Biochemical Reaction Networks: Network Structures and Dynamical Properties

By

Tung D. Nguyen

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

(Mathematics)

at the

UNIVERSITY OF WISCONSIN – MADISON

2021

Date of final oral examination: May 3, 2021

The dissertation is approved by the following members of the Final Oral Committee:

Professor David F. Anderson, Professor, Mathematics

Professor Gheorghe Craciun, Professor, Mathematics

Professor Benedek Valko, Professor, Mathematics

Professor Sebastien Roch, Professor, Mathematics

Abstract

Title: Biochemical Reaction Networks: Network Structures and Dynamical Properties

Reaction networks are commonly used to model a variety of physical systems ranging from the microscopic world like cell biology and chemistry, to the macroscopic world like epidemiology and evolution biology. At its core, a reaction network model consists of two components: the network component, and its associated dynamics. The dynamical systems associated with reaction networks usually come from one of two types: a deterministic model utilizing ordinary differential equations (ODEs) or a stochastic model utilizing continuous-time Markov chains. Not surprisingly, there is a strong connection between the network structure and the qualitative behavior of the associated dynamical system, both in the deterministic and stochastic modeling regimes.

A major question in the theory of reaction networks concerns this connection: given a reaction network model that has certain special structures, what are the qualitative properties of its dynamical systems? Regarding this question, two main contributions will be presented in this thesis. In the reaction network literature, many results established qualitative behavior of reaction networks under the common assumption that the dynamical system is governed by mass action kinetics. As many networks in practice do not follow the law of mass action, the first contribution relaxed this assumption and extended three existing results to the setting of non-mass action kinetics. The second contribution lies in the study of strongly endotactic networks. We utilized "tier structure"-an analytical tool to study the dynamical properties of strongly endotactic

networks in both the deterministic and stochastic models.

Another question arises naturally from a mathematical standpoint, and is also inspired by the recent emergence of network science and network biology: how prevalent or common are these special network structures? To address this question quantitatively for the structure deficiency zero, we first developed two random graph frameworks including an Erdős Rényi framework and a stochastic block model framework to generate random reaction networks. Under these two frameworks, we then studied the scaling limit of the probability that a random network has deficiency zero as the number of species goes to infinity.

Acknowledgements

First and foremost, I am deeply grateful to my thesis advisor, David F. Anderson for introducing me to the field of reaction networks where I discovered many interesting problems and met many wonderful people. I am thankful for his continuous support and encouragement that helped me overcome many difficulties I faced during my research. I am grateful for many valuable lessons on writing and presenting that I learned from him. I would like to also express my gratitude to Dr. Daniele Cappelletti, my friend and collaborator, for having many inspiring ideas and discussions that helped me solve my thesis problems and come up with future directions.

Thanks to Professor Gheorghe Craciun, Professor Benedek Valko and Professor Sebastien Roch for appearing in my thesis defense, and for their great classes that I attended during my time at University of Wisconsin-Madison.

I also would like to thank the academic and administrative staff of Department of Mathematics. Especially, thanks to Kathie Brohaugh for all the administrative helps, and thanks to Justin Sukiennik for helping me develop and polish my teaching skills.

Last but not least, I would like to thank my parents for always being understanding, encouraging and caring. I am forever grateful to them for their unconditional support throughout my life.

List of Figures

Contents

Chapter 1

Introduction

Reaction networks are used to model a variety of physical systems from microscopic processes such as chemical reactions and protein interactions, to macroscopic phenomena such as the spread of disease and the evolution of species. In reaction networks, the interacting agents (such as biochemical molecules, animal species, human populations) are referred to by a common term "species". These networks take the form of directed graphs in which the vertices, often termed *complexes* in the domains of interest, are linear combinations of the species over the non-negative integers and the directed edges are termed *reactions*. See Figure 1 for an example of a reaction network.

