UNIVERSITÄT
POTSDAM

A model reduction approach to finding robustness in large-scale biochemical reaction networks

Chris Kindler

WiSe 2017/2018

Selbstständigkeitserklärung

Hiermit erkläre ich, daß ich die vorliegende Arbeit selbstständig angefertigt, nicht anderweitig zu Prüfungszwecken vorgelegt und keine anderen als die angegebenen Hilfsmittel verwendet habe. Sämtliche wissentlich verwendete Textausschnitte, Zitate oder Inhalte anderer Verfasser wurden ausdrücklich als solche gekennzeichnet.

Potsdam, 11. März 2018	C. Kindler
Ort, Datum	Chris Kindler

Zusammenfassung

Robustheit von biochemischen Systemen, i.e. ihre Kapazität unter diversen umweltlichen Bedingungen normale Funktionalität zu erhalten, ist ein Schwerpunkt der Systemanylyse in Bioinformatik und Systembiologie. Die Art der Robustheit, welche wir in dieser Arbeit besprechen werden fokussiert sich auf Spezieskonzentrationen. Eine Robuste Spezies ist solch eine, die für jeden Gleichgewichtszustand ihre Konzentration beibehält. Vorhandene Methoden zur Erkennung von robusten Spezies in biochemischen Reaktionsnetzwerken sind nicht auf große Netzwerke anwendbar. In dieser Arbeit werden wir mit Hilfe von Modellreduzierung einen Ansatz schaffen, welcher es erlaubt auch in solchen Netzwerken robuste Spezies zu finden. Der Aufbau der Arbeit ist wie folgt. Nach einer kurzen Einführung in die Systembiologie und die vielen mathematischen Anwendungen darin werden wir die nötige allgemeine Theorie der Reaktionsnetzwerke besprechen. Danach werden wir die existierenden Ansätze zum Finden robuster Spezies besprechen und anschließen werden wir unsere Modellreduzierung charakterisieren. Nachdem wir einen Weg finden, um Ergebnisse aus dem reduzieren Netzwerk zum originalen Netzwerk abzuleiten demonstrieren wir unseren Ansatz an biologischen Systemen aus den wirklichen Leben.

Contents

Abstract

Robustness of biochemical systems, i.e. their capacity to maintain normal functionality under diverse environmental conditions, has become a focal point of study in systems biology and bioinformatics. In the context of biochemical reaction networks, robustness research of the kind that we aim to contribute to focuses on species concentrations, with a robust species being one that maintains the same concentration across all equilibrium points of the network. While this research has produced a number of interesting results, the scope of our knowledge on this subject remains intriguingly limited. In this thesis we expand on existing work in an attempt to generalize our understanding of mechanisms that confer robustness to species in a given network. After giving a brief introduction into the field of systems biology and the many applications of mathematics therein we proceed to characterize the central problem discussed in this thesis, 'Can we identify biochemical network structures and sub-structures that constrain species concentrations in a way which allows us to infer robustness?'. In order to answer that question we go on to explore the general theory behind reaction networks and discuss some of the more important early results. Then we review the more recent theory on robustness in these networks and finally we outline our findings, which aim to open up the possibility of investigating robustness to a substantially larger number of networks than was possible before. We achieve this by first extending mathematical findings from past work to a more general class of networks. This in itself may still be restricting as, with some other considerations we will discuss, lack of scalability might make it impossible to apply these results in large-scale networks. However, using crucial properties of the aforementioned extension, we create an algorithm which can be used to analyze biological networks with hundreds of reactions. Using real world examples that range from small and well studied to large and previously unexplored networks we demonstrate how this new insight allows us to find robustness across a wider spectrum and in a less computationally expensive manner than any other method available.

1 Introduction

Mathematical models have found application in the field of biology since the late 19th century [18, 17, 14]. Population growth models were used to understand the effects of natural selection [24, 36], reaction-diffusion models helped to explain how physical processes and constraints affect the shape of organisms [33, 34] and stochastic processes were applied to a variety of problems in genetics and other areas [12, 4]. And as biological research continued to progress, from abstract conceptualizations to in-depth explanations, from localized to system-wide analysis, finding and studying biomathematical models would become more and more important to understand the mechanisms behind life [41]. In molecular biology, this exact paradigm shift, away from localized qualitative models and towards quantitative models of system dynamics, and the associated move towards more mathematical and computational approaches saw the emergence of a new field, that of 'systems biology' [37, 2].

As the name suggests, systems biology, operating on the molecular level, examines biological structures as a whole rather than focusing on their individual parts [21, 25].

Biological knowledge is used in combination with technological and analytical innovations in order to gather, process and interpret huge swaths of data. This approach already gave rise to a considerable number of new insights into the structure, dynamics and properties of biological systems [35, 38] and will only become more important as computational and experimental technologies advance. In this thesis we hope to add to this pool of knowledge by helping to get one step closer to answering one of biology's most interesting questions: 'How does life adapt and flourish in the face of at times drastic environmental changes?'

2 Problem Description

Now that we have a rough understanding of the general context we will be working in our next step will be to lay the ground work for a mathematical discussion of biochemical systems and then take a look at the specific problem at the centre of this thesis. At its most basic, a biochemical system describes the interactions between a set of molecules $\mathcal{S}$ in some environment. Mathematically each molecule $s \in \mathcal{S}$, commonly referred to as biochemical species or simply species, has an associated concentration $c_s \in \mathbb{R}$ which through interaction with the other species changes over time. Putting all the distinct concentrations into a single concentration vector $c \in \mathbb{R}^{\mathcal{S}}$, the dynamics of the system are given by some function f such that

$$\frac{d}{dt}c(t) = f(c)$$

which, depending on the biochemical system, can vary in shape to some degree [15, 6, 31]. Mathematical analysis of biochemical systems therefore is the investigation of a set of ordinary differential equations given by f and results include identification of steady states and behavior at and around those steady states, such as stability and, most important to us, robustness.

As a biological term, robustness describes a system's ability to maintain normal functionality in the face of environmental changes [23]. It is a critical component to the adaptability of life and unfolds itself in many different ways [40, 22]. Robustness differs from stability in that in a robust system a small perturbation at steady state may lead to a transition towards a different steady state. What robustness guarantees is that functionality in maintained. Mathematically we can capture this by choosing a function $g : \mathbb{R}^{\mathcal{S}} \to X$ and defining a system to be robust with respect to g if for any practical steady state c^* and any steady state c^{**} that is reached upon perturbation of the system at c^*, we have $g(c^*) = g(c^{**})$. We specify c^* to be practical since many steady states of the underlying dynamical system, such as those involving negative concentrations, are not feasible or interesting in application and will thus not be of any concern to us. In general, this means we will only focus on those steady states where every species is contributing, i.e. every species concentration is positive

Definition 1. For a biochemical system with dynamics

$$\frac{d}{dt}c(t) = f(c)$$

a concentration vector $c^* \in \mathbb{R}^{\mathcal{S}}$ is a **positive steady state** if $f(c^*) = 0$ and for every species $s \in \mathcal{S}$ we have $c_s^* > 0$.

For the specific type of robustness we are interested in, which is known as absolute concentration robustness, we will use $g(c) = c_s$ for some species s.

Definition 2. A system exhibits **absolute concentration robustness (ACR)** in species s if for any two positive steady states $c^*, c^{**} \in \mathbb{R}^{\mathcal{S}}_+$ we have $c_s^* = c_s^{**}$.

Having this property in a system means that even upon environmental perturbation, any function resting solely upon the concentration of species s will be maintained. Our objective in this thesis will be to find a way to identify ACR in a subset of biochemical systems.

3 Preliminaries

In this section we will define the subset of biochemical systems considered in this thesis and then discuss the relevant general theory pertaining these systems.

3.1 Biochemical Reaction Networks

As we mentioned above, a biochemical system describes the interactions between a set of species $\mathcal{S}$ as a system of ordinary differential equations

$$\frac{d}{dt}c(t) = f(c)$$

For any non-trivial system, the function f, called the species formation rate function, is non-linear in c which means a direct analysis of the system's dynamics is not possible. However, the structure of the interactions between species, which can be decomposed into elementary reactions such as

$$A + 2B \to C$$

creates linearities which we can use in order to analyze the system. Each of these reactions $y \to y'$ is characterized by its input y, output y', and speed $\kappa(c; y, y')$. The input, referred to as substrate, and the output or product are each a complex of species, i.e. for each species s there is an associated (non-negative) stoichiometry $y(s)$ and $y'(s)$ in the substrate and product complexes. In the example of a reaction above we have

$$\begin{array}{lll} y(A) = 1 & y(B) = 2 & y(C) = 0 \\ y'(A) = 0 & y'(B) = 0 & y'(C) = 1 \end{array}$$

Collectively the elementary reactions in a biochemical system form a reaction network giving us the following formalization.

Definition 3. A **biochemical reaction network** is a quadruple $\{\mathcal{S}, \mathcal{C}, \mathcal{R}, \kappa\}$ consisting of a finite set of species $\mathcal{S}$, a finite set of complexes $\mathcal{C} \subset \mathbb{R}^{\mathcal{S}}$, a finite set of reactions $\mathcal{R} \subset \mathcal{C} \times \mathcal{C}$ denoted $y \to y'$ for $y, y' \in \mathcal{C}$ and a kinetic $\kappa : \mathbb{R}^{\mathcal{S}} \times \mathcal{R} \to \mathbb{R}$ determining the speed of each reaction for a given concentration.

The species formation rate function of a biochemical system, given its reaction network, is composed as follows. For each species s, a reaction $y \to y'$ creates y'_s molecules of s and uses up y_s. The change in s which a single reaction $y \to y'$ creates therefore is $y'_s - y_s$. Since the speed of each reaction $y \to y'$ for a concentration c is given by $\kappa(c; y \to y')$, the aggregate rate of change in species s will be

$$f_s(c) = \sum_{y \to y' \in \mathcal{R}} \kappa(c; y \to y')(y'_s - y_s)$$

and thus

$$f(c) = \sum_{y \to y' \in \mathcal{R}} \kappa(c; y \to y')(y' - y)$$

There are a number of different kinetics [7, 28, 3, 5] that are used when modeling biochemical systems, among them mass action kinetics, the one we will use throughout our thesis. Mass action kinetics, which has been shown to successfully model chemical and biochemical processes in the past [11], is defined as

$$\kappa(c; y \to y') = k_{y \to y'} \prod_{s \in \mathcal{S}} c_s^{y_s}$$

meaning that the speed of each reaction $y \to y'$ is determined by the concentrations of the species in the substrate complex and a positive rate constant $k_{y \to y'} \in \mathbb{R}_+$. A common convention here is to use

$$c^y := \prod_{s \in \mathcal{S}} c_s^{y_s}$$

so that the kinetics can be written as

$$\kappa(c; y \to y') = k_{y \to y'} c^y$$

and we get

$$f(c) = \sum_{y \to y' \in \mathcal{R}} k_{y \to y'} c^y$$

When not explicitly stated otherwise any network we discuss from here on out will be considered to be a mass action network, i.e. have mass action kinetics, and our results will be pertaining to the set of biochemical systems that are modeled using a mass action networks. Let us now take a look at three network examples we will use throughout this thesis.

Example 1. The first example will be $N = \{\{A\}, \{A, 0\}, \{A \to 0, 0 \to A\}$. This Network has just one species, A, and two reactions, one which adds A to the system and one which removes it. The structure of each network can be visualized as a directed graph with the complexes as its vertices and the reactions as its edges. The structural graph for this network would therefore be

$$A \rightleftharpoons 0$$

The rate constants $k_{y\to y'}$ can be interpreted as weights for the edges of this graph. For this network the species formation rate function is given as

$$f(c_A) = k_{0\to A}c^0(1-0) + k_{A\to 0}c^A(0-1)$$

using $k_1 := k_{0\to A}$ and $k_2 := k_{A\to 0}$ as a shorthand we therefore get

$$f(c_A) = k_1 - k_2 c_A$$

For any steady state c_A^* of this network we will have to have $f(c_A^*) = 0$. This gives us

$$c_A^* = \frac{k_1}{k_2}$$

as the only steady state of the system. Since the network only has a single steady state, any robustness property, and in particular ACR, will trivially hold. To have a meaningful definition of ACR we will therefore require at least two steady states.

Example 2. Our next simple example, taken from [29], is given by

$$\begin{aligned}\mathcal{S} &= \{A, B\}\\ \mathcal{C} &= \{A+B, 2B, A, B\}\\ \mathcal{R} &= \{A+B \to 2B, B \to A\}\end{aligned}$$

We have two species involved in two reactions via four complexes and the structural graph is

$$\begin{aligned}A + B &\to 2B\\ B &\to A\end{aligned}$$

Furthermore, given a concentration c, the speed of the two reactions are

$$\begin{aligned}\kappa(c; A+B\to 2B) &= k_1 c_A^1 c_B^1 c_C^0\\ &= k_1 c_A c_B\end{aligned}$$

where k_1 is used as a shorthand for $k_{A+B\to 2B}$ and

$$\begin{aligned}\kappa(c; B\to A) &= k_2 c_A^0 c_B^1 c_C^0\\ &= k_2 c_B\end{aligned}$$

We can see that the first reaction consumes one molecule of A and the second reaction produces one molecule of A. Therefore the rate of change in the concentration of species A will be

$$\begin{aligned}f_A(c) &= \dot{c}_A\\ &= -\kappa(c; A+B\to 2B) + \kappa(c; B\to A)\\ &= -k_1 c_A c_B + k_2 c_B\end{aligned}$$

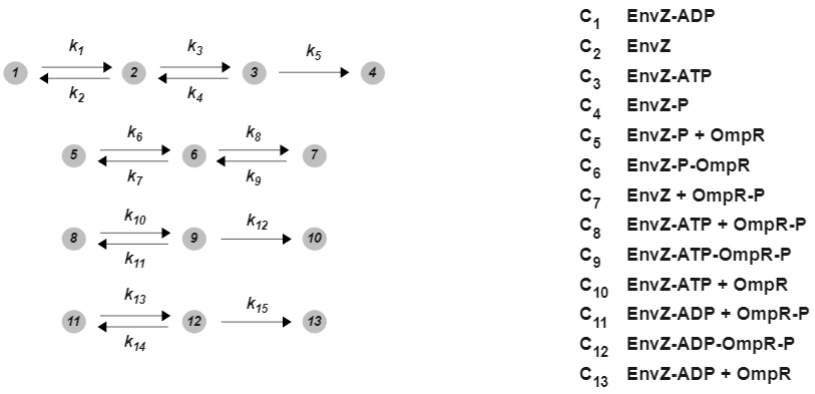

Figure 1: The EnvZ/OmpR two-component osmoregulator in E. coli, taken from [19]

i.e. the speed of the second reaction minus the speed of the first. Similarly for species B we get

$$\begin{aligned} f_B(c) &= \dot{c}_B \\ &= +\kappa(c; A + B \to 2B) - \kappa(c; B \to A) \\ &= k_1 c_A c_B - k_2 c_B \end{aligned}$$

and so

$$f(c) = \begin{pmatrix} -k_1 c_A c_B + k_2 c_B \\ k_1 c_A c_B - k_2 c_B \end{pmatrix}$$

which means that the steady states are either at an arbitrary $c_B^* \neq 0$ and $c_A^* = \frac{k_1}{k_2}$ or at $c_B^* = 0$ and an arbitrary c_A^*. As mentioned above we will only want to consider those steady states interesting in application , i.e. positive steady states. These are given by

$$c^* = \begin{pmatrix} k_1/k_2 \\ \theta \end{pmatrix}$$

with $\theta > 0$ being some positive constant and we can see that the system will be robust with respect to $g(c) = c_A$, meaning that the network exhibits ACR in species A.

Example 3. The last example, taken from [19] and given by Figure 1, is a network with 9 species, 13 complexes and 15 reactions describing the EnvZ/OmpR two-component osmoregulator in E. coli. This is still a rather small network and we can already see that finding all steady states in a non-linear dynamic system with nine equations will be extremely elaborate if it is even possible. If we want to find ACR in this network, we are going to need another method of searching for it, one that relies on the linearities arising from the graph structure of the network. The rest of this section is dedicated to setting the groundwork for developing such a method.

