A. Wheeler, P. Reddy, R. J. Konopka, C. P. Kyriacou, M. Rosbash, and J. C. Hall, *P-element transformation with period locus dna restores rhythmicity to mutant, arrhythmic drosophila melanogaster*, Cell, 39 (1984), pp. 369–376.

[146] M. Zhu, R. Namdar, J. Stambaugh, J. Starke, A. Bulpitt, S. Berning, and C. Peloquin, *Population pharmacokinetics of ethionamide in patients with tuberculosis*, Tuberculosis, 82 (2002), pp. 91–96.