At its core, a reaction network model consists of two components: the network (or graph) component, and the dynamics under such a graph. The dynamical systems associated with reaction networks usually come from one of two types. When the abundances of the constituent species in a system are low, randomness plays an important role in the interaction between species, and thus the abundances are modeled stochastically as a continuous-time Markov chain. However, when the abundances of the species are high, such randomness is averaged out and the concentrations are instead modeled deterministically by a system of ordinary differential equations (ODEs). Not surprisingly, in both models there is a strong connection between the network structure and the qualitative behavior of the dynamical system. Certain network structures such as *deficiency zero*

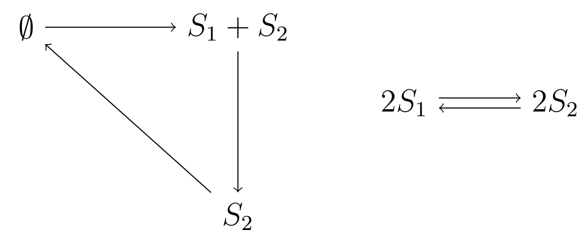

Figure 1: A reaction network with two species: S_1 and S_2. The vertices are linear combinations of the species over the integers, and are termed complexes. The directed edges are termed reactions and determine the net change in the counts of the species due to one instance of the reaction.

and *strong endotacticity* (which will be detailed in later sections) ensure many desirable behaviors of the dynamical systems including existence and stability of equilibria.

Focusing on this, my work revolves around two objectives:

1. Establish the dynamical properties of reaction systems whose associated networks have certain structures.

2. Examine how prevalent these structures are among random reaction networks.

Specifically, my work in [5] and [13] addressed the first objective. In the reaction network literature, many results (see [1, 2, 6, 7, 8, 16, 18]) established qualitative behavior of reaction networks under the common assumption that the dynamical system is governed by mass action kinetics. As many networks in practice do not follow the law of mass action, in [13] we relaxed this assumption and extended three existing results [6, 7, 16] to the setting of non-mass action kinetics. In [5], we utilized an analytical tool to study the dynamical properties of strongly endotactic networks in both the deterministic and stochastic models. My main contribution in this project was proving that strongly endotactic networks satisfy a Lyapunov-like condition that leads to a Large Deviation Principle (LDP).

The second objective not only arises naturally from a mathematical standpoint, but it is also inspired by the recent emergence of network science and network biology, where many important network features are examined via randomized networks. In [12], we developed an Erdős-Rényi framework to generate random reaction networks, and utilized it to understand the prevalence of deficiency zero-the network structure most central to reaction network theory. With this as a starting point, in [14] we considered a stochastic block model framework which can be adapted to study deficiency zero in different settings where reaction networks may have vastly different structures.

The remainder of this thesis is organized in chronological order of my work. In Chapter 2, I will provide the necessary background on reaction networks including key definitions, the associated dynamical systems, and several relevant classical results. In Chapter 3, which corresponds to my work in [13], I will start with established results on complex balanced reaction networks with mass action kinetics, then extend these results to the setting of non-mass action kinetics. In Chapter 4, which follows my work in [5], I will formally introduce strongly endotactic network, then explain the main analytic tool-tier structures and use it to analyze strongly endotactic networks. In Chapter 5, which follows my work in [12] and [14], I will set up two frameworks: Erdős-Rényi and stochastic block model to generate random reaction networks, and use them to quantify the prevalence of deficiency zero structure among random networks. Finally, the Appendix includes several technical lemmas needed for the proofs of various theorems presented in the thesis.