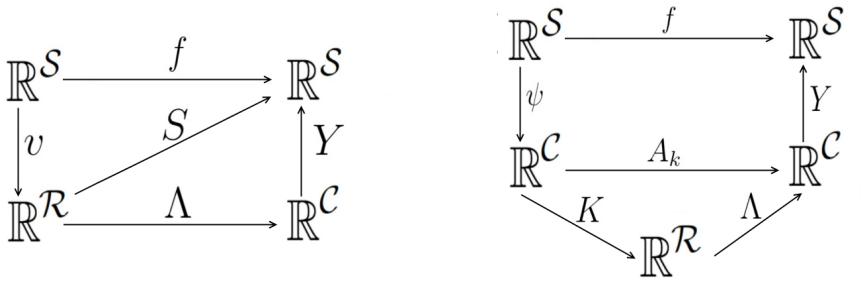

Figure 2: Two decompositions of the species formation rate function, the left for general kinetics and the right for mass action kinetics.

3.2 Deficiency

Let us start by going back to the general species formation rate function

$$f(c) = \sum_{y \to y' \in \mathcal{R}} \kappa(c; y \to y')(y' - y)$$

We can see that this function is linear in $\kappa(c; y \to y')$ and so can be rewritten as

$$f(c) = Sv(c)$$

where the columns of the stoichiometric matrix $S \in \mathbb{R}^{\mathcal{S}\times\mathcal{R}}$ are $y' - y$ for each reaction $y \to y'$ and

$$v_{y\to y'}(c) = \kappa(c; y \to y')$$

Furthermore if we define $\omega^y \in \mathbb{R}^{\mathcal{C}}$ as the unit vector corresponding to y, i.e.

$$\omega^y_{y'} = \begin{cases} 1 & \text{if } y = y' \\ 0 & \text{otherwise} \end{cases}$$

and $Y \in \mathbb{R}^{\mathcal{S}\times\mathcal{C}}$ such that its columns are the complexes y then

$$Y\omega^y = y$$

With that we can factorize S as

$$S = Y\Lambda$$

where the columns of $\Lambda \in \mathbb{R}^{\mathcal{C}\times\mathcal{R}}$ are $\omega^{y'} - \omega^y$ for each reaction $y \to y'$, meaning that Λ is the incidence matrix of the graph. With this we have another way of writing the species formation rate function as

$$f(c) = Y\Lambda v(c)$$

where Y describes the composition of the complexes, Λ the structure of the reaction graph and $v(c)$ the speed of the reactions. With that we can define the first important property for analyzing reaction networks.

Definition 4. The **structural deficiency** $\delta_s \in \mathbb{N}$ of a reaction network is given by

$$\delta_s = \dim\big(\ker(Y) \cap \mathrm{Im}(\Lambda)\big)$$

The reason we are interested in the structural deficiency of a network is that we can use it as a measure of the influence the structure of the graph has on set of steady states. For example, if we have a network with $\delta_s = 0$ then every v such that $Y\Lambda v = 0$ will also satisfy $\Lambda v = 0$ and since for every steady state c^* of the network we have

$$0 = f(c^*) = Y\Lambda v(c^*)$$

we must have

$$\Lambda v(c^*) = 0$$

and therefore the steady states are determined exclusively by the graph and the kinetics and not by the composition of the complexes. Knowing that the structure of a graph and therefore Λ always follow certain rules, this makes it possible to infer a number of implications on the set of steady states from existence and uniqueness to stability and robustness. We will outline some of those later on.

Let us now proceed to the next important property. Suppose that the kinetics of a network are mass action and recall that for such a network we have

$$\kappa(c; y \to y') = k_{y\to y'} c^y$$

With $\psi : \mathbb{R}^{\mathcal{S}} \to \mathbb{R}^{\mathcal{C}}$ defined as

$$\psi_y(c) = c^y$$

we can now rewrite $v(c)$ as $v(c) = K\psi(c)$ where $K \in \mathbb{R}^{\mathcal{R}\times\mathcal{C}}$ and

$$K_{y\to y',y''} = \begin{cases} k_{y\to y'} & \text{if } y = y'' \\ 0 & \text{otherwise} \end{cases}$$

for each $y \to y' \in \mathcal{R}$ and $y'' \in \mathcal{C}$. Defining $A_k = \Lambda K$ gives us

$$f(c) = YA_k\psi(c)$$

and as before we will be interested in $\ker(Y) \cap \mathrm{Im}(A_k)$

Definition 5. The **dynamic deficiency** $\delta_d \in \mathbb{N}$ of a reaction network is given by

$$\delta_d = \dim\big(\ker(Y) \cap \mathrm{Im}(A_k)\big)$$

Since $\mathrm{Im}(A_k) \subset \mathrm{Im}(\Lambda)$ any network will have $\delta_d \leq \delta_s$. The same reasoning that made us interested in structural deficiency applies here, since A_k describes the structure and kinetics of the network.

3.2.1 Structural Deficiency

Let us take a more in-depth look at the structural deficiency. First we want to find an easier way to calculate δ_s. Recall that the definition we used above was

$$\delta_s = \dim\big(\ker(Y) \cap \mathrm{Im}(\Lambda)\big)$$

which has no immediately apparent neat way to be solved. However, since Y and Λ are defined in a way so that $S = Y\Lambda$ and we can also write the above as

$$\begin{aligned}\delta_s &= \dim\big(\ker(Y) \cap \mathrm{Im}(\Lambda)\big) \\ &= \dim\big(\ker(Y\Lambda)\big) - \dim\big(\ker(\Lambda)\big) \\ &= \big(|\mathcal{R}| - rank(Y\Lambda)\big) - \big(|\mathcal{R}| - rank(\Lambda)\big) \\ &= rank(\Lambda) - rank(Y\Lambda)\end{aligned}$$

and thus another way to find the structural deficiency is to find $rank(\Lambda) - rank(S)$. Let us take a look at $rank(\Lambda)$. By definition the columns of Λ are $\omega^{y'} - \omega^{y}$ for each $y \to y' \in \mathcal{R}$ and therefore

$$\begin{aligned}rank(\Lambda) &= \dim\big(\mathrm{Im}(\Lambda)\big) \\ &= \dim\big(\ker(\Lambda^T)^{\perp}\big) \\ &= \dim(\mathbb{R}^{\mathcal{C}}) - \dim\big(\ker(\Lambda^T)\big) \\ &= |\mathcal{C}| - \dim\big(\{x \in \mathbb{R}^{\mathcal{C}} \,|\, \forall\, y \to y' \in \mathcal{R} \,:\, \langle x, \omega^{y'} - \omega^{y}\rangle = 0\}\big)\end{aligned}$$

Any x that satisfies

$$\forall\, y \to y' \in \mathcal{R} \,:\, \langle x, \omega^{y'} - \omega^{y}\rangle = 0$$

will have to have $x_y = x_{y'}$ for all y and y' that are directly linked $y \leftrightsquigarrow y'$, i.e. either $y \to y' \in \mathcal{R}$ or $y' \to y \in \mathcal{R}$. By extension if y is indirectly linked to y', meaning that there exist $y_1, ..., y_n \in \mathcal{C}$ such that

$$y \leftrightsquigarrow y_1\,,\; y_i \leftrightsquigarrow y_{i+1}\,,\; y_n \leftrightsquigarrow y' \in \mathcal{R}$$

then also $x_y = x_{y'}$. This gives us the following intuitive definition for linked complexes

Definition 6. We say that $y, y' \in \mathcal{C}$ are **linked** $y \sim y'$ if $y = y'$, $y \leftrightsquigarrow y'$ or there exist $y_1, ..., y_n \in \mathcal{C}$ such that

$$y \leftrightsquigarrow y_1\,,\; y_i \leftrightsquigarrow y_{i+1}\,,\; y_n \leftrightsquigarrow y' \in \mathcal{R}$$

It is easy to check that this is an equivalence relation and thus partitions the set of complexes into equivalence classes

$$\mathcal{C} = \bigcup_{i=1}^{l} L_i$$

The sets L_i, describing the connected components of the undirected graph, are called linkage classes and from above we know that if y and y' are in the same linkage class, then $x_y = x_{y'}$ so that if we define $\omega^{L_i} \in \mathbb{R}^{\mathcal{C}}$ such that

$$\omega_y^{L_i} = \begin{cases} 1 & \text{if } y \in L_i \\ 0 & \text{otherwise} \end{cases}$$

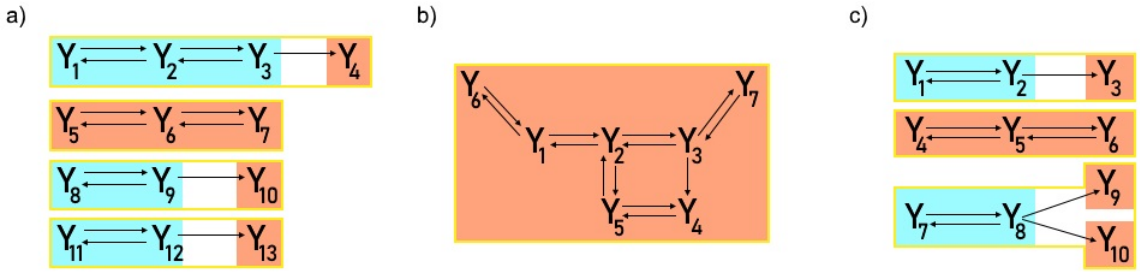

Figure 3: Three Examples for the structure of reaction networks. Linkage classes are marked with yellow, terminal strong linkage classes are highlighted with orange, non-terminal strong linkage classes with blue

then $x \in \text{span}\{\omega^{L_1}, ..., \omega^{L_l}\}$. Furthermore choosing any linkage class L_i by definition if $y \to y' \in \mathcal{R}$ then y and y' are linked and thus either both are in L_i or neither. In both cases we have

$$\langle \omega^{L_i}, \omega^{y'} - \omega^{y} \rangle = 0$$

which combined with the above gives us

$$\text{span}\{\omega^{L_1}, ..., \omega^{L_l}\} = \{x \in \mathbb{R}^{\mathcal{C}} \,|\, \forall\, y \to y' \in \mathcal{R} \,:\, \langle x, \omega^{y'} - \omega^{y} \rangle = 0\}$$

and so we finally get

$$rank(\Lambda) = |\mathcal{C}| - l$$

Using the common conventions $n := |\mathcal{C}|$ and $s := rank(S)$ this gives us the new way to calculate the structural deficiency we were looking for, namely

$$\delta_s = n - l - s$$

Next let us go back to where we characterized linkage classes. The way we defined a link between two complexes was to disregard the directions of the reactions and see if there is a connection in the graph between them. As it turns out, later on the directions of these connections will also be very much of interest to us and specifically we will be interested in those connections that all follow the same direction. In order to capture this we define the following new type of link

Definition 7. We say that $y, y' \in \mathcal{C}$ are **strongly linked** if $y = y'$ or there exists $y_1, ..., y_n, y'_1, ..., y'_m \in \mathcal{C}$ such that

$$y \to y_1\,,\; y_i \to y_{i+1}\,,\; y_n \to y' \in \mathcal{R}$$

and

$$y' \to y'_1\,,\; y'_i \to y'_{i+1}\,,\; y'_n \to y \in \mathcal{R}$$

In other words, two complexes are strongly linked if there exists a cycle in the graph that contains both of them. As before, it is easy to check that this is an equivalence relation and partitions the complexes into equivalence classes

$$\mathcal{C} = \bigcup_{i=1}^{p} P_i$$

The sets P_i characterize the strongly connected components of the graph and are called strong linkage classes. If in a strong linkage class P_i there are no outgoing reactions, meaning that, for any complex $y \in P_i$ in that class, $y \to y' \in \mathcal{R}$ implies that $y' \in P_i$, then P_i is called terminal, and if this is not the case it is non-terminal. Having this we can get to the final important structural property in this section.

Definition 8. A reaction network is **weakly reversible** if all strong linkage classes are terminal.

Equivalently we can say that a network is weakly reversible if the linkage classes coincide with the strong linkage classes, i.e. for every two linked complexes there is a cycle including both of them. All of the above concepts are visually quite intuitive and Figure 3 shows three examples that should give us an understanding sufficient enough to move on from structural to kinetic properties.

3.2.2 Dynamic Deficiency

Recall that within the mass action framework, the species formation rate function is given by

$$f(c) = \sum_{y \to y' \in \mathcal{R}} k_{y \to y'} \psi_y(c)(y' - y)$$

where $\psi_y(c) = c^y$. Since this function is linear in $\psi(c)$ we could rewrite it as

$$f(c) = Y A_k \psi_y(c)$$

and defined the dynamic deficiency as

$$\delta_d = \dim \big(\ker(Y) \cap \mathrm{Im}(A_k) \big)$$

Analogous to the last section, to make more sense of this construct, we will want to study the latter part of this equation, namely the matrix A_k. To start off, let us take a look at two different ways to characterize A_k. The first we already discussed when drawing the connection from structural to dynamic deficiency. Here we used the fact that coming from

$$f(c) = Y \Lambda v(c)$$

we have $v(c) = K\psi(c)$ and thus

$$f(c) = Y \Lambda K \psi(c)$$

and A_k can be characterized as the product of two matrices with relatively simple structure. The other way we can represent A_k is as a difference of two equally simple matrices.

Proposition 1. *For the adjacency matrix of graph weighted by the rate constants* $\Phi \in \mathbb{R}^{\mathcal{C} \times \mathcal{C}}$ *such that*

$$\Phi_{y,y'} = \begin{cases} k_{y \to y'} & \text{if } y \to y' \in \mathcal{R} \\ 0 & \text{otherwise} \end{cases}$$

and $\Delta(x)$ *such that*

$$\Delta(x)_{y,y'} = \begin{cases} (x)_y & \text{if } y = y' \\ 0 & \text{otherwise} \end{cases}$$

we have

$$A_k = \Phi^T - \Delta(\Phi 1)$$

The matrix $\Delta(\Phi 1)$ *is the degree matrix of the weighted graph, meaning that* A_k *is its negative Laplacian [39].*

The proof can be found in the Appendix. Let us quickly think about what this means for the structure of A_k. First off, for any entry (y, y') outside of the diagonals, by definition $\Delta(\Phi 1)_{y,y'}$ will be zero and so A_k will coincide with Φ^T. Since $\Phi^T_{y,y'} = \Phi_{y',y}$ is equal to the rate constant of the reaction from y' to y if such a reaction exists, and zero if it does not, $(A_k)_{y,y'} = \Phi_{y',y}$ can be seen as a measure of direct throughput from complex y' to y. In the diagonals we find that

$$\begin{aligned} (A_k)_{y,y} &= \Phi_{y,y} - \Delta(\Phi 1)_{y,y} \\ &= \Phi_{y,y} - (\Phi 1)_y \\ &= \Phi_{y,y} - \sum_{y' \in \mathcal{C}} \Phi_{y,y'} \\ &= \sum_{y' \neq y} \Phi_{y,y'} \end{aligned}$$

which by analogy is the combined throughput flowing out of y. If we think of $\psi(c)_y$ as a measure of pressure at node y this is rather intuitive since the rate of change in that node $(A_k \psi(c))_y$ will be determined by the flow coming from of incoming nodes $\Phi_{y',y}\psi_{y'}(c)$ and the combined outgoing flow

$$\Big(\sum_{y' \neq y} \Phi_{y,y'} \Big) \psi_y(c)$$

Having some intuition about the structure of A_k we can now move on to study its properties in some detail. Since many results in chemical reaction network theory depend on the specific structure of $\ker(A_k)$, let us start there. Going back to our previous analogy, for any $\psi \in \mathbb{R}^{\mathcal{C}}$ to be in $\ker(A_k)$, the incoming and outgoing flow in every node will have to be the same. If we apply pressure at any node $\psi_y > 0$ that has outgoing reactions, the nodes these reactions lead to will have incoming flow and will thus need to have pressure in order to create outgoing flow. In other terms

$$y \to y' \in \mathcal{R} \Rightarrow (\psi_y > 0 \Rightarrow \psi_{y'} > 0)$$

It follows that if in a strong linkage class P_i any node is 'pressured', then all must be, i.e.

$$y, y' \in P_i \Rightarrow (\psi_y > 0 \Rightarrow \psi_{y'} > 0)$$

Furthermore, we would expect the same principle of incoming and outgoing flow canceling out that we have for single nodes has to be true on a higher level, meaning that

if a strong linkage class is pressured, then all strong linkage classes 'upstream' would have incoming flow and thus would need to create outgoing flow. If we assume this to be true and take any terminal strong linkage class P_j, and another strong linkage class downstream being pressured, then P_j has incoming flow but no way to create outgoing flow since it is terminal. This means that strong linkage classes downstream from terminal strong linkage classes can not have pressure. Mathematically we get the following.