Chapter 2

Background on reaction networks

2.1 Reaction networks and key definitions

Let $\{S_1, \dots, S_n\}$ be a set of n *species* undergoing a finite number of reaction types. We denote a particular reaction by $y \to y'$, where y and y' are linear combinations of the species on $\{0, 1, 2, \dots\}$ representing the number of molecules of each species consumed and created in one instance of that reaction, respectively. The linear combinations y and y' are often called *complexes* of the system. For a given reaction, $y \to y'$, the complex y is called the *source complex* and y' is called the *product complex*. A complex can be both a source complex and a product complex. We may associate each complex with a vector in $\mathbb{Z}^n_{\geq 0}$, whose coordinates give the number of molecules of the corresponding species in the complex. As is common in the reaction network literature, both ways of representing complexes will be used interchangeably throughout the paper. For example, if the system has 2 species $\{S_1, S_2\}$, the reaction $S_1 + S_2 \to 2S_2$ has $y = S_1 + S_2$, which is associated with the vector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$, and $y' = 2S_2$, which is associated with the vector $\begin{bmatrix} 0 \\ 2 \end{bmatrix}$. Viewing the complexes as vectors, the *reaction vector* associated to the reaction $y \to y'$ is simply $y' - y \in \mathbb{Z}^n$, which gives the state update of the system due to one occurrence of the reaction.

Definition 2.1. *For $n \geq 0$, let $\mathcal{S} = \{S_1, ..., S_n\}$, $\mathcal{C} = \cup_{y \to y'} \{y, y'\}$, and $\mathcal{R} = \cup_{y \to y'} \{y \to y'\}$ be the sets of species, complexes, and reactions respectively. The triple $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ is called a reaction network. When $n = 0$, in which case $\mathcal{S} = \mathcal{C} = \mathcal{R} = \emptyset$, the network is termed the empty network.* $\triangle$

Remark 2.2. *It is common to assume, and we shall do so throughout, that each species of a given reaction network appears with a positive coefficient in at least one complex, and each complex takes part in at least one reaction (as either a source or a product complex). Thus, a reaction network $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ is fully specified if we know $\mathcal{R}$. In this case, we call $\mathcal{S}$ and $\mathcal{C}$ the set of species and the set of complexes* associated with $\mathcal{R}$.

To each reaction network $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$, there is a unique directed graph constructed in the obvious manner: the vertices of the graph are given by $\mathcal{C}$ and a directed edge is placed from y to y' if and only if $y \to y' \in \mathcal{R}$. Each connected component of the graph is called a *linkage class*. We denote by ℓ the number of linkage class. Note that by definition the directed graph associated to a reaction network contains only vertices corresponding to elements in $\mathcal{C}$ involved in some reaction, i.e., the degree of all vertices is at least 1 and so isolated vertices are not present in the associated network.

Remark 2.3. *Note that since each linkage class must consist of at least two complexes, we have the bound $\ell \leq \frac{|\mathcal{C}|}{2}$.*

Definition 2.4. *A reaction network $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ is called weakly reversible if each connected component of the associated directed graph is strongly connected.*

Definition 2.5. *The linear subspace $S = span\{y' - y\}$ generated by all reaction vectors is called the stoichiometric subspace of the network. For $c \in \mathbb{R}^n_{\geq 0}$ we say $c + S = \{x \in$*

$\mathbb{R}^n | x = c + s$ for some $s \in S\}$ is a stoichiometric compatibility class, $(c + S) \cap \mathbb{R}^n_{\geq 0}$ is a non-negative stoichiometric compatibility class, and $(c + S) \cap \mathbb{R}^m_{>0}$ is a positive stoichiometric compatibility class. Denote $dim(S) = s$.

Definition 2.6. *A vertex, $y \in \mathbb{Z}^n_{\geq 0}$, is called binary if $\sum_{i=1}^n y_i = 2$. A vertex is called unary if $\sum_{i=1}^n y_i = 1$. The vertex $\vec{0} \in \mathbb{Z}^n$ is said to be of zeroth order.*

Definition 2.7. *A reaction network $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ is called binary if each vertex is binary, unary, or of zeroth order.*

The following type of network will play a key role in Chapter 5.

Definition 2.8. *A reaction network is called paired if each of its connected components contains precisely two vertices. A reaction network is called i-paired if it is paired and contains i connected components.*

2.2 Dynamical systems

The dynamical systems associated with reaction networks usually come from one of two types. When the abundances of the constituent species in a system are low, randomness plays an important role in the interaction between species, and thus the abundances are modeled stochastically as a continuous-time Markov chain. However, when the abundances of the species are high, such randomness is averaged out and the concentrations are instead modelled deterministically by a system of ordinary differential equations (ODEs).

2.2.1 Stochastic model

The most common stochastic model for a reaction network $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ treats the system as a continuous time Markov chain whose state at time t, $X(t) \in \mathbb{Z}^n_{\geq 0}$, is a vector giving the number of molecules of each species present with each reaction modeled as a possible transition for the chain. The model for the reaction $y \to y'$ is determined by the source and product complexes of the reaction, and a function $\lambda_{y\to y'}$ of the state that gives the transition intensity, or rate, at which the reaction occurs. In the biological and chemical literature, transition intensities are referred to as propensities.

Given that the reaction $y \to y'$ happens at time t, the state is updated by the addition of the reaction vector $y' - y$,

$$X(t) = X(t-) + y' - y.$$

A common choice for the intensity functions $\lambda_{y\to y'}$ is to assume the system satisfies the stochastic version of *mass action kinetics*. In this case, the functions have the form

$$\lambda_{y\to y'}(x) = \kappa_{y\to y'} \prod_{i=1}^{n} \frac{x_i!}{(x_i - y_i)!} 1_{\{x_i \geq y_i\}} \tag{2.1}$$

where $\kappa_{y\to y'} > 0$ is called the *rate constant*. Under the assumption of mass action kinetics and a non-negative initial condition, it follows that the dynamics of the system is confined to the particular non-negative stoichiometric compatibility class determined by the initial value $X(0)$, namely $X(t) \in (X(0) + S) \cap \mathbb{R}^n_{\geq 0}$.

Simple book-keeping implies that $X(t)$ satisfies

$$X(t) = X(0) + \sum_{y\to y' \in \mathcal{R}} R_{y\to y'}(t)(y' - y),$$

where $R_{y\to y'}(t)$ gives the number of times reaction $y \to y'$ has occurred by time t. Kurtz showed that X can be represented as the solution to the stochastic equation

$$X(t) = X(0) + \sum_{y \to y' \in \mathcal{R}} Y_{y \to y'} \left(\int_0^t \lambda_{y \to y'}(X(s))ds \right) (y' - y), \tag{2.2}$$

where the $Y_{y \to y'}$ are independent unit-rate Poisson process [32].

Another way to characterize the models of interest is via Kolmogorov's forward equation, termed the chemical master equation in the biology and chemistry literature, which describes how the distribution of the process changes in time. Letting $p_\mu(x,t)$ give the probability that $X(t) = x$ assuming an initial distribution of μ, the forward equation is

$$\frac{d}{dt} p_\mu(x,t) = \sum_{y \to y' \in \mathcal{R}} \lambda_{y \to y'}(x - (y' - y)) p_\mu(x - (y' - y), t) - \sum_{y \to y' \in \mathcal{R}} \lambda_{y \to y'}(x) p_\mu(x,t).$$

Constant solutions to the forward equation, i.e. those satisfying

$$\sum_{y \to y' \in \mathcal{R}} \pi(x - y' + y) \lambda_{y \to y'}(x - y' + y) = \pi(x) \sum_{y \to y' \in \mathcal{R}} \lambda_{y \to y'}(x)$$

are stationary measures for the process, and if they are summable they can be normalized to give a stationary distribution. Assuming the associated stochastic model is non-explosive, stationary distributions characterize the long-time behavior of the stochastically modeled system.

2.2.2 Deterministic model

Under the classical scaling (for more details, see [10, 11, 31]) the continuous time Markov chain model of the previous section becomes

$$x(t) = x(0) + \sum_{y \to y' \in \mathcal{R}} \left(\int_0^t f_{y \to y'}(x(s))ds \right) (y' - y) \tag{2.3}$$

where

$$f_{y \to y'}(x) = \kappa_{y \to y'} x^y \tag{2.4}$$

where $\kappa_{y\to y'} > 0$ is the rate constant, and where for two vectors $u, v \in \mathbb{R}^m_{\geq 0}$ we denote $u^v = \prod_i u_i^{v_i}$ with the convention $0^0 = 1$. Later, we will also utilize the notation uv for the vector whose ith component is $u_i v_i$.