Proposition 2. *Let $T_1, ..., T_t$ be the terminal strong linkage classes of the network corresponding to A_k. There exists a basis $\xi_1, ..., \xi_t$ of $\ker(A_k)$ such that for every $i = 1, ..., t$ we have*

$$supp(\xi_i) \subset T_i$$

and $(\xi_i)_y > 0$ for $y \in T_i$

The proof [13] is mathematically somewhat more involved but follows the basis logic we just laid out here. As mentioned above, this structural quality of the kernel of A_k is the basis for some interesting results [16, 10, 9] one of which we will discuss next.

3.3 The Deficiency Zero Theorem

The Deficiency Zero Theorem [10] is one of the earliest and most impactful results is chemical reaction network theory. As the name suggests the theorem will deal with networks that have a deficiency of zero. We might ask which type of deficiency, however, as we will see, under the conditions of the theorem, dynamic and structural deficiency will coincide. For now let us assume we are dealing with networks of dynamic deficiency zero $\delta_d = 0$. Let $c^* \in \mathbb{R}^{\mathcal{S}}$ be a steady state in such a network, i.e. $f(c^*) = 0$. Then we know that

$$f(c^*) = YA_k\psi(c^*) = 0$$

which means

$$\psi(c^*) \in \ker(YA_k)$$

Now it is possible that either $\psi(c^*)$ is in the kernel of A_k or that is in not and $A_k\psi(c^*)$ is in the kernel of Y. However, since

$$\delta_d = \dim(\mathrm{Im}(A_k) \cap \ker(Y)) = 0$$

the second option is impossible and we must have

$$\psi(c^*) \in \ker(A_k)$$

This means that for any basis $\{\xi_1, ..., \xi_t\}$ of $\ker(A_k)$ we can find coefficients λ_i such that

$$\psi(c^*) = \sum_{i=1}^{t} \lambda_i \xi_i$$

From Proposition 2 we know that $\{\xi_1, ..., \xi_t\}$ can be chosen in such a way that ξ_i takes positive values in the terminal strong linkage class T_i and is zero everywhere else, which means that for $\psi(c^*)$ we have

$$\begin{aligned} supp(\psi(c^*)) &\subset \bigcup_{i=1}^{t} supp(\xi_i) \\ &= \bigcup_{i=1}^{t} T_i \end{aligned}$$

so that $\psi(c^*)$ has to be zero for any complex y that is in a non-terminal strong linkage class. For a positive steady state c^*, we have

$$\psi_y(c^*) = \prod_{s \in \mathcal{S}} (c_s^*)^{y_s} > 0$$

for any complex y which means that for a positive steady state to exist in a network with $\delta_d = 0$, we must have

$$\mathcal{C} = supp(\psi(c^*)) \subset \bigcup_{i=1}^{t} T_i$$

implying that for every $y \in \mathcal{C}$ there has to exist a terminal strong linkage class T_i such that $y \in T_i$. In other words, there can be no non-terminal strong linkage classes, which was the definition of a weakly reversible network (Section 3.2.1 Definition 8). Since in a weakly reversible network the linkage classes coincide with strong linkage classes and all strong linkage classes are terminal, the number of linkage classes l has to be equal to the number of terminal strong linkage classes t. That means for A_k we have

$$\begin{aligned} \dim(\mathrm{Im}(A_k)) &= |\mathcal{C}| - \dim(\ker(A_k)) \\ &= n - t \\ &= n - l \end{aligned}$$

which is the same as $\dim(\mathrm{Im}(\Lambda))$ as we know from Section 3.2.1. This, together with $\mathrm{Im}(A_k) \subset \mathrm{Im}(\Lambda)$ then implies that

$$\mathrm{Im}(A_k) = \mathrm{Im}(\Lambda)$$

and therefore

$$\delta_d = \dim(\ker(Y) \cap \mathrm{Im}(A_k)) = \dim(\ker(Y) \cap \mathrm{Im}(\Lambda)) = \delta_s$$

which means in a weakly reversible network structural and dynamic deficiency will always be the same. What we just found is only one part of the full theorem which reads as follows

Theorem 1. *Let $N = \{\mathcal{S}, \mathcal{C}, \mathcal{R}, \kappa\}$ be a reaction network endowed with mass action kinetics of structural or dynamic deficiency zero. There exists a positive steady state $c^* \in \mathbb{R}^{\mathcal{C}}$ in N if and only if the network is weakly reversible. Any other positive concentration vector $c \in \mathbb{R}^{\mathcal{C}}$ is a steady state if and only if $\log(c^*) - \log(c) \in \ker(S^T)$. Every steady state is locally stable against perturbations in $\mathrm{Im}(S)$.*

Since our results will only use the arguments we presented above we will not discuss the proof any further. A full proof can be found in [10]. robustness can be found in smaller networks.

4 Existing Approaches to Determine ACR

There are two main results that give us a way to search for robust species in biochemical reaction networks. The first [29] gives us a structural condition to identify complexes $y, y' \in \mathcal{C}$ that are coupled, meaning that there exists a $\theta \in \mathbb{R}$ such that

$$\frac{(c^*)^y}{(c^*)^{y'}} = \theta$$

for any positive steady state c^*. Robustness follows if y and y' only differ in a single species, i.e.

$$\frac{(c^*)^y}{(c^*)^{y'}} = (c_s^*)^\alpha$$

for some $\alpha \in \mathbb{R}$. The second approach [19] uses symbolic computations in order to find complexes $y^1, ..., y^k \in \mathcal{C}$ for which there exist coefficients $a \in \mathbb{R}^k$ such that

$$\sum_{i=1}^{k} a_i (c^*)^{y^i} = 0$$

for all steady states c^*. Given this 'invariance', certain structural relations between the complexes will again make it possible to infer robustness.

4.1 Coupled Complexes

In a reaction network N, let us take any positive steady state c^*, then by definition $f(c^*) = 0$ and

$$Y A_k \psi(c^*) = 0$$

Assume now that the structural deficiency of the network is zero. Then, from our discussion of Theorem 1, we already know that the network is weakly reversible, that $\delta_s = \delta_d$, and that $\psi(c^*) \in \ker(A_k)$ since

$$\delta_d = \dim(\ker(y) \cap \mathrm{Im}(A_k)) = 0$$

Further we know that we can choose a basis $\{\xi_1, ..., \xi_l\}$ of $\ker(A_k)$ such that $supp(\xi_i) = L_i$ and $(\xi_i)_y \geq 0$ for all $i = 1, ..., l$ where L_l are the linkage classes of the network. This gives us

$$\psi(c^*) = \sum_{i=1}^{l} \lambda_i \xi_i$$

and in particular for $y \in L_j$

$$\psi_y(c^*) = \sum_{i=1}^{l} \lambda_i (\xi_i)_y$$

$$= \lambda_j (\xi_j)_y$$

since $(\xi_i)_y = 0$ for all $i \neq j$. That means that if we choose any two complexes $y, y' \in L_j$ in the same linkage class, then we get

$$\begin{aligned} \frac{\psi_y(c^*)}{\psi_{y'}(c^*)} &= \frac{\lambda_j (\xi_j)_y}{\lambda_j (\xi_j)_{y'}} \\ &= \frac{(\xi_j)_y}{(\xi_j)_{y'}} \end{aligned}$$

and since the basis of $\ker(A_k)$ can be chosen independently of the steady state the following theorem follows immediately.

Theorem 2. *In a reaction network of structural or dynamic deficiency zero, for any two complexes $y, y' \in \mathcal{C}$ in the same linkage class, there exists some $\theta \in \mathbb{R}$ such that*

$$\frac{(c^*)^y}{(c^*)^{y'}} = \theta$$

holds for every positive steady state c^*.

What this means for us is that in a deficiency zero network, if there are two complexes $y, y' \in \mathcal{C}$ in the same linkage class that differ only in a single species, then the network is robust with respect to that species, since we have

$$(c_s^*)^\alpha = \frac{(c^*)^y}{(c^*)^{y'}} = \theta$$

and therefore $c_s^* = \sqrt[\alpha]{\theta}$ for any positive steady state c^*. An interesting corollary to this theorem is that such a network can not conserve mass [30], meaning that there is no set of weights $w \in \mathbb{R}^{\mathcal{S}}$ such that $w_s > 0$ for all $s \in \mathcal{S}$ and

$$w^T S = 0$$

The reason here is if we have two complexes y, y' in the same linkage class, then there exists a set of reactions $y_1 \to y_1', ..., y_k \to y_k'$ linking these complexes, i.e.

$$y' - y = \sum_{i=1}^{k} \varepsilon_i (y_i' - y_i)$$

for some $\varepsilon_i \in \{-1, 1\}$. Since for everyone of those reactions $y_i' - y_i$ is a columns in S if such a w existed, then we have

$$\langle w, y_i' - y_i \rangle = 0$$

and so

$$\langle w, y' - y \rangle = 0$$

If $y' - y = \alpha s$ for some $\alpha > 0$ then $w_s \alpha = 0$ and therefore $w_s = 0$. A similar argument to the one we just used for identifying coupled complexes in deficiency zero networks can be used for networks that have a deficiency one one.

Theorem 3. *In a reaction network of structural or dynamic deficiency one, for any two complexes $y, y' \in \mathcal{C}$ that are both in non-terminal strong linkage classes, there exists some $\theta \in \mathbb{R}$ such that*

$$\frac{(c^*)^y}{(c^*)^{y'}} = \theta$$

holds for every positive steady state c^*.

Instead of discussing the proof, which is relegated to the Appendix, let us take a look at the first example of a reaction network we discussed.

$$\begin{aligned} A + B &\to 2B \\ B &\to A \end{aligned}$$

The species formation rate function of this network was given by

$$f(c) = \begin{pmatrix} -k_1 c_A c_B + k_2 c_B \\ k_1 c_A c_B - k_2 c_B \end{pmatrix}$$

We already remarked that for any positive steady state c^* we would have

$$0 = -k_1 c_A^* c_B^* + k_2 c_B^*$$

or

$$k_1 c_A^* c_B^* = k_2 c_B^*$$

which since $c_B^* > 0$ simplifies to $c_A^* = \frac{k_2}{k_1}$, i.e. species A exhibits absolute concentration robustness. Let us see if we can get to this result without looking at the dynamic equations of the network. First we will need to determine one of the two types of deficiency. Since the dynamic deficiency is in general hard to calculate let us use the structural deficiency for which we have the simple formula

$$\delta_s = n - l - s$$

where n was the number of complexes, l the number of linkage classes, and s the rank of S. That the number of complexes is four and the number of linkage classes is two is apparent by just looking at the reactions. Further we can also easily check that

$$S = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$$

so that $s = 1$ and we get $\delta_s = 4 - 2 - 1 = 1$. From Theorem 3 we therefore know that $A + B$ and B must be coupled, since they are both in non-terminal strong linkage classes. This, combined with the fact that $A + B$ and B differ only in species A gives us exactly what we were expecting, i.e. that there is ACR in A. So now we have our first way to check if a network has robust species. However, since this will only ever work if the network we are considering has a deficiency lower than two, which in larger networks is rarely the case, another approach is needed.

4.2 Invariants

With that let us get to the next result, which will not depend on deficiency at all. As mentioned above, we want to find complexes $y^1, ..., y^k \in \mathcal{C}$ for which there exist coefficients $a \in \mathbb{R}^k$ such that

$$\sum_{i=1}^{k} a_i (c^*)^{y^i} = 0$$

for all steady states c^*. In the following we will, given a set of complexes, call such a set of coefficients a an invariant. Since $\psi(c)$ is defined such that $\psi_y(c) = c^y$, by definition any invariant on the complexes $y^1, .., y^k$ will have to satisfy

$$\sum_{i=1}^{k} a_i \psi_{y^i}(c^*) = 0$$

With c^* being a steady state we furthermore have

$$YA_k\psi(c^*) = 0$$

which gives us a relatively easy way to look for a suitable invariants. Namely, by finding if there exists $a' \in \mathbb{R}^{\mathcal{C}}$ such that

$$a' \in \ker(YA_k)^{\perp}$$

and

$$supp(a') \subset \{y^1, ..., y^k\}$$

because if that holds true then we will have

$$\begin{aligned} 0 &= \sum_{y \in \mathcal{C}} a'_y \psi_y(c^*) \\ &= \sum_{i=1}^{k} a'_{y^i} \psi_{y^i}(c^*) \end{aligned}$$

for any steady state c^*, so that we can choose a to be the entries of a' on $y^1, ..., y^k$

$$a = a'|_{y^1,...,y^k}$$

and a will be an invariant. In that case we will say that the invariant a is of type 1. Since $\operatorname{span}\{\psi(c^*) \mid f(c^*) = 0\}$ does not have to be equal to $\ker(YA_k)$, it is possible for there to be those invariants that are not of type 1. However their nature makes it exceedingly difficult to find them in most networks [19] and so we will in this section concentrate exclusively on type 1 invariants. Further, without any loss in generality, moving on we can assume that the complexes are ordered in such a way, that $\mathcal{C} = \{y^1, ..., y^n\}$ meaning that when using $y^1, ..., y^k$ we are talking about the first k complexes in $\mathcal{C}$. This makes it more natural for $a \in \mathbb{R}^k$ to define $(a, 0) \in \mathbb{R}^{\mathcal{C}}$ such that with $a' = (a, 0)$ we have

$$a'_{y^i} = \begin{cases} a_i & \text{if } i \leq k \\ 0 & \text{otherwise} \end{cases}$$

With that let us assume that a is a type 1 invariant for $y^1, ..., y^k$. From above we have

$$(a, 0) \in \ker(YA_k)^\perp$$

which is the same as

$$(a, 0) \in \mathrm{Im}((YA_k)^T)$$

meaning that for some $b \in \mathbb{R}^{\mathcal{S}}$ we must have $(a, 0) = (YA_k)^T b$. Conversely if there exists $b \in \mathbb{R}^{\mathcal{S}}$ so that

$$((YA_k)^T b)_{y^i} = 0$$

for $i > k$ then $(YA_k)^T b = (a, 0)$ for some $a \in \mathbb{R}^k$ and b thus defines an invariant. It follows that a is a type 1 invariant if and only if $(a, 0) = (YA_k)^T b$. If we split

$$YA_k = K|N$$

where K are the first k columns and N the latter $n - k$ then such a b will have to satisfy $N^T b = 0$ and we get that the space of type 1 invariants I_k for the complexes $y^1, ..., y^k$ is given by

$$I_k = (YA_k)^T \ker(N^T)$$

and is thus isomorphic to $\ker(N^T)/\ker((YA_k)^T)$. One way to check for non-trivial type 1 invariants would therefore be to calculate $\mathrm{rank}(YA_k)$ and then for each subset $Z \subset \{1, ..., n\}$ remove those columns from YA_k. If the resulting matrix has a lower rank than YA_k then the space of type 1 invariants on the complexes with indices in Z is non-trivial, since

$$\begin{aligned} \dim(I_Z) &= \dim(\ker(N_Z^T)) - \dim(\ker((YA_k)^T)) \\ &= \mathrm{rank}(YA_k) - \mathrm{rank}(N_Z) \end{aligned}$$