We say that the deterministic system (2.3) has *deterministic mass action kinetics* if the rate functions $f_{y\to y'}$ have the form (2.4). The system 2.3 is equivalent to the system of ODEs

$$\dot{x} = \sum_{y\to y'\in\mathcal{R}} \kappa_{y\to y'} x^y (y' - y). \tag{2.5}$$

The trajectory with initial condition x_0 is confined to the non-negative stoichiometric compatibility class $(x_0 + S) \cap \mathbb{R}^n_{\geq 0}$.

2.3 Special network structures and classical results

2.3.1 Complex balance reaction networks

Some mass action systems have complex balanced equilibria [26, 28], which has been shown to play an important role in many biological mechanisms [17, 23, 29, 34]. An equilibrium point c is said to be complex balanced if for all $z \in \mathcal{C}$, we have

$$\sum_{y\to y'\in\mathcal{R}:y=z} \kappa_{y\to y'} c^y = \sum_{y\to y'\in\mathcal{R}:y'=z} \kappa_{y\to y'} c^y, \tag{2.6}$$

where the sum on the left is over reactions for which z is the product complex and the sum on the right is over reactions for which z is the source complex.

In [28] it was shown that if there exists a complex balanced equilibrium $c \in \mathbb{R}^n_{\geq 0}$ then

1. There is one, and only one, positive equilibrium point in each positive stoichiometric compatibility class.

2. Each such equilibrium point is complex balanced.

3. Each such complex balanced equilibrium point is locally asymptotically stable relative to its stoichiometric compatibility class.

In [19], a proof is presented showing global stability relative to the stoichiometric compatibility class. Because of the above, we say that a system is complex balanced if it admits a complex balanced equilibrium.

Complex balanced systems are also of interest in the stochastic setting. In particular, the following theorem in [8] provides an explicit form for the stationary distribution of complex balanced systems.

Theorem 2.9. *Let $\{\mathcal{S}, \mathcal{R}, \mathcal{C}\}$ be a reaction network. Suppose that when modeled deterministically with mass action kinetics and rate constants $\{\kappa_{y\to y'}\}$ the system is complex balanced with a complex balanced equilibrium $c \in \mathbb{R}^n_{\geq 0}$. Then the stochastically modeled system with intensities (2.1), with the same rate constants $\{\kappa_{y\to y'}\}$, admits the stationary distribution*

$$\pi(x) = \prod_{i=1}^{n} \frac{c_i^{x_i}}{x_i!} e^{-c_i}, \quad x \in \mathbb{Z}^n_{\geq 0}. \tag{2.7}$$

See also [6], which shows that these systems are non-explosive, implying π yields the limiting distributions of the process. This stationary distribution will become the starting point and the main inspiration for results in Chapter 3.

While complex balanced networks have many interesting properties, in practice it is not an easy task to check if a network is complex balanced. Fortunately, there are classical results in the field going all the way back to the seminal works of Horn, Jackson, and Feinberg in 1972 [21, 26, 28] that give a condition to ensure a reaction network is complex balanced for all choices of rate constants. The condition is based on a network quantity termed the *deficiency*, which can be easily computed from each reaction network.

Theorem 2.10. *If the reaction system is weakly reversible and has a deficiency of zero, then for any choice of rate constants* $\{\kappa_{y\to y'}\}$ *the deterministically modeled system with mass action kinetics is complex balanced.*

The network structure *deficiency zero* will be the main focus for Chapter 5, and the next section will provide more information on deficiency of a reaction network.

2.3.2 Deficiency of a reaction network

We start the section with a formal definition on deficiency of a reaction network.

Definition 2.11. *The deficiency of a chemical reaction network* $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ *is* $\delta = |\mathcal{C}| - \ell - s$*, where* $|\mathcal{C}|$ *is the number of complexes,* ℓ *is the number of linkage classes, and* s *is the dimension of the stoichiometric subspace of the network.*

For each $j \leq \ell$*, we let* $\mathcal{C}_j$ *denote the collection of complexes in the* j*th linkage class,* s_j *be the corresponding dimension of the span of the reaction vectors of that component, and define* $\delta_j = |\mathcal{C}_j| - 1 - s_j$ *to be the deficiency of that component.*

Remark 2.12. *From the definition of deficiency, the empty network has deficiency zero.*

We collect a number of basic properties of deficiency in the following lemma.

Lemma 2.13. *Let* $n \geq 1$ *and let* $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ *be a reaction network with* n *species.*

(a) δ *does not depend upon the direction of the edges.*

(b) $s_j \leq |\mathcal{C}_j| - 1$*, and so* $\delta_j \geq 0$.

(c) $s \leq |\mathcal{C}| - \ell$, *and so* $\delta \geq 0$.

(d) $\delta = 0$ *if and only if both the following conditions hold:*

(i) $s_j = |\mathcal{C}_j| - 1$ *for each* $j \leq \ell$ *(equivalently,* $\delta_j = 0$ *for each* $j \leq \ell$*).*

(ii) $\sum_{j=1}^{\ell} s_j = s$.

(e) *If* $\delta = 0$, *then*

$$|\mathcal{C}| \leq 2n.$$

(f) *Suppose the reaction network is paired, and that* ζ_j *is a reaction vector from the* j*th connected component. Then* $\delta = 0$ *if and only if* $\{\zeta_j\}$ *are linearly independent.*

(g) *(Monotonicity of deficiency.) Let* $\{\widehat{\mathcal{S}}, \widehat{\mathcal{C}}, \widehat{\mathcal{R}}\}$ *and* $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ *be two reaction networks with* $\widehat{\mathcal{R}} \setminus \mathcal{R} = \{y \to y'\}$, *a single reaction. Let* $\hat{\delta}$ *and* δ *be the deficiencies of the two networks. Then*

$$\hat{\delta} \geq \delta.$$

(h) *Suppose the complexes of* $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ *are either unary or of zeroth order, then* $\delta = 0$.

(i) *Let* $\tilde{\mathcal{R}}$ *be a subset of* $\mathcal{R}$ *in which precisely one reaction of each reversible pair is removed. If* $\tilde{\mathcal{R}}$ *consists of linearly independent reaction vectors, then* $\delta = 0$.

Proof. (a) This follows from the definition of deficiency.

(b) This follows from the observation that a cycle within a connected component implies a dependency among the reaction vectors.

(c) This follows from (b) since $\mathcal{C} = \cup_{j=1}^{\ell} \mathcal{C}_j$ and $s \leq \sum_{j=1}^{\ell} s_j$.

(d) This follows in a straightforward manner from (b) and (c).

(e) From the definition of deficiency $\delta = |\mathcal{C}| - \ell - s$, the fact that $s \leq n$, and $\ell \leq \frac{|\mathcal{C}|}{2}$ (from Remark 2.3), we have

$$\delta \geq |\mathcal{C}| - \frac{|\mathcal{C}|}{2} - n = \frac{|\mathcal{C}|}{2} - n.$$

Since the reaction network has deficiency zero, we therefore have

$$0 \geq \frac{|\mathcal{C}|}{2} - n, \tag{2.8}$$

which implies $|\mathcal{C}| \leq 2n$.

(f) Since the reaction network is paired, we have $s_j = 1$ and $|\mathcal{C}_j| = 2$ for each $j \leq \ell$. Thus condition (i) in (d) is satisfied. Since $s_j = 1$, condition (ii) in (d) holds if and only if all ζ_j are linearly independent.