One thing to note is that this method requires us to calculate $\ker((YA_k)^T)$, however we would much rather work with $\ker(YA_k)$ since we already know the shape of all but δ_d of its elements and have an efficient way to calculate them by partitioning A_k. Let us therefore quickly discuss a second method of finding invariants which uses $\ker(YA_k)$ instead. By definition, if $\{\beta_1, ..., \beta_p\}$ is a basis of $\ker(YA_k)$, then

$$\langle \beta_i, (a, 0) \rangle = 0$$

for any type 1 invariant a so that if we define B as the matrix that has β_i as its rows, then $(a, 0) \in \ker(B)$. Further, if we split

$$B = B'|X$$

such that B' contains the first k columns of B then a is an invariant if and only if $a \in \ker(B')$ and we get

$$\dim(I_k) = k - \mathrm{rank}(B')$$

Let $q = \mathrm{rank}(B')$ then, using elementary row operations [1] on B' we can bring it into a form where all but the upper q rows are zero without changing its image. Using the same elementary row operations on B gives us a change of basis of $\ker(YA_k)$ such that

in all but q of its elements the first k entries are zero. Without loss of generality we assume $\{\beta_1, .., \beta_p\}$ is exactly that basis and we have

$$B = \begin{pmatrix} B' & X_1 \\ 0 & X_2 \end{pmatrix}$$

Now, by definition $q = \mathrm{rank}(B')$, so that we can find q linearly independent columns of B'. Let $U \subset \{1, ..., k\}$ be the subset of corresponding indices and V the remaining $k - q$ indices and define B'_U as the $q \times q$ sub-matrix of B' containing the columns with indices in U and B'_V the sub-matrix containing the remaining columns, then a is a type 1 invariant if and only if

$$B'_U a_U + B'_V a_V = 0$$

Since we chose U in such a way that B'_U is invertible, this means

$$a_U = -(B'_U)^{-1} B'_V a_V$$

By choosing any $a_V \in \mathbb{R}^{k-q}$ we can therefore generate an invariant using the above formula. Finally we can use that by definition of an invariant

$$\langle \psi_U(c^*), a_U \rangle = -\langle \psi_V(c^*), a_V \rangle$$

holds for any steady state c^* so that with the above we have

$$\begin{aligned} \langle \psi_V(c^*), a_V \rangle &= -\langle \psi_U(c^*), a_U \rangle \\ &= -\langle \psi_U(c^*), -(B'_U)^{-1} B'_V a_V \rangle \\ &= \langle (B'_V)^T ((B'_U)^{-1})^T \psi_U(c^*), a_V \rangle \end{aligned}$$

which, given that a_V can be chosen arbitrarily, means

$$\psi_V(c^*) = (B'_V)^T ((B'_U)^{-1})^T \psi_U(c^*)$$

or

$$0 = (B'_V)^T ((B'_U)^{-1})^T \psi_U(c^*) - \psi_V(c^*)$$

Each of the $k - q$ rows of this matrix equation therefore describes an invariant with coefficients given by $(B'_V)^T ((B'_U)^{-1})^T$ involving one complex in the index set V and up to all of the complexes in index set U. To illustrate let us assume we have a network with $\delta_d = 0$. In that case we have

$$\ker(Y A_k) = \ker(A_k)$$

which using Proposition 2 means that we can choose B such that $B_{i,y} = 0$ if $y \notin L_i$. Assuming $y_1, y_2 \in \mathcal{C}$ are in the same linkage class we therefore have

$$B' = (B_{1,1}\, B_{1,2})$$

which choosing $U = \{1\}$ and $V = \{2\}$ gives us $B'_U = B_{1,1}$ and $B'_U = B_{1,2}$ and thus

$$\psi_{y_2}(c^*) = \frac{B_{1,2}}{B_{1,1}} \psi_{y_1}(c^*)$$

or

$$\frac{\psi_{y_2}(c^*)}{\psi_{y_1}(c^*)} = \frac{B_{1,2}}{B_{1,1}}$$

showing that in a deficiency zero network, any two y_1 and y_2 in the same linkage class are coupled.

5 Robustness in Large-Scale Networks

Both of the approaches we presented above are practical in small-scale networks but completely unusable when trying to analyze larger-scale networks. The reasons here are that the result on coupling requires a deficiency of zero or one to be applicable which is simply not the case in interesting large-scale networks. The approach of finding invariants might not depend on a low deficiency to be useful, however what it does require is biological knowledge helping to guide our search in the right direction since calculating all possible invariants will be too computationally expensive. In this section we will therefore attempt to further the research in this field by working our way up to an algorithm designed to search large reaction networks for robust species. We will use ideas from past research in combination with new ones in order to work past the difficulties large networks present, the first step of which will be breaking them down into manageable chunks.

5.1 Network Reduction

This reduction of the network, provided that it keeps important properties intact, will make further analysis much more feasible. The arguments we will discuss here will take roughly the following form. Let N be a reaction network with species formation rate function f and let N_{red} be another network with rate function f_{red}. Let further all steady states c^* of N be steady states of N_{red}, meaning that all zeroes of f are also zeros of f_{red}. Then any function g that is constant on the steady states of N_{red}, i.e.

$$g(c^*) = g(c^{**})$$

for all c^*, c^{**} such that $f_{red}(c^*) = f_{red}(c^{**}) = 0$ will also be constant on the steady states of N. With that in mind let us take a look at two ways in which we can reduce a network in order to make it more workable.

5.1.1 Species Removal

Assume we have the mass action network from Example 2

$$\begin{aligned} A + B &\to 2B \\ B &\to A \end{aligned}$$

with

$$f(c) = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} v(c)$$

Now let us consider one of the concentrations as fixed, for example $c_B = \theta$, and ask ourselves if we can find a network N_{-B} that has

$$f_{-B}(c_A^*) = 0$$

for every $c_A^* \in \mathbb{R}$ such that

$$f\begin{pmatrix} c_A^* \\ \theta \end{pmatrix} = 0$$

The most straightforward possibility for f_{-B} is just removing the row corresponding to B in S, so that we have

$$f_{-B}(c_A) = \begin{pmatrix} -1 & 1 \end{pmatrix} v \begin{pmatrix} c_A \\ \theta \end{pmatrix}$$

since then $f_{-B}(c_A) = f_A(c_A, \theta)$ which clearly satisfies the condition. What we now want to find out whether there exists a network N_{-B} that has

$$S_{-B} = \begin{pmatrix} -1 & 1 \end{pmatrix}$$

as its stoichiometric matrix and

$$v_{-B}(c_A) = v(c_A, \theta)$$

as its kinetics. Firstly we have

$$v_{-B}(c_A) = \begin{pmatrix} k_1 c_A \theta \\ k_2 \theta \end{pmatrix}$$

meaning that if we assume the reduced network still is mass action, then, since θ is a constant, the substrate complex in the first reaction will have to be $y_1 = A$ and in the second $y_2 = 0$. Furthermore S_{-B} tells us that in the first reaction consumes one molecule of A while the second produces one molecule. This gives us the quite natural reduced network N_{-B} as

$$A \leftrightharpoons 0$$

which is the network from Example 1 with

$$v_{-B}(c_A) = \begin{pmatrix} (k_{-B})_1 c_A \\ (k_{-B})_2 \end{pmatrix}$$

where $(k_{-B})_1 = k_1\theta$ and $(k_{-B})_2 = k_2\theta$. In effect this reduced network therefore just kept the reactions preserved and eliminated the reduced species by removing it from the complexes and adding the lost value to the kinetics. In general, species reduction will work quite similarly. To show this, suppose now we have a general mass action network N and want to remove species s. Firstly the set of species in the new network is

$$\mathcal{S}_{-s} = \mathcal{S} \setminus \{s\}$$

Now, kinetically, as we have seen, mass action makes it quite easy to fix a concentration since

$$\begin{aligned} v_{y \to y'}(c) &= k_{y \to y'} \prod_{s' \in \mathcal{S}} c_{s'}^{y_{s'}} \\ &= k_{y \to y'} c_s^{y_s} \prod_{s \neq s'} c_{s'}^{y_{s'}} \end{aligned}$$

and if $c_s = \theta$ we have a new mass action kinetics with

$$(k_{-s})_{y_{-s} \to y'_{-s}} = k_{y \to y'} \theta^{y_s}$$

Structurally this gives us that all reactions can remain unchanged, except that any complex y will now have to be y_{-s} meaning that we delete the entry in y corresponding to species s. This gives us that the set of complexes in the new network will be

$$\mathcal{C}_{-s} = \{y_{-s} \in \mathbb{R}^{(\mathcal{S}_{-s})} : y \in \mathcal{C}\}$$

and

$$\mathcal{R}_{-s} = \{y_{-s} \to y'_{-s} : y \to y' \in \mathcal{R}\}$$

This is underscored by the fact that we have $S = Y\Lambda$ meaning that $S_{-s} = Y_{-s}\Lambda$ where Y defines the complexes in N and Λ the reactions between these complexes. As we can see in the new network Λ can remain unchanged, meaning that the reaction structure is preserved and in the complexes, the row Y_s is removed, meaning that the new complexes will be just like the old ones, only with the entry corresponding to s removed. One thing we have not mentioned yet is that when removing s, two complexes which differ only in that species will fall together meaning they will be the same complex in the reduced network. On the one hand, this is good since that is where the reduction in size of the network is coming from, on the other hand, this can introduce one problem, namely duplicate reactions. Assume we have two reactions $y \to y', y'' \to y''' \in \mathcal{R}$ in N and assume further that $y_{-s} = y''_{-s}$ and $y'_{-s} = y'''_{-s}$. If this is the case, then they will correspond to the same reaction in N_{-s}. The relevant part of the species formation rate function is

$$f(c) = \sum_{y \to y' \in \mathcal{R}} k_{y \to y'} c^y (y' - y)$$

and therefore the important part for us is

$$k_{y \to y'} c^y (y' - y) + k_{y'' \to y'''} c^{y''} (y''' - y'')$$

which translates to

$$k_{y \to y'} \theta^{y_s} c_{-s}^{y_{-s}} (y'_{-s} - y_{-s}) + k_{y'' \to y'''} \theta^{y''_s} c_{-s}^{y''_{-s}} (y'''_{-s} - y''_{-s})$$

Since $y_{-s} = y''_{-s}$ and $y'_{-s} = y'''_{-s}$ will be

$$(k_{y \to y'} \theta^{y_s} + k_{y'' \to y'''} \theta^{y''_s}) c_{-s}^{y_{-s}} (y'_{-s} - y_{-s})$$

we will have to have

$$k_{y_{-s} \to y'_{-s}} = k_{y \to y'} \theta^{y_s} + k_{y'' \to y'''} \theta^{y''_s}$$

What this means is that we simply can remove the column in S_{-s} corresponding to the reaction $y'' \to y'''$ and add up the entries in v_{-s} corresponding to both reactions into $(v_{-s})_{y \to y'}$. And while the original equation with duplicate equations

$$f_{-s}(c_{-s}) = S_{-s} v_{-s}(c_{-s})$$

still holds, this new definition of S_{-s} and v_{-s} lets us define the reduced network that has steady states at $f_{-s}(c^*_{-s}) = 0$.

Lemma 1. *Let $N = \{\mathcal{S}, \mathcal{C}, \mathcal{R}, k\}$ be a mass action network. For $\theta \in \mathbb{R}$ define $N_{-s}(\theta) = \{\mathcal{S}_{-s}, \mathcal{C}_{-s}, \mathcal{R}_{-s}, k_{-s}(\theta)\}$*

$$\mathcal{S}_{-s} = \mathcal{S} \setminus \{s\}$$

$$\mathcal{C}_{-s} = \{y_{-s} \in \mathbb{R}^{(\mathcal{S}_{-s})} : y \in \mathcal{C}\}$$

$$\mathcal{R}_{-s} = \{y_{-s} \to y'_{-s} : y \to y' \in \mathcal{R}\}$$

and

$$k_{y_{-s} \to y'_{-s}}(\theta) = \sum_{y'' \to y''' \in \mathcal{R} \,:\, y_{-s} = y''_{-s}, y'_{-s} = y'''_{-s}} k_{y'' \to y'''} \theta^{y''_s}$$

then for any steady state c^* of N such that $c^*_s = \theta$ we get that c^*_{-s} is a steady state of N_{-s}.

Now the question that remains is how we can use this to analyze larger networks. In Example 2 the original network has a structural deficiency of one while the reduced network has a structural deficiency of zero. Using Theorem 2 we are able to infer concentration robustness in species A in the reduced network. This means we know that in the original network at steady state for any fixed concentration of species B, the concentration of A will also have to be constant, i.e. for any two steady states $(\pi_1, \theta), (\pi_2, \theta)$ we have $\pi_1 = \pi_2$. If we further can show that in the reduced network the concentration c^*_A at steady state does not depend on θ, then we have that for any two steady states $(\pi_1, \theta_1), (\pi_2, \theta_2)$ of the original network we must have $\pi_1 = \pi_2$ which means species A is also robust in the original network. Before we discuss the theory that makes it possible for us to do that let us take a quick look at another way to reduce a network.

5.1.2 Species Merging

Having discussed our concept of species removal in some detail we will keep this section relatively brief as it is much the same. The goal is to reduce a network N by merging two species A, B. This will be possible by fixing the ratio of their concentrations

$$c^*_A = \theta c^*_B$$

since then the species formation rate function of N can be rewritten as

$$\begin{aligned} f(c^*) &= \sum_{y \to y' \in \mathcal{R}} k_{y \to y'} (c^*)^y (y' - y) \\ &= \sum_{y \to y' \in \mathcal{R}} k_{y \to y'} (c^*_A)^{y_A} (c^*_{-A})^y_{-A} (y' - y) \\ &= \sum_{y \to y' \in \mathcal{R}} k_{y \to y'} (\theta c^*_B)^{y_A} (c^*_{-A})^y_{-A} (y' - y) \\ &= \sum_{y \to y' \in \mathcal{R}} k_{y \to y'} \theta^{y_A} (c^*_B)^{y_A + y_B} (c^*_{-A,B})^{y_{-A,B}} (y' - y) \end{aligned}$$

The reduced network will not have species A but instead A will be merged into species B

$$(y_{A\cup B})_B = y_A + y_B$$

for all complexes $y \in \mathcal{C}$ while keeping all other species the same. With that let c^* be a steady state of N then

$$\begin{aligned} f(c^*) &= \sum_{y\to y'\in\mathcal{R}} k_{y\to y'}\theta^{y_A}(c^*_B)^{y_{A\cup B}}(c^*_{-A,B})^{y_{-A,B}}(y'-y) \\ &= \sum_{y\to y'\in\mathcal{R}} k_{y\to y'}\theta^{y_A}(c^*_{-A})^{y_{A\cup B}}(y'-y) \end{aligned}$$

which means defining

$$f_{A\cup B}(c_{-A}) = \sum_{y\to y'\in\mathcal{R}} k_{y\to y'}\theta^{y_A}(c_{-A})^{y_{A\cup B}}(y'_{A\cup B} - y_{A\cup B})$$

gives us a network akin to the one we extracted in the last section, the only change being that instead of removing a row from S this time around we sum up two rows. For any steady state c^* of the original network such $c^*_A = \theta c^*_B$, will imply c^*_{-A} to be a steady state of the reduced network since for any species s other than B we have

$$\begin{aligned} (f_{A\cup B})_s(c^*_{-A}) &= \sum_{y\to y'\in\mathcal{R}} k_{y\to y'}\theta^{y_A}(c^*_{-A})^{y_{A\cup B}}((y'_{A\cup B})_s - (y_{A\cup B})_s) \\ &= \sum_{y\to y'\in\mathcal{R}} k_{y\to y'}\theta^{y_A}(c^*_{-A})^{y_{A\cup B}}(y'_s - y_s) \\ &= f_s(c^*) \\ &= 0 \end{aligned}$$

and for $s = B$

$$\begin{aligned} (f_{A\cup B})_B(c^*_{-A}) &= \sum_{y\to y'\in\mathcal{R}} k_{y\to y'}\theta^{y_A}(c^*_{-A})^{y_{A\cup B}}((y'_{A\cup B})_B - (y_{A\cup B})_B) \\ &= \sum_{y\to y'\in\mathcal{R}} k_{y\to y'}\theta^{y_A}(c^*_{-A})^{y_{A\cup B}}(y'_A + y'_B - y_A - y_B) \\ &= f_A(c^*) + f_B(c^*) \\ &= 0 \end{aligned}$$

Lemma 2. *Let $N = \{\mathcal{S}, \mathcal{C}, \mathcal{R}, k\}$ be a mass action network. For $\theta \in \mathbb{R}$ define $N_{s_1\cup s_2}(\theta) = \{\mathcal{S}_{s_1\cup s_2}, \mathcal{C}_{s_1\cup s_2}, \mathcal{R}_{s_1\cup s_2}, k_{s_1\cup s_2}(\theta)\}$*

$$\mathcal{S}_{s_1\cup s_2} = \mathcal{S}\setminus\{s_1\}$$

$$\mathcal{C}_{s_1\cup s_2} = \{y_{s_1\cup s_2} \in \mathbb{R}^{(\mathcal{S}_{-s})} : y \in \mathcal{C}\}$$

$$\mathcal{R}_{s_1\cup s_2} = \{y_{s_1\cup s_2} \to y'_{s_1\cup s_2} : y \to y' \in \mathcal{R}\}$$

and

$$k_{y_{s_1\cup s_2}\to y'_{s_1\cup s_2}}(\theta) = \sum_{y''\to y'''\in\mathcal{R}\,:\, y_{s_1\cup s_2}=y''_{s_1\cup s_2}, y'_{s_1\cup s_2}=y'''_{s_1\cup s_2}} k_{y''\to y'''}\theta^{y''_{s_1}}$$

then for any steady state c^* of N such that $c^*_{s_1} = \theta c^*_{s_2}$ we get that $c^*_{-s_1}$ is a steady state of $N_{s_1\cup s_2}$.