(g) Let ℓ, s and $\hat{\ell}, \hat{s}$ be the number of connected components and dimension of the stoichiometric subspace of $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ and $\{\widehat{\mathcal{S}}, \widehat{\mathcal{C}}, \widehat{\mathcal{R}}\}$, respectively.

- Case 1: $y, y' \in \mathcal{C}$ and y and y' are from the same connected component. In this case, we have $|\widehat{\mathcal{C}}| = |\mathcal{C}|$ and $\hat{\ell} = \ell$. Since y and y' are from the same connected component, the reaction vector $y' - y$ can be written as the linear combination of the remaining reaction vectors from its connected component. Therefore adding $y \to y'$ to $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ does not increase the dimension of its stoichiometric subspace. Thus $\hat{s} = s$ and $\hat{\delta} = \delta$.

- Case 2: $y, y' \in \mathcal{C}$ and y and y' are from different connected components. In this case, we have $|\widehat{\mathcal{C}}| = |\mathcal{C}|$ and $\hat{\ell} = \ell - 1$. Since we are adding one reaction to $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ to obtain $\{\widehat{\mathcal{S}}, \widehat{\mathcal{C}}, \widehat{\mathcal{R}}\}$, we add at most 1 dimension to the stoichiometric subspace of $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$. Thus $\hat{s} \leq s + 1$ and

$$\hat{\delta} = |\widehat{\mathcal{C}}| - \hat{\ell} - \hat{s} \geq |\mathcal{C}| - (\ell - 1) - (s + 1) = \delta.$$

- Case 3: $y \in \mathcal{C}$ and $y' \notin \mathcal{C}$ or vice versa. In this case, we have $|\widehat{\mathcal{C}}| = |\mathcal{C}| + 1$, and $\hat{\ell} = \ell$. Similar to the previous case, we must have $\hat{s} \leq s + 1$, and thus

$$\hat{\delta} = |\widehat{\mathcal{C}}| - \hat{\ell} - \hat{s} \geq |\mathcal{C}| + 1 - \ell - (s + 1) = \delta.$$

- Case 4: $y, y' \notin \mathcal{C}$. In this case, we have $|\widehat{\mathcal{C}}| = |\mathcal{C}| + 2$, and $\hat{\ell} = \ell + 1$. Similar to the previous cases, we still have $\hat{s} \leq s + 1$ and thus

$$\hat{\delta} = |\widehat{\mathcal{C}}| - \hat{\ell} - \hat{s} \geq |\mathcal{C}| + 2 - (\ell + 1) - (s + 1) = \delta. \qquad \square$$

(h) The proof of this part is similar to the proof of Lemma 2.13(g), and thus it is omitted for the sake of brevity. The result in this part is well-known.

(i) Again, the proof of this part is similar to the proof of Lemma 2.13(g), and thus it is omitted for the sake of brevity.

Definition 2.14. *Let $R = \{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ be a reaction network, and $\tilde{\mathcal{R}} \subset \mathcal{R}$. Then we denote by $\pi_{\tilde{\mathcal{R}}}(R)$ the reaction network whose set of reactions is $\tilde{\mathcal{R}}$, and whose species and vertices are the subsets of $\mathcal{S}$ and $\mathcal{C}$ that are associated with $\tilde{\mathcal{R}}$, according to Remark 2.2.*

Note that in Definition 2.14, $\pi_{\tilde{\mathcal{R}}}(R)$ can be thought of as a "sub-network", or a projection of R onto the subset of species, vertices, and reactions associated with $\tilde{\mathcal{R}}$. The following corollary is a direct consequence of Lemma 2.13(g).

Corollary 2.15. *Let $R = \{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ be a reaction network, and $\tilde{\mathcal{R}} \subset \mathcal{R}$. Then*

$$\delta_{\pi_{\tilde{\mathcal{R}}}(R)} \leq \delta_R.$$

In particular, if $\pi_{\tilde{\mathcal{R}}}(R)$ has a positive deficiency, then R also has a positive deficiency.