When in future sections we discuss reduced networks, we will mean those networks obtained through a series of species removal and species merging.

5.1.3 Properties of Reduced Networks

Let us take a look at two useful properties that arise in reduced networks. For a network N let $N' = N'(\theta_1, ..., \theta_k)$ be a network reduced k times at each stage i either fixing a species concentration $c_{s_i} = \theta_i$ or the ration of two $c_{s_i} = \theta_i c_{s'_i}$. Let further δ'_s be the structural deficiency of N'.

Proposition 3. *If $\delta'_s = 0$ then N has positive steady states only if N' is weakly reversible.*

Proof. Assume there exists a steady state c^* of N. Let

$$N^* = N^*(c^*_{s_1}, ..., c^*_{s_k})$$

be the network upon the same series of reduction, however instead of fixing the corresponding concentration or ratio thereof to θ_i it is fixed to $c^*_{s_i}$. By Lemmas 1 and 2, $c^*_{-s_1,...,s_k}$ will be a steady state of N^* and since c^* is positive so will $c^*_{-s_1,...,s_k}$. Further, by definition of reduced networks, we know that the species, complexes and reactions of the reduced network do not depend on the exact value the concentration or ratio is fixed to meaning that $S' = S^*$, $C' = C^*$ and $R' = R^*$. This gives us $Y' = Y^*$ and $\Lambda' = \Lambda^*$ and therefore $\delta'_s = \delta^*_s$. Now, if $\delta'_s = 0$ then $\delta^*_s = 0$ and with Theorem 1 we know that N^* is weakly reversible since it has a positive steady state. Since $\Lambda' = \Lambda^*$, we further know that N^* will be weakly reversible if and only if N' is weakly reversible. The claim follows immediately. □

Proposition 4. *Let $\mathcal{L} = \{c^* : f(c^*) = 0\}$ be the set of steady states in N. Let further $L(\theta_1, ..., \theta_k)$ be the set of all concentration vectors $\hat{c}$ compatible with the reduction, meaning that if in the i-th step we fix the value of $c_{s_i} = \theta_i$ then $\hat{c}_{s_i} = \theta_i$ for any $\hat{c} \in L(\theta_1, ..., \theta_k)$ and if we fixed the ratio $c_{s_i} = \theta_i c_{s'_i}$ then $\hat{c}_{s_i} = \theta_i \hat{c}_{s'_i}$ for any $\hat{c} \in L(\theta_1, ..., \theta_k)$. If*

$$\mathcal{L} \subset L(\theta_1, ..., \theta_k)$$

then for any steady state c^* of N, $c^*_{s_1,...,s_k}$ will be a steady state of N'.

Proof. By definition of a N', for any steady state $c^* \in \mathcal{L}$ of N compatible with the reductions from N to N', the concentration vector $c^*_{s_1,...,s_k}$ will be a steady state of N'. In other words, for any

$$c^* \in \mathcal{L} \cap L(\theta_1, ..., \theta_k)$$

$c^*_{s_1,...,s_k}$ is a steady state of N'. If $\mathcal{L} \subset L(\theta_1, ..., \theta_k)$ then

$$\mathcal{L} \cap L(\theta_1, ..., \theta_k) = \mathcal{L}$$

and the claim of the proposition follows. □

5.2 Structural Couplings

For a network N let $N' = N'(\theta_1, ..., \theta_k)$ be a reduced network. Assuming the network N' has a deficiency of one or zero can use Theorems 2 or 3 in order to find coupled complexes and robust species in N'. However, using these theorems gives us results of the form

$$\frac{c^*_y}{c^*_{y'}} = \pi(k')$$

and since $k' = k'(\theta_1, .., \theta_k)$ depends on the fixed values this will be generalizable to the original network only if all steady states are compatible with the reduction as shown in Proposition 4. Since this is generally not the case we will need to find a new way to extract couplings, preferable one that depends on the structure rather than the kinetics since the structure is maintained when changing the fixed values θ_i.

Lemma 3. *For a network $N = \{\mathcal{S}, \mathcal{C}, \mathcal{R}, k\}$ let $N'(\theta) = \{\mathcal{S}', \mathcal{C}', \mathcal{R}', k'(\theta)\}$ a network reduced once. Assume that two reactions $y_1 \to y_1'$ and $y_2 \to y_2'$ in the reduced network N' are coupled, meaning that for any steady state $c^* \in \mathbb{R}^{\mathcal{S}'}$ we have*

$$v'_{y_1 \to y_1'}(c^*) = \pi v'_{y_2 \to y_2'}(c^*)$$

where π does not depend on the kinetics k' of the reduced network. If for $y \in \mathcal{C}$ we define $\hat{y}$ as the complex upon reduction, i.e. either y_{-s} or $y_{s \cap s'}$ depending on the method of reduction we used, then for any steady state $c^{**} \in \mathbb{R}^{\mathcal{S}}$ of original network N then

$$\sum_{y \to y' \in \mathcal{R} \,:\, \hat{y} = y_1 \,,\, \hat{y}' = y_1'} v_{y \to y'}(c^{**}) = \pi \sum_{y \to y' \in \mathcal{R} \,:\, \hat{y} = y_2 \,,\, \hat{y}' = y_2'} v_{y \to y'}(c^{**})$$

meaning that in the original network the sum of reactions which yielded $y_1 \to y_1'$ and $y_2 \to y_2'$ upon reduction will be coupled.

Proof. If the coupling in the reduced network does not depend on the kinetics k' then it will can not depend on θ and must hold across every reduction of the same type $N'(\theta)$ regardless of the specific value of the fixed constant θ. By definition for any $N'(\theta)$ we have

$$v'_{y_1 \to y_1'}(c^{**}_{-s}) = \sum_{y \to y' \in \mathcal{R} \,:\, \hat{y} = y_1 \,,\, \hat{y}' = y_1'} v_{y \to y'}(c^{**})$$

and the same for $y_2 \to y_2'$. Since for any steady state c^{**} of N we know that c^*_{-s} is a steady state of $N'(c^*_s)$ the claim follows. □

This Lemma identifies couplings which can be extrapolated from a reduced network to the original network. All we need is coupled reactions in the reduced network. In order to get those we follow the ideas in [29] which we presented in Section 4.1. Therefore let us quickly remember the arguments used there. For any network we could decompose the species formation rate function into

$$f(c) = Y A_k \psi(c)$$

Finding the structure of $\ker(A_k)$ then allowed us to infer

$$\psi_y(c^*) = \pi\psi_{y'}(c^*)$$

with $\pi = \pi(k)$ when $\delta_d = \dim(\mathrm{Im}(A_k) \cap \ker(Y))$ was either zero or one. Analogously we can decompose f into

$$f(c) = Y\Lambda v(c)$$

meaning that if we can find $\ker(\Lambda)$ by analogy we will be able to infer

$$v_{y_1 \to y_1'}(c^*) = \pi v_{y_2 \to y_2'}(c^*)$$

if $\delta_s = \dim(\mathrm{Im}(\Lambda) \cap \ker(Y))$ is zero or one. With that let us take a look at $\ker(\Lambda)$. We recall that Λ gives us the structure of the network graph having for each reaction $y \to y' \in \mathcal{R}$ a column with

$$\Lambda_{y \to y'} = \omega^y - \omega^{y'}$$

Let us assume there are k complexes $y_1, ..., y_k \in \mathcal{C}$ such that we have a set of reactions $\{r_1, ..., r_k\} \subset \mathcal{R}$ for which either $r_i = y_i \to y_{i+1}$ or $r_i = y_{i+1} \to y_i$ where $y_{k+1} := y_1$ for convenience. Let η_i denote the direction of the i-th reaction, i.e.

$$\eta_i = \begin{cases} 1 & \text{if } r_i = y_i \to y_{i+1} \\ -1 & \text{otherwise} \end{cases}$$

Defining η that way makes it so that multiplying η_i by the column Λ_{r_i} corresponding to the i-th reaction in the cycle gives us either

$$\eta_i\Lambda_{r_i} = \omega^{y_{i+1}} - \omega^{y_i}$$

if the reaction goes from y_i to y_{i+1} or

$$\eta_i\Lambda_{r_i} = (-1)(\omega^{y_i} - \omega^{y_{i+1}})$$

if it goes from y_{i+1} to y_i. In any case we have

$$\begin{aligned} \sum_{i=1}^{k} \eta_i\Lambda_{r_i} &= \sum_{i=1}^{k} \omega^{y_{i+1}} - \omega^{y_i} \\ &= \omega^{y_{k+1}} - \omega^{y_1} \\ &= \omega^{y_1} - \omega^{y_1} \\ &= 0 \end{aligned} \tag{1}$$

which means defining $v \in \mathbb{R}^{\mathcal{R}}$ such that

$$v_r = \begin{cases} \eta_i & \text{if } r = r_i \\ 0 & \text{otherwise} \end{cases}$$

gives us $v \in \ker(\Lambda)$. Let Υ be the set of all v that can be found in the above way, meaning that we have complexes $y_1, ..., y_k$ and reactions $r_1, ..., r_k$ such that

$$r_i = y_i \to y_{i+1} \text{ or } r_i = y_{i+1} \to y_i$$

and directions

$$\eta_i = \begin{cases} 1 & \text{if } r_i = y_i \to y_{i+1} \\ -1 & \text{otherwise} \end{cases}$$

for each of the reactions and find υ as

$$\upsilon_r = \begin{cases} \eta_i & \text{if } r = r_i \\ 0 & \text{otherwise} \end{cases}$$

Then $\mathrm{span}(\Upsilon)$ is called the cycle space of the graph and $\upsilon \in \mathrm{span}(\Upsilon)$ a cycle function. As a shorthand we will call υ a cycle and say that a reaction $r \in \mathcal{R}$ is involved in cycle υ if $\upsilon_r \neq 0$. The following two results define the relationship between the incidence matrix of a graph and its cycle space and give a way to decompose the cycle space [39].

Lemma 4. *For any network we have* $\ker(\Lambda) = \mathrm{span}(\Upsilon)$.

By definition of Υ and Eq. 1 we already showed that $\mathrm{span}(\Upsilon) \subset \ker(\Lambda)$. The remainder of the proof for this lemma can be found in the Appendix. For now let us concentrate on how we can use this. The lemma tells us that in a network with

$$\dim(\ker(\Lambda)) = \gamma$$

any set of independent cycles $\Gamma = \{\upsilon_1, ..., \upsilon_\gamma\}$ is a basis of $\ker(\Lambda)$.

Lemma 5. *Assume we have two reactions* $r_1, r_2 \in \mathcal{R}$ *such that any cycle that involves the one also involves the other, i.e.*

$$\upsilon_{r_1} = 0 \iff \upsilon_{r_2} = 0$$

Then there exists a set of independent cycles

$$\Gamma' = \{\upsilon'_1, ..., \upsilon'_\gamma\}$$

such that $(\upsilon'_i)_{r_1} = (\upsilon'_i)_{r_2} = 0$ *for all* $i > 1$.

Again the proof can be found in the Appendix. With that we have a result akin to Proposition 2 and can therefore make claims similar to those in Theorems 2 and 3.

Theorem 4. *Let* N *be a network with* $\delta_s = 0$ *then any two reactions* $y_1 \to y'_1$ *and* $y_2 \to y'_2$ *as described in Lemma 4 are coupled, meaning that there exists a constant* π *such that for any positive steady state* c^* *we have*

$$\frac{v_{y_1 \to y'_1}(c^*)}{v_{y_2 \to y'_2}(c^*)} = \pi$$

If N *is instead a network with* $\delta_s = 1$ *then any two reactions* $y_1 \to y'_1$ *and* $y_2 \to y'_2$ *not involved in any cycles are coupled.*

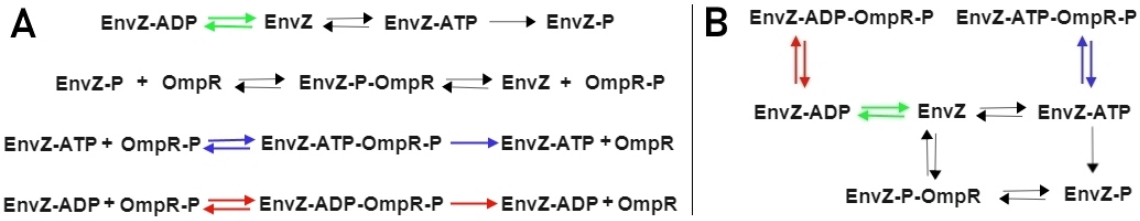

Figure 4: The left reaction network (A), taken from [19], shows the EnvZ/OmpR two-component osmoregulator in E. coli from Example 3. The right network (B) shows the left network upon removal of species $OmpR$ and $OmpR - P$

The proof can be found in the Appendix. Since π does not depend on k, if for a network N we have a reduced network $N' = N'(\theta_1, ..., \theta_k)$ with $\delta'_s = 0$ and with Theorem 4 we find

$$\frac{v_{y_1 \to y'_1}(c^*)}{v_{y_2 \to y'_2}(c^*)} = \pi$$

then this holds true for any combination of fixed constants $\theta_1, ..., \theta_k$. With Lemma 3 this means we can extrapolate these findings to the original network. A small corollary to this result is that for a positive steady state to exist, π has to be positive. The proof of Theorem 4 shows that this means any two reaction $y_1 \to y'_1$ and $y_2 \to y'_2$ as described in Lemma 4 have to have the same direction in the cycles they are involved in which implies $\pi = 1$.

Having this theorem we could theoretically create an algorithm that reduces a network to a deficiency zero or one network, finds coupled reactions in that network, and extrapolates them to the larger network. Let us take the network from Example 3 to illustrate (Figure 4 A). It has a deficiency of two and we so Theorems 2 or 3 are not applicable. What we can do however is reduce the network and find couplings that way. In Figure 4 B we can see the network N', which is N upon removal of species OmpR and OmpR-P. Now N' is of structural deficiency zero, meaning we can apply the first part of Theorem 4. Therefore the reactions marked in the same colour will be coupled, since any cycle involving the one will also have to involve the other. In the original network we marked the reactions that upon species removal combine into these reactions in the same colour. For example we see that the green reactions have a one to one relation between networks and are therefore also coupled in the original network giving us that the species EnvZ-ADP and EnvZ are coupled, meaning their ratio is fixed at every steady state.

If we now merge these two species by fixing their ratio to exactly this value then from Proposition 4 we know that any steady state of the original network will have the reduced state be a steady state of the reduced network and if we find ACR in the reduced network it is also found in the original network. We note that it is not important for us to know this fixed value since we are not interested in the exact concentration of the ACR species, but instead are making a qualitative claim.

Using that approach in the case of this network will however not give us any further results, so instead let us go back to investigating N'. As a next step let us merge all species connected by coloured reaction, i.e. EnvZ-ADP-OmpR-P into EnvZ-ADP

(red), EnvZ-ADP into EnvZ (green) and EnvZ-ATP-OmpR-P into EnvZ-ATP. What we remain with is the core of black reactions. The cycles in that network lie in a way so that we cannot utilize Theorem 4. Let us discuss an alternative approach.

5.3 Complex Balance

Since we have a network with $\delta_s = 0$, for any steady state c^* we must have $\Lambda v(c^*) = 0$. In particular for every row Λ^y of Λ we must have

$$\langle \Lambda^y, v(c^*) \rangle = 0$$

By the definition of Λ we know that the entry of Λ at (y, r) is 1 if y is the product of reaction r and -1 if it is the substrate, meaning that we have

$$\Lambda^y_{y' \to y''} = \begin{cases} 1 & \text{if } y = y'' \\ -1 & \text{if } y = y' \\ 0 & \text{otherwise} \end{cases}$$

It follows that

$$\langle \Lambda^y, v(c^*) \rangle = \sum_{y' \to y \in \mathcal{R}} v_{y' \to y}(c^*) - \sum_{y \to y' \in \mathcal{R}} v_{y \to y'}(c^*)$$

which with the above gives us

$$\sum_{y' \to y \in \mathcal{R}\,,\, y' \in \mathcal{C}} v_{y' \to y}(c^*) = \sum_{y \to y' \in \mathcal{R}\,,\, y' \in \mathcal{C}} v_{y \to y'}(c^*)$$

In that case we say that y is a balanced complex and if every complex in a network is balanced, we say that the network is complex-balanced. We just showed that every network with $\delta_s = 0$ is complex-balanced [8]. Using this together with Lemma 3 yields the following theorem.

Theorem 5. *Let N be a reaction network and let $N' = N'(\theta_1, ..., \theta_k)$ be a reduced network of N. For $y \in \mathcal{C}$ we define $\hat{y} \in \mathcal{C}'$ as the complex y upon reduction. If the structural deficiency of N' is zero, then for every complex $y^{red} \in \mathcal{C}'$ we have*

$$\sum_{y' \in \mathcal{C}\,,\, y \to y' \in \mathcal{R} \,:\, y^{red} = \hat{y}} v_{y \to y'}(c^*) = \sum_{y' \in \mathcal{C}\,,\, y' \to y \in \mathcal{R} \,:\, y^{red} = \hat{y}} v_{y' \to y}(c^*)$$

for every positive steady state c^* of N.

Now with this we know that if we reduce a network to a deficiency zero network, then for every complex in that reduced network, the sum of all original complexes that correspond to that reduced complex will be balanced in sum. We will illustrate that in the example from above but before that let us mention the following consequence of Theorem 5.

Corollary 1. *Let N be a reaction network and let N' be a reduced network of N. For $y \in \mathcal{C}$ we define $y^{red} \in \mathcal{C}'$ as the complex y upon reduction. If the structural deficiency of N' is zero, and there exists a complex $\hat{y} \in \mathcal{C}'$ in N' such that there is exactly one complex $y \in \mathcal{C}$ in N with $y^{red} = \hat{y}$, then y is a balanced complex in N.*

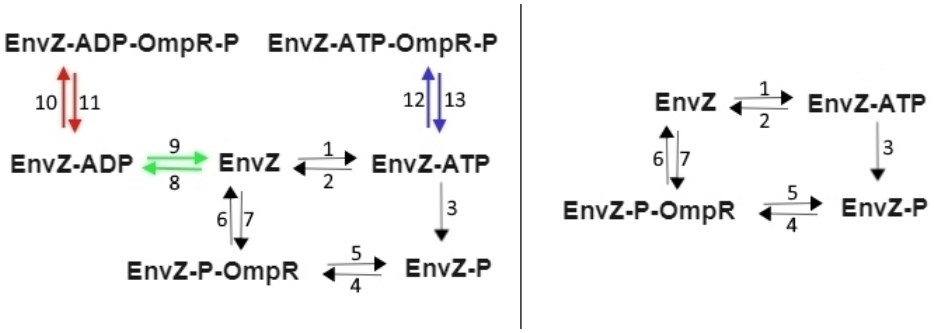

Figure 5: The left network (A) is a reduced network introduced in Figure 4 and the right (B) is the network upon further reduction by merging species.

With that let us return to the example from Figure 4. We already saw how the network looks upon removal of two species and what couplings we could extract from that with the help of Theorem 4. We also mentioned that we will continue reduction by merging species. Both reduced networks (Figure 5) have a structural deficiency of zero so that in both cases we can apply Theorem 5. Using this we can for example see that in both cases EnvZ-ATP is balanced. For the first network the corresponding complexes in the original network are

$$\{\text{EnvZ-ATP, EnvZ-ATP + OmpR-P, EnvZ-ATP + OmpR}\}$$

i.e. those that are just EnvZ-ATP in addition to one or more of the removed species. In the second network we merged EnvZ-ATP with EnvZ-ATP-OmpR-P meaning the respective corresponding complex sets merge as well and we get that now the set of complexes corresponding to EnvZ-ATP is

$$\{\text{EnvZ-ATP, EnvZ-ATP + OmpR-P, EnvZ-ATP + OmpR, EnvZ-ATP-OmpR-P}\}$$

Theorem 5 tells us that both of these sets are balanced in sum. The corresponding equation from Theorem 5 was

$$\sum_{y' \in \mathcal{C}\,,\, y \to y' \in \mathcal{R} \,:\, y^{red} = \hat{y}} v_{y \to y'}(c^*) = \sum_{y' \in \mathcal{C}\,,\, y' \to y \in \mathcal{R} \,:\, y^{red} = \hat{y}} v_{y' \to y}(c^*)$$

Assuming we have a reaction r in the original network such that both sides, i.e. substrate and product, correspond to the same complex $\hat{y}$ in the reduced network, then for that complex we will find $v_r(c^*)$ on both sides of this equation and for all others it will be on neither. We can therefore safely disregard it. The same logic applies to the following

Proposition 5. *If for two sets of complexes we have $C_1 \subset C_2 \subset \mathcal{C}$ and both C_1 and C_2 are balanced in sum, then $C_2 \setminus C_1$ is balanced in sum.*

Proof. Since C_2 is balanced in sum we have

$$\sum_{y \in C_2\,,\, y' \in \mathcal{C}\,,\, y \to y' \in \mathcal{R}} v_{y \to y'}(c^*) = \sum_{y \in C_2\,,\, y' \in \mathcal{C}\,,\, y' \to y \in \mathcal{R}} v_{y' \to y}(c^*)$$

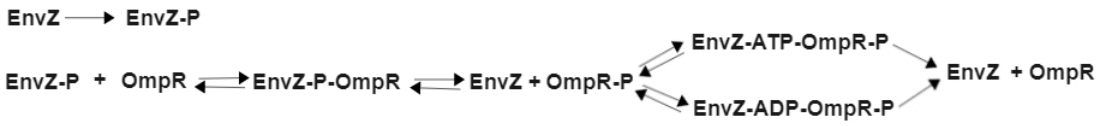

Figure 6: The network from Figure 4 upon merging of species EnvZ, EnvZ-ATP and EnvZ-ADP.

and since C_1 is a subset of C_2 we can rewrite this as

$$\begin{aligned}&\sum_{y\in C_1\,,\,y'\in\mathcal{C}\,,\,y\to y'\in\mathcal{R}} v_{y\to y'}(c^*) + \sum_{y\in C_2\setminus C_1\,,\,y'\in\mathcal{C}\,,\,y\to y'\in\mathcal{R}} v_{y\to y'}(c^*)\\ =&\sum_{y\in C_1\,,\,y'\in\mathcal{C}\,,\,y'\to y\in\mathcal{R}} v_{y'\to y}(c^*) + \sum_{y\in C_2\setminus C_1\,,\,y'\in\mathcal{C}\,,\,y'\to y\in\mathcal{R}} v_{y'\to y}(c^*)\end{aligned}$$

which with C_1 being balanced in sum yields

$$\sum_{y\in C_2\setminus C_1\,,\,y'\in\mathcal{C}\,,\,y\to y'\in\mathcal{R}} v_{y\to y'}(c^*) = \sum_{y\in C_2\setminus C_1\,,\,y'\in\mathcal{C}\,,\,y'\to y\in\mathcal{R}} v_{y'\to y}(c^*)$$

as claimed. □

In our example this means $\{EnvZ-ATP\}$ is balanced in sum. Since this set contains just a single complex we get that EnvZ-ATP is a balanced complex in the original network. From the resulting equations we can see that EnvZ-ATP is coupled to EnvZ, i.e.

$$c^*_{EnvZ-ATP} = \theta^*_1 c^*_{EnvZ}$$

for every steady state c^* in N. From the last section we already know that EnvZ is coupled to EnvZ-ADP, i.e.

$$c^*_{EnvZ-ADP} = \theta^*_2 c^*_{EnvZ}$$

Figure 6 shows the structure of $N'' = N''(\theta^*_1, \theta^*_2)$, the network upon merging EnvZ-ATP and EnvZ by fixing their ratio to θ^*_1 and then merging EnvZ-ADP and EnvZ by fixing their ratio to θ^*_2. The resulting network N'' has a deficiency of one, meaning that we can apply Theorem 3 and find that all complexes in non-terminal strong linkage classes are coupled in that network. Particularly we find EnvZ and EnvZ + OmpR-P to be coupled, meaning that we have ACR in the species OmpR-P in the reduced network $N''(\theta^*_1, \theta^*_2)$. Since any steady state from the original network is compatible with that reductions of N'', Proposition 4 tells us that we also have ACR in OmpR-P in the original network.

5.4 Kinetic Restrictions

Before we go ahead and apply our results to more networks we will discuss one final approach to find coupled reactions. Let us use the right network in Figure 5 illustrate the idea. To start off, this network has four species and for each of those there is exactly one complex with that species, meaning that Y will be the identity and we will have

$$S = \Lambda = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 1 & -1 \\ 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & -1 & 1 \end{pmatrix}$$

and obviously therefore $\delta_s = 0$. Next, by looking at the graph we can easily find the following independent cycles $\upsilon_1, ..., \upsilon_4$.

$$\begin{pmatrix}1\\1\\0\\0\\0\\0\\0\end{pmatrix}, \begin{pmatrix}0\\0\\0\\1\\1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\0\\0\\0\\0\\1\\1\end{pmatrix}, \begin{pmatrix}1\\0\\1\\1\\0\\1\\0\end{pmatrix}$$

Now, since these cycles are a basis of $\ker(\Lambda)$ and therefore $\ker(S)$, any steady state will have to satisfy

$$v(c^*) = a_1\upsilon_1 + a_2\upsilon + a_3\upsilon_3 + a_4\upsilon_4$$

which was how we got Theorem 4. Now we want to involve the kinetics of the system, meaning we want to use

$$f(c^*) = Y\Lambda K\psi(c^*)$$

or rather $v(c^*) = K\psi(c^*)$ and therefore

$$K\psi(c^*) = a_1\upsilon_1 + a_2\upsilon + a_3\upsilon_3 + a_4\upsilon_4$$

which means we can restrict the space of possible coefficients $a_1, ..., a_4$. In our example we have K

$$K = \begin{pmatrix}k_1 & 0 & 0 & 0\\ 0 & k_2 & 0 & 0\\ 0 & k_3 & 0 & 0\\ 0 & 0 & k_4 & 0\\ 0 & 0 & 0 & k_5\\ 0 & 0 & 0 & k_6\\ k_7 & 0 & 0 & 0\end{pmatrix}$$

This means that any v such that $v = Kx$ will have the form

$$v = (k_1x_1, k_2x_2, k_3x_2, k_4x_3, k_5x_4, k_6x_4, k_7x_1)^T$$

Since $k_i > 0$ this means that

$$\mathrm{Im}(K) = \left\{v : \frac{v_2}{v_3} = \frac{k_2}{k_3}, \frac{v_5}{v_6} = \frac{k_5}{k_6}, \frac{v_7}{v_1} = \frac{k_7}{k_1}\right\}$$

These are exactly the junctions in the graph. Define a as the vector containing all the coefficients a_i and Υ as the matrix containing all the cycles υ_i, i.e. $a = (a_1, a_2, a_3, a_4)^T$ and $\Upsilon = (\upsilon_1, \upsilon_2, \upsilon_3, \upsilon_4)$ then we know that

$$\Upsilon a \in \mathrm{Im}(K)$$

If we further define $\pi_1 = \frac{k_2}{k_3}, \pi_2 = \frac{k_5}{k_6}, \pi_4 = \frac{k_7}{k_1}$ and $\Upsilon_{2,5,7}$ as the matrix containing the second, fifth and first rows of Υ in that order and $\Upsilon_{3,6,1}$ analogously we get

$$(\Upsilon_{2,5,7})a = \begin{pmatrix}\pi_1 & 0 & 0\\ 0 & \pi_2 & 0\\ 0 & 0 & \pi_3\end{pmatrix}(\Upsilon_{3,6,1})a$$

which plugging in the values of Υ gives us

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} \pi_1 & 0 & 0 \\ 0 & \pi_2 & 0 \\ 0 & 0 & \pi_3 \end{pmatrix} \begin{pmatrix} a_4 \\ a_3 + a_4 \\ a_1 + a_4 \end{pmatrix}$$

Solving this yields

$$a = \begin{pmatrix} \pi_1 \\ \pi_2(\pi_3(\pi_1 + 1) + 1) \\ \pi_3(\pi_1 + 1) \\ 1 \end{pmatrix} \alpha$$

for some $\alpha \in \mathbb{R}$ and we finally get

$$v(c^*) = \Upsilon a = \begin{pmatrix} \pi_1 + 1 \\ \pi_1 \\ 1 \\ \pi_2(\pi_3(\pi_1 + 1) + 1) + 1 \\ \pi_2(\pi_3(\pi_1 + 1) + 1) \\ \pi_3(\pi_1 + 1) + 1 \\ \pi_3(\pi_1 + 1) \end{pmatrix} \alpha$$

It now follows for example that

$$\begin{aligned} \frac{v_{r_1}(c^*)}{v_{r_2}(c^*)} &= \frac{\pi_1 + 1}{\pi_1} \\ &= \frac{k_2/k_3 + 1}{k_2/k_3} \\ &= \frac{k_2 + k_3}{k_2} \end{aligned}$$

We therefore get that in the reduced network EnvZ and EnvZ-ATP are coupled with constant ratio $\frac{k_2+k_3}{k_2}$. This result can be extrapolated to the original network since neither k_2 nor k_3 of the reduced network depend on the fixed constants $\theta_1, ..., \theta_5$.

We can generalize this idea in order to analyze networks where the above methods do not give much or any useful information. Let N be a reaction network and δ_s be its structural deficiency. Let further $\Gamma = \{\upsilon_1, ..., \upsilon_\gamma\}$ be a basis of $\ker(\Lambda)$ and $x_1, ..., x_{\delta_s}$ be a basis of $\ker(Y) \cap \mathrm{Im}(\Lambda)$. Define $\Delta = \{\upsilon'_1, ..., \upsilon'_{\delta_s}\}$ to be any set such that $x_i = \Lambda\upsilon'_i$. Then $\Gamma \cup \Delta$ is a basis of $\ker(Y\Lambda)$ and any steady state will have to fulfill

$$v(c^*) = \sum_{i=1}^{\gamma} a_i \upsilon_i + \sum_{i=1}^{\delta_s} a'_i \upsilon'_i$$

Let Υ be the matrix that has υ_i as its columns and Υ' the matrix that has υ'_i as its columns. Let further R_1, R_2 be the ordered subsets of the set of reactions containing the reactions contributing to junctions of the graph, meaning that if we number the junctions, for junction j, the j-th entries in R_1 and R_2 are exactly the reactions involved in that junction. Then for $v(c^*) \in \mathrm{Im}(K)$ we need to have

$$\Upsilon_{R_1} a_{R_1} + \Upsilon'_{R_1} a'_{R_1} = \mathrm{diag}(\pi)(\Upsilon_{R_2} a_{R_2} + \Upsilon'_{R_2} a'_{R_2})$$

where

$$\pi_j = \frac{k_{r_j}}{k_{r'_j}}$$

for r_j, r'_j respectively the j-th entry in R_1 and R_2. In general determining the space of possible $v(c^*)$ is too computationally expensive to be useful. However using Theorems 2 we can find networks that yield results using this approach. Let $\delta_s = 0$ then from Theorem 2 we know that any two complexes in the same linkage class are coupled and therefore any two reactions in the same linkage class are coupled. It follows that

$$v(c^*) = \alpha_1 v_{L_1} + ... + \alpha_l v_{L_l}$$

where v_{L_i} is different from zero only on reactions in the linkage class L_i. Furthermore, we can choose the set of cycles Γ in a way so that every $v_i \in \Gamma$ is different from zero only on a single linkage class meaning that Υ in made up of blocks for each linkage class. By extension, the same holds for Υ_{R_1} and Υ_{R_2}. Since N has $\delta_s = 0$ we have

$$v(c^*) = \Upsilon a$$

and the space of possible $v(c^*)$ is kinetically restricted to

$$\Upsilon_{R_1} a_{R_1} = \mathrm{diag}(\pi)\Upsilon_{R_2} a_{R_2}$$

Since $\mathrm{diag}(\pi)$ is a diagonal matrix and Υ_{R_1} and Υ_{R_2} are block matrices, we can solve this system separately for each linkage class and find

$$v_{L_i}(c^*) = \alpha_i v_{L_i}$$

For each v_{L_i} we are now able to find which π_i and by extension which rate constants contribute to each entry. Let N be a new network and $N' = N'(\theta_1, ..., \theta_k)$ be a reduced network with $\delta'_s = 0$. If for a linkage class L_i in N' we find v_{L_i} such that $v'_{L_i}(c^*) = \alpha_i v_{L_i}$ and for two reactions r_1, r_2 in that linkage class we find that no rate constant that depends on $\theta_1, ..., \theta_k$ contributes to $(v_{L_i})_{r_1}$ or $(v_{L_i})_{r_2}$ then the corresponding reactions will be coupled in N since for any set of fixed values of $\theta_1, ..., \theta_k$, in the reduced network $N'(\theta_1, ..., \theta_k)$ the reactions r_1 and r_2 are coupled with the same constant.

6 A Network Reduction Algorithm

The approaches we presented to retrieve information from reduced networks were based on finding those reduced networks with a structural deficiency of either one or zero. Since calculating the deficiency for all possible reductions is computationally not feasible, in this section we will develop a heuristic for deciding which consecutive reductions will have the highest potential impact on deficiency.

For a network N, its structural deficiency δ_s is given by

$$\delta_s = n - l - s$$

where n is the number of complexes, l is the number of linkage classes, and s is the rank of the stoichiometric matrix S. Let $N' = N'(\theta)$ be the network upon one reduction and

$$\delta'_s = n' - l' - s'$$

its structural deficiency. Assuming that the reduction was the removal of species A. Then by definition we know that

$$\mathrm{Im}(S') = \mathrm{Im}(S_{-A})$$

where S_{-A} is the stoichiometric matrix of the original network upon removal of the row corresponding to species A. With that we also know

$$\mathrm{rank}(S') = \mathrm{rank}(S_{-A})$$

which since s is the dimension of the row space of S, i.e.

$$s = \mathrm{rank}(S) = \dim(\{S_B \,:\, B \in \mathcal{S} \setminus \{A\}\} \cup \{S_A\})$$

yields that either $s' = s$ if the row S_A corresponding to A lies in the row space of S_{-A} or $s - 1$ if it does not. Similarly if the reduction was the merging of species A and B, then

$$\mathrm{Im}(S') = \mathrm{Im}(S_{A\cup B})$$

where $S_{A\cup B}$ was S upon adding the row corresponding to A to that corresponding to B and then removing the row corresponding to A. Again we get that $s' = s$ if S_A is in the row space of $S_{A\cup B}$ and $s - 1$ if that is not the case. In conclusion know that s' will at most be 1 different from s. We will therefore focus on n' and l' when choosing a reduction. If for a complex $y \in \mathcal{C}$ we define $\hat{y}$ as y upon the reduction that yields N', then by definition

$$n' = |\mathcal{C}'| = |\{y' \,|\, y' = \hat{y}\}|$$

which ordering the set of complexes $\mathcal{C} = \{y_1, ..., y_n\}$ in N gives us

$$\begin{aligned} n' &= |\{y' \,|\, y' = \hat{y}_i\}| \\ &= |\{y_i \,|\, \forall j > i \,:\, \hat{y}_j \neq \hat{y}_i\}| \\ &= n - |\{y_i \,|\, \exists j > i \,:\, \hat{y}_j = \hat{y}_i\}| \end{aligned}$$

Let therefore

$$D' := \{y_i \,|\, \exists j > i \,:\, \hat{y}_j = \hat{y}_i\}$$

be the set of 'duplicates' upon reduction and $d' = |D'|$, then $n' = n - d'$. Similarly the number of linkage classes l' in N' will be determined by complexes in different linkage classes of N yielding the same complex upon reduction, i.e.

$$\begin{aligned} l' &= l - |\{L_i \,|\, \exists j > i\,,\, y_i \in L_i\,,\, y_j \in L_j \,:\, \hat{y}_i = \hat{y}_j\}| \\ &= l - |\{y \in \mathcal{C} \,|\, \exists j \in \{1, ..., l\}\,,\, y' \in L_j \,:\, y \in L_i\,,\, j > i\,,\, \hat{y} = \hat{y}'\}| \end{aligned}$$

Define

$$E' = \{y \in \mathcal{C} \,|\, \exists j \in \{1, ..., l\}\,,\, y' \in L_j \,:\, y \in L_i\,,\, j > i\,,\, \hat{y} = \hat{y}'\}$$

the set of complexes such that if y is in the i-th linkage class then there exists $j > i$ and $y' \in L_j$ which yields the same complex upon reaction as y. Let $e' = |E'|$ then we get $l' = l - e'$ and therefore finally

$$\delta'_s = n' - l' - s'$$

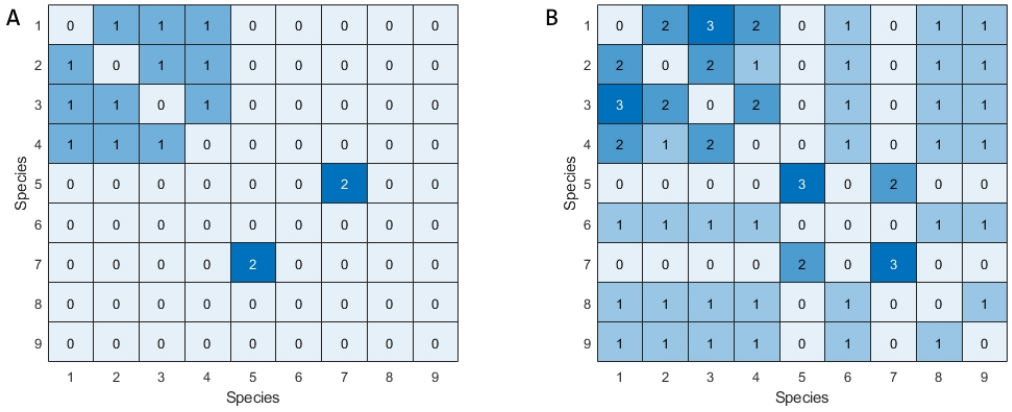

Figure 7: Heat maps for the matrices $P_d - P_e$ (A) and P_d (B) for the network from Example 3.

$$= n - d' - (l - e') - s'$$
$$= \delta_s - d' + e' + (s - s')$$

Throughout reduction we will want to maximize $d' - e'$.

Our heuristic will therefore be as follows. Define two matrices $P_d, P_e \in \mathbb{R}^{\mathcal{S}\times\mathcal{S}}$. For every two complexes y_i, y_j, if their difference $y_i - y_j$ is zero at every species except a single species s, then upon removal of that species, these complexes fall together. For two such complexes y_i, y_j, if they are in the same linkage class, then reduction will remove complexes within that linkage class and they contribute to D but not to E. We increase $P_d(s, s)$. If they are in different linkage classes, reduction will remove complexes by combining linkage classes, contributing to both D and E. We increase $P_d(s, s)$ and $P_e(s, s)$. If $y_i - y_j$ is different from zero in exactly two species s_1, s_2, then merging these species will combine the complexes. Analogous to the case of species removal, if y_i and y_j are in the same linkage class we increase $P_d(s_1, s_2)$ and if they are not we increase both $P_d(s_1, s_2)$ and $P_e(s_1, s_2)$. With that the entries of $P_d - P_e$ will give us a measure of the immediate impact of reduction with the corresponding species. Further P_e gives us a measure of potential impact at later reductions as combining linkage classes in the current reduction will make within linkage class complex reduction more likely in later reductions. Choosing our reduction target we will therefore weigh both of these matrices such that early reduction depend more heavily on d' while for later reductions will be chosen according to $d' - e'$.

In Figure 7 we can see these matrices for the network from Example 3. The matrix $P_d - P_e$ (A) suggest that merging of species OmpR (5) and OmpR-P (7) will have the most immediate benefit as a first reduction. The matrix P_d (B) suggests that a merging of species EnvZ-ADP (1) and EnvZ-ATP (3) or the removal of species OmpR or OmpR-P will have the highest potential benefit in reductions after the first.

7 Results

What follows are the results of network reduction when applied to two large-scale biochemical systems [20]. A more complete study of these networks with the methods

Figure 8: Upper left (A) shows the graph of the H. pylori network. Upper right shows a the graph of a reduced network after 81 steps. Under that (C) we have the heat map of P_d for the original H. pylori network.

discussed above is possible but due to the time constraints of this thesis not included. However these initial results sufficiently demonstrate the possibilities as well as the limits of network reduction as an approach identifying robustness in large-scale networks. All implementations can be found in the attached digital files or at `bit.ly/2p6lTAs`.

Helicobacter pylori [`bigg.ucsd.edu/models/iIT341`]
The first system is a metabolic reconstruction of Helicobacter pylori strain 26695, a human gastric pathogen infecting almost half of the world population [32]. The model can be found at the link provided above. The reaction network of the system contains 485 species in 993 complexes with 554 reactions between them and has a structural deficiency of 85. Its graph is shown in Figure 8 (A). A first test of our heuristic shows that removing species 19 combines 44 complexes across the network, a visualization of the complete matrix P_d can be found in Figure 8 (C). Following the procedure outlined above, after 81 reduction steps, we find a network with a structural deficiency of one. Using Theorem 4 we know that any two reactions not in a cycle are coupled. Since this network, its graph is shown in Figure 8 (B), has only three cycles, there are many of these reactions and a large amount of information can be extrapolated to the original network. As an example we find that in the reduced network, the complexes corresponding to

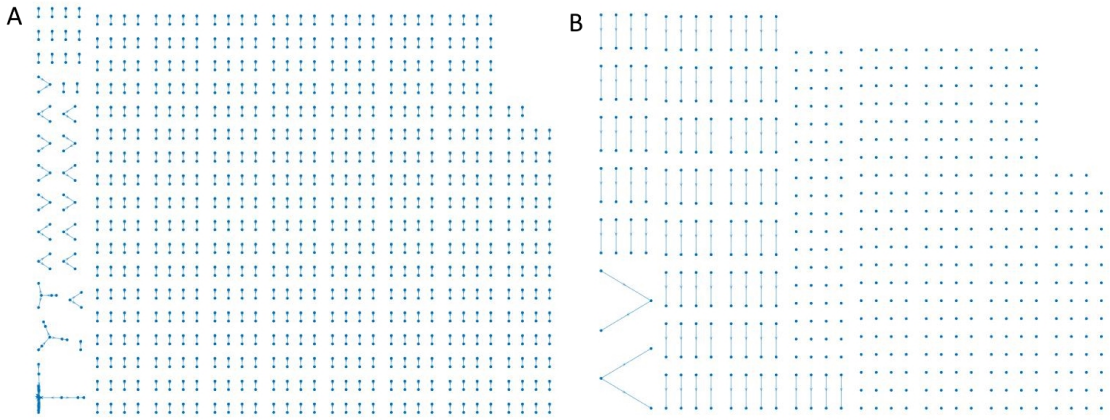

Figure 9: Left (A) shows the graph of the original Th. maritima network. Right (B) shows the graph of the reduced deficiency zero network.

the single complex sets $\{45\}$ and $\{993\}$ are substrates of reactions not in cycles. This means that in the original network, the corresponding reactions, which have complexes 45 and 993 will also be coupled. Since those complexes are different only in species 100 (7,8-Diaminononanoate), we have showed ACR in that species.

Thermotoga maritima [bigg.ucsd.edu/models/iLJ478]
The next network is a reconstruction of the central metabolic network of the bacterium Thermotoga maritima [42]. It has 650 species, 1330 complexes and 754 reactions with a structural deficiency of 142. Its graph is shown in Figure 9 (A). Using the reduction heuristic for 100 steps gives us a network with a structural deficiency of zero. The graph of this network is shown in Figure 9 (B). We can see that this network is not weakly reversible. In Proposition 3 we learned that if reduction yields a network of deficiency zero that is not weakly reversible, then the original network can have no positive steady states. This makes further analysis of this network with respect to robustness unnecessary. We made an attempt to find a reduced network of deficiency zero for the last network, however were not able to find one. It is possible that our approach of searching for reductions of a targeted deficiency is not exhaustive, however this result would suggests that the H. pylori network has positive steady states.

Escherichia coli [bigg.ucsd.edu/models/e_coli_core,/iJO1366]
We finally tested the approach two further networks, both genome-scale reconstructions of Escherichia coli. For the first [27], we found a reduced network with structural a deficiency of zero that is not weakly reversible, meaning that this network does not allow for positive steady states as in the Th. maritima network. The second [26], a genome-scale reconstruction of the metabolic network of Escherichia coli K-12 MG1655 has 1805 species and 2583 reaction. The current implementation of our algorithm applied to this network is too slow to provide meaningful results. This suggests that the approach in its current form will be limited to networks with a number of species in the hundreds.

References

[1] Kohei Adachi. Elementary matrix operations. In *Matrix-Based Introduction to Multivariate Data Analysis*, pages 3–16. Springer, 2016.

[2] Lilia Alberghina and Hans V Westerhoff. *Systems biology: definitions and perspectives*, volume 13. Springer Science & Business Media, 2007.

[3] Carlene Perpetua P Arceo, Editha C Jose, Angelyn R Lao, and Eduardo R Mendoza. Reaction networks and kinetics of biochemical systems. *Mathematical biosciences*, 283:13–29, 2017.

[4] Paul C Bressloff. *Stochastic processes in cell biology*, volume 41. Springer, 2014.

[5] Vijaysekhar Chellaboina, Sanjay P Bhat, Wassim M Haddad, and Dennis S Bernstein. Modeling and analysis of mass-action kinetics. *IEEE Control Systems*, 29(4), 2009.

[6] Athel Cornish-Bowden and Athel Cornish-Bowden. *Fundamentals of enzyme kinetics*, volume 510. Wiley-Blackwell Weinheim, Germany, 2012.

[7] Smadar Ben-Tabou de Leon and Eric H Davidson. Modeling the dynamics of transcriptional gene regulatory networks for animal development. *Developmental biology*, 325(2):317–328, 2009.

[8] Martin Feinberg. Complex balancing in general kinetic systems. *Archive for rational mechanics and analysis*, 49(3):187–194, 1972.

[9] Martin Feinberg. Chemical reaction network structure and the stability of complex isothermal reactors—ii. multiple steady states for networks of deficiency one. *Chemical Engineering Science*, 43(1):1–25, 1988.

[10] Martin Feinberg. The existence and uniqueness of steady states for a class of chemical reaction networks. *Archive for Rational Mechanics and Analysis*, 132(4):311–370, 1995.

[11] Robin E Ferner and Jeffrey K Aronson. Cato guldberg and peter waage, the history of the law of mass action, and its relevance to clinical pharmacology. *British journal of clinical pharmacology*, 81(1):52–55, 2016.

[12] Ronald Aylmer Fisher. *The genetical theory of natural selection: a complete variorum edition*. Oxford University Press, 1999.

[13] Jeremy Gunawardena. Chemical reaction network theory for in-silico biologists. *Notes available for download at http://vcp. med. harvard. edu/papers/crnt. pdf*, 2003.

[14] Daniel Sander Hoffmann. The dawn of mathematical biology. *arXiv preprint arXiv:1511.01455*, 2015.

[15] Neal S Holter, Amos Maritan, Marek Cieplak, Nina V Fedoroff, and Jayanth R Banavar. Dynamic modeling of gene expression data. *Proceedings of the National Academy of Sciences*, 98(4):1693–1698, 2001.

[16] Fritz Horn. Necessary and sufficient conditions for complex balancing in chemical kinetics. *Archive for Rational Mechanics and Analysis*, 49(3):172–186, 1972.

[17] Alston S Householder. Advances and applications of mathematical biology, 1941.

[18] Giorgio Israel. *Essays on the history of mathematical biology*. Giorgio Israel, 2013.

[19] Robert L Karp, Mercedes Pérez Millán, Tathagata Dasgupta, Alicia Dickenstein, and Jeremy Gunawardena. Complex-linear invariants of biochemical networks. *Journal of theoretical biology*, 311:130–138, 2012.

[20] Zachary A King, Justin Lu, Andreas Dräger, Philip Miller, Stephen Federowicz, Joshua A Lerman, Ali Ebrahim, Bernhard O Palsson, and Nathan E Lewis. Bigg models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic acids research*, 44(D1):D515–D522, 2015.

[21] Hiroaki Kitano. Systems biology: a brief overview. *Science*, 295(5560):1662–1664, 2002.

[22] Hiroaki Kitano. Biological robustness. *Nature Reviews Genetics*, 5(11):826, 2004.

[23] Hiroaki Kitano. Towards a theory of biological robustness. *Molecular systems biology*, 3(1):137, 2007.

[24] Fritz Müller. Ituna and thyridia: a remarkable case of mimicry in butterflies. *Trans. Entomol. Soc. Lond*, 1879:20–29, 1879.

[25] Denis Noble. *The music of life: biology beyond genes*. Oxford University Press, 2008.

[26] Jeffrey D Orth, Tom M Conrad, Jessica Na, Joshua A Lerman, Hojung Nam, Adam M Feist, and Bernhard Ø Palsson. A comprehensive genome-scale reconstruction of escherichia coli metabolism—2011. *Molecular systems biology*, 7(1):535, 2011.

[27] Jeffrey D Orth, Ronan MT Fleming, and Bernhard O Palsson. Reconstruction and use of microbial metabolic networks: the core escherichia coli metabolic model as an educational guide. *EcoSal plus*, 2010.

[28] Irwin H Segel. Enzyme kinetics: behavior and analysis of rapid equilibrium and steady state enzyme systems, 1975.

[29] Guy Shinar and Martin Feinberg. Structural sources of robustness in biochemical reaction networks. *Science*, 327(5971):1389–1391, 2010.

[30] Guy Shinar and Martin Feinberg. Design principles for robust biochemical reaction networks: what works, what cannot work, and what might almost work. *Mathematical biosciences*, 231(1):39–48, 2011.

[31] Ralf Steuer, Thilo Gross, Joachim Selbig, and Bernd Blasius. Structural kinetic modeling of metabolic networks. *Proceedings of the National Academy of Sciences*, 103(32):11868–11873, 2006.

[32] Ines Thiele, Thuy D Vo, Nathan D Price, and Bernhard Ø Palsson. Expanded metabolic reconstruction of helicobacter pylori (iit341 gsm/gpr): an in silico genome-scale characterization of single-and double-deletion mutants. *Journal of bacteriology*, 187(16):5818–5830, 2005.

[33] Darcy Wentworth Thompson et al. On growth and form. *On growth and form.*, 1942.

[34] Alan Mathison Turing. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 237(641):37–72, 1952.

[35] Julio Vera, Olaf Wolkenhauer, and Ulf Schmitz. Cancer systems biology: Current achievements in–omics data analysis, network reconstruction and mathematical modelling. *eLS*, 2014.

[36] Vito Volterra. Fluctuations in the abundance of a species considered mathematically, 1926.

[37] Ludwig Von Bertalanffy. General system theory. *New York*, 41973(1968):40, 1968.

[38] Marian Walhout, Marc Vidal, and Job Dekker. *Handbook of systems biology: concepts and insights*. Academic Press, 2012.

[39] Douglas Brent West et al. *Introduction to graph theory*, volume 2. Prentice hall Upper Saddle River, 2001.

[40] James Michael Whitacre. Biological robustness: paradigms, mechanisms, and systems principles. *Frontiers in genetics*, 3:67, 2012.

[41] B. J. Zeng. From positive to synthetic science. *Communication on Transgenic Animals*, 11, 1995.

[42] Ying Zhang, Ines Thiele, Dana Weekes, Zhanwen Li, Lukasz Jaroszewski, Krzysztof Ginalski, Ashley M Deacon, John Wooley, Scott A Lesley, Ian A Wilson, et al. Three-dimensional structural view of the central metabolic network of thermotoga maritima. *Science*, 325(5947):1544–1549, 2009.

A Proofs

Proof of Proposition 1 Again, what we want to show is that for $\Phi \in \mathbb{R}^{\mathcal{C}\times\mathcal{C}}$ defined such that

$$\Phi_{y,y'} = \begin{cases} k_{y\to y'} & \text{if } y \to y' \in \mathcal{R} \\ 0 & \text{otherwise} \end{cases}$$

we have

$$A_k = \Phi^T - \Delta(\Phi 1)$$

Proof. The species formation rate function can be written as

$$\begin{aligned} f(c) &= \sum_{y\to y'\in\mathcal{R}} k_{y\to y'}\psi_y(c)(y'-y) \\ &= \sum_{y\to y'\in\mathcal{R}} k_{y\to y'}\psi_y(c)(Y\omega_{y'} - Y\omega_y) \\ &= Y\sum_{y\in\mathcal{C}}\Big(\sum_{y'\in\mathcal{C}} k'_{y'\to y}\psi_{y'}(c) - \sum_{y'\in\mathcal{C}} k'_{y\to y'}\psi_y(c)\Big)\omega_y \end{aligned}$$

where

$$k'_{y\to y'} = \begin{cases} k_{y\to y'} & \text{if } y \to y' \in \mathcal{R} \\ 0 & \text{otherwise} \end{cases}$$

This means that for any $x \in \mathbb{R}^{\mathcal{C}}$ we must have

$$(A_k x)_y = \sum_{y'\in\mathcal{C}} k'_{y'\to y}x_{y'} - \sum_{y'\in\mathcal{C}} k'_{y\to y'}x_y$$

and so we have $A_k = \Phi^T - \Delta(\Phi 1)$ since

$$\begin{aligned} (\Phi^T x)_y &= \sum_{y'\in\mathcal{C}} \Phi_{y',y}x_{y'} \\ &= \sum_{y'\in\mathcal{C}} k'_{y'\to y}x_{y'} \end{aligned}$$

and

$$\begin{aligned} (\Delta(\Phi 1)x)_y &= \sum_{y'\in\mathcal{C}} \Delta(\Phi 1)_{y,y'}x_{y'} \\ &= (\Phi 1)_y x_y \\ &= (\sum_{y'\in\mathcal{C}} k'_{y\to y'})x_y \end{aligned}$$

□

Proof of Lemma 4 What we wanted to show was that in a network with

$$\dim(\ker(\Lambda)) = \gamma$$

there are at most γ independent cycles and any set of γ independent cycles

$$\Gamma = \{v_1, ..., v_\gamma\}$$

will be a basis of $\ker(\Lambda)$. We already showed that for any cycle v, we have $v \in \ker(\Lambda)$ and by definition

$$\dim(\text{span}(\Gamma)) = \gamma$$

so that the only thing that remains to show is that there are γ independent cycles in such a network and not more.

Proof. From Section 3.2.1 we know that $\text{rank}(\Lambda) = n - l$ so that

$$\dim(\ker(\Lambda)) = r - \text{rank}(\Lambda) = r - (n - l)$$

where r is the number of reactions, n the number of complexes and l the number of linkage classes. Assuming we have a network with

$$\dim(\ker(\Lambda)) = 0$$

then $r = n - l$. For any linkage class L_i, let n_i be the number of complexes in that class. Since the linkage classes partition the set of complexes we have $n = \sum_{i=1}^{l} n_i$. Furthermore the number of reactions needed to link n_i complexes is at least $n_i - 1$ meaning that the number of reactions will have to be at least

$$r \geq \sum_{i=1}^{l} n_i - 1 = n - l$$

Now let us consider what happens if there is a cycle in that network. Since after removing any reaction in that cycle from the network, all complexes in the cycle would still be linked, we have at least one reaction that is not needed to link the linkage classes and so the number of reactions in that network greater than $n_j - 1$ which means we get

$$r > \sum_{i=1}^{l} n_i - 1 = n - l$$

Since this is not possible, there can be no cycles in the network. Now we can argue by induction. A network with only a single reaction will necessarily have

$$\dim(\ker(\Lambda)) = 0$$

and no cycles. Any other network can be created by adding reactions to a network with fewer reactions. Let therefore N be a network with

$$\dim(\ker(\Lambda)) = \gamma$$

and assume $\Gamma = \{v_1, ..., v_\gamma\}$ is a basis of $\ker(\Lambda)$. Add a single reaction $r = y \to y'$ to the network and define $\omega_r = \omega_{y'} - \omega_y$. Then in the new network N_+ the incidence matrix is given by $\Lambda_+ = (\Lambda|\omega_r)$. This means for every $v \in \ker(\Lambda)$ we have

$$v_+ = (v|0) \in \ker(\Lambda_+)$$

Now there are two possibilities. First, assume that y and y' were in different linkage classes in N. Then $l_+ = l - 1$ and since $r_+ = r + 1$ and $n_+ = n$ we get

$$\begin{aligned} \dim(\ker(\Lambda_+)) &= r_+ - n_+ + l_+ \\ &= r - n + l \\ &= \dim(\ker(\Lambda)) \end{aligned}$$

which means Γ is a basis of $\ker(\Lambda)$. Furthermore since in N there is no chain of reactions from y' to y, the new network can have no cycle involving r and the claim follows. If instead y and y' were in the same linkage class in N, then arguing along the same lines we get

$$\dim(\ker(\Lambda_+)) = \dim(\ker(\Lambda)) + 1$$

Since there exists a chain of reactions $r_1, ..., r_k$ in N that link y and y', the new reactions will create a non-trivial cycle v_+. Since r is involved in that cycle, it cannot be in Γ and we get $\{v_+\} \cup \Gamma$ is a basis of Λ_+. Let v'_+ be any other cycle involving r, then

$$v''_+ = (v_+)_r v'_+ - (v'_+)_r v_+$$

has $(v''_+)_r = 0$ and therefore $v''_+ \in \text{span}(\Gamma)$ which means

$$v'_+ \in \text{span}(\{v_+\} \cup \Gamma)$$

which completes the proof. □

Proof of Lemma 5 What we wanted to show was that if we have two reactions r_1, r_2 such that any cycle v that involves the one also involves the other then there exists a set of independent cycles

$$\Gamma' = \{v'_1, ..., v'_\gamma\}$$

such that $(v'_i)_{r_1} = (v'_i)_{r_2} = 0$ for all $i > 1$.

Proof. Let Γ be a set of independent cycles

$$\Gamma = \{v_1, ..., v_\gamma\}$$

Assume there are is one or less cycle in Γ such that $(v_k)_{r_1} \neq 0$. Reorder Γ into Γ' such that $v'_1 = v_k$ and we are done since for any $i > 2$ we have a $j \neq k$ such that $v'_i = v_j$ and therefore

$$(v'_i)_{r_1} = 0 \Rightarrow (v'_i)_{r_2} = 0$$

Assuming there are more than one cycle in Γ such that $(v_i)_{r_1} \neq 0$. Without loss of generality let $(v_1)_{r_1} \neq 0$. For any cycles v_i, v_j with $(v_i)_{r_1} \neq 0$ and $(v_j)_{r_1} \neq 0$ we must have

$$(v_i)_{r_1}(v_j)_{r_2} = (v_j)_{r_1}(v_i)_{r_2}$$

since otherwise we can create a cycle

$$v = (v_j)_{r_2}(v_i) - (v_i)_{r_2}(v_j)$$

such that $v_{r_1} \neq 0$ and $v_{r_2} = 0$. Using this we can find a new set of independent cycles

$$\Gamma' = \{v'_1, ..., v'_\gamma\}$$

by taking $v'_1 := v_1$ and for $i > 1$ take

$$v'_i = (v'_1)_{r_1} v_i - (v_i)_{r_1} v'_1$$

With that we have

$$(v'_i)_{r_1} = (v'_1)_{r_1}(v_i)_{r_1} - (v_i)_{r_1}(v'_1)_{r_1} = 0$$

and

$$(v'_i)_{r_2} = (v'_1)_{r_1}(v_i)_{r_2} - (v_i)_{r_1}(v'_1)_{r_2} = 0$$

since either $(v_i)_{r_2} = (v_i)_{r_1} = 0$ or

$$\frac{(v_i)_{r_1}}{(v_i)_{r_2}} = \frac{(v'_1)_{r_1}}{(v'_1)_{r_2}}$$

In any case, Γ satisfies the requirement that $(v'_i)_{r_1} = (v'_i)_{r_2} = 0$ for $i > 1$. □

Proof of Theorem 4 The theorem had two parts which we will prove separately. The first was in connection to networks with a structural deficiency of zero and claimed that for any two reactions $y_1 \to y'_1$ and $y_2 \to y'_2$ as described in Lemma 4 will be coupled, meaning that there exists some constant θ such that for any positive steady state c^* we have

$$\frac{v_{y_1 \to y'_1}(c^*)}{v_{y_2 \to y'_2}(c^*)} = \theta$$

Proof. Let c^* be a positive steady state of N, then we have

$$Y \Lambda v(c^*) = 0$$

and since

$$\delta_s = \dim(\ker(Y) \cap \mathrm{Im}(\Lambda)) = 0$$

we must have

$$\Lambda v(c^*) = 0$$

This means for any basis $\Gamma = \{\upsilon_1, ..., \upsilon_\gamma\}$ of $\ker(\Lambda)$ we can find coefficients λ_i such that

$$v(c^*) = \sum_{i=1}^{\gamma} \lambda_i \upsilon_i$$

Now if we choose this basis as in Lemma 4, i.e. such that

$$(\upsilon_i)_{y_1 \to y_1'} = 0$$

and

$$(\upsilon_i)_{y_2 \to y_2'} = 0$$

for all $i > 1$ then we get

$$v_{y_1 \to y_1'}(c^*) = \lambda_1 (\upsilon_1)_{y_1 \to y_1'}$$

and

$$v_{y_2 \to y_2'}(c^*) = \lambda_1 (\upsilon_1)_{y_2 \to y_2'}$$

which means

$$\frac{v_{y_1 \to y_1'}(c^*)}{v_{y_2 \to y_2'}(c^*)} = \frac{(\upsilon_1)_{y_1 \to y_1'}}{(\upsilon_1)_{y_2 \to y_2'}}$$

Since the basis of Λ can be chosen independently of c^*, the claim follows. □

Let us therefore move on to $\delta_s = 1$. There we claimed that any two reactions $y_1 \to y_1'$ and $y_2 \to y_2'$ not involved in any cycles are coupled.

Proof. As before we assume c^* to be a positive steady state and $\Gamma = \{\upsilon_1, ..., \upsilon_\gamma\}$ to be a basis of $\ker(\Lambda)$. Since

$$v_{y \to y'}(c^*) = k_{y \to y'}(c^*)^y$$

if c^* is positive then so is $v(c^*)$. If now there exists a reaction $y \to y'$ that is not involved in any cycle then for $\upsilon \in \ker(\Lambda)$ we have

$$(\upsilon_i)_{y \to y'} = 0$$

so that $v(c^*) \notin \ker(\Lambda)$. Since

$$\delta_s = \dim(\ker(Y) \cap \mathrm{Im}(\Lambda)) = 1$$

this gives us that for any element $v \in \ker(Y\Lambda)$ we can find $\lambda_0, \lambda_1, ..., \lambda_\gamma$ such that

$$v = \lambda_0 v(c^*) + \sum_{i=1}^{\gamma} \lambda_i \upsilon_i$$

In particular let c^{**} be any positive steady state, then

$$v(c^{**}) = \lambda_0 v(c^*) + \sum_{i=1}^{\gamma} \lambda_i \upsilon_i$$

and for any reaction $y \to y'$ not involved in any cycle,

$$\begin{aligned} v_{y\to y'}(c^{**}) &= \lambda_0 v_{y\to y'}(c^*) + \sum_{i=1}^{\gamma} \lambda_i (v_i)_{y\to y'} \\ &= \lambda_0 v_{y\to y'}(c^*) \end{aligned}$$

or equivalently

$$\frac{v_{y\to y'}(c^{**})}{v_{y\to y'}(c^*)} = \lambda_0$$

which means that if we have two reactions $y_1 \to y_1'$ and $y_2 \to y_2'$ not involved in any cycles we get

$$\frac{v_{y_1\to y_1'}(c^{**})}{v_{y_1\to y_1'}(c^*)} = \frac{v_{y_2\to y_2'}(c^{**})}{v_{y_2\to y_2'}(c^*)}$$

or

$$\frac{v_{y_1\to y_1'}(c^{**})}{v_{y_2\to y_2'}(c^{**})} = \frac{v_{y_1\to y_1'}(c^*)}{v_{y_2\to y_2'}(c^*)}$$

which means that $\frac{v_{y_1\to y_1'}(c)}{v_{y_2\to y_2'}(c)}$ is constant across all steady states as claimed. □