To end this section, we will illustrate the concept of deficiency via two examples.

Example 1 (Enzyme kinetics [8])**.** Consider a reaction network with species $\{S, E, SE, P\}$ and associated graph

$$S + E \leftrightarrows SE \leftrightarrows P + E$$

$$E \leftrightarrows \emptyset \leftrightarrows S.$$

In this example, the reaction network has $|\mathcal{C}| = 6$ vertices, there are $\ell = 2$ connected components, and the dimension of the stochiometric subspace is $s = 4$. Thus the deficiency is

$$\delta = 6 - 2 - 4 = 0.$$

The following example demonstrates that it is sometimes most natural to use Lemma 2.13(f) to verify that a network has a deficiency of zero.

Example 2 (Binary, 3-paired)**.** Consider a reaction network with species $\{S_1, S_2, \ldots, S_9\}$ and associated graph

$$S_1 + S_2 \rightleftarrows S_3 + S_4$$

$$S_1 + S_3 \rightleftarrows S_5 + S_6$$

$$S_6 + S_7 \rightleftarrows S_8 + S_9.$$

This network is *paired* in the sense of Definition 2.8. Moreover, there is linear independence among the connected components, which can be seen easily since each connected component has a species not found in any other connected component. Hence, Lemma 2.13(f) implies that the deficiency of this network is zero.

Chapter 3

Complex balanced reaction networks with non-mass action kinetics

Recall from Chapter 2 that if a network is complex balanced when modeled deterministically, the associated stochastic model under mass action kinetics admits a stationary distribution which is a product of Poissons. Many subsequent work follows from this result [6, 7, 16]. In [16], the converse was shown to be true: if the stationary distribution of a stochastically modeled network is given by a product of Poissons, then the network is complex balanced. In [6] it was shown that complex balanced networks are non-explosive. Lastly, in [7] it was shown that the limit of this stationary distribution under classical scaling (for more detail on the scaling see [10, 11, 31]) is a well known Lyapunov function.

Motivated by the fact that not all networks in practice follow the law of mass action, in this section we considered a more generalized setting in which the kinetics are not necessarily mass action. Specifically, we assumed the stochastic intensity function is of the form

$$\lambda_{y\to y'}(x) = \kappa_{y\to y'} \prod_{i=1}^{n} \theta_i(x_i)\theta_i(x_i-1)\dots\theta_i(x_i-y_i+1)$$

where $\theta_i : \mathbb{Z} \to \mathbb{R}_{\geq 0}$. This is a more generalized version of mass action kinetics (2.1), since we recover (2.1) with θ_i all being identity. It was proven in [8] that the stochastic model under this kinetics admits a product form stationary measure, albeit not of Poissons.

In this section, we will extend the three aforementioned results in [6, 7, 16] to stochastically modelled networks under this non-mass action kinetics. Notably, the stationary distribution in this setting does not converge under the classical scaling. Thus, we will construct a modified scaling under which the stationary distribution converges to a Lyapunov function of a related ODE system.

3.1 Stationary distribution of complex balanced reaction networks with mass action kinetics and related results

We start with the main theorem in [8], which provides an explicit form for the stationary distribution of complex balanced systems.

Theorem 3.1. *Let $\{\mathcal{S}, \mathcal{R}, \mathcal{C}\}$ be a reaction network. Suppose that when modeled deterministically with mass action kinetics and rate constants $\{\kappa_{y\to y'}\}$ the system is complex balanced with a complex balanced equilibrium $c \in \mathbb{R}^n_{\geq 0}$. Then the stochastically modeled system with intensities (2.1), with the same rate constants $\{\kappa_{y\to y'}\}$, admits the stationary distribution*

$$\pi(x) = \prod_{i=1}^{n} \frac{c_i^{x_i}}{x_i!} e^{-c_i}, \quad x \in \mathbb{Z}^n_{\geq 0}. \tag{3.1}$$

See also [6], which shows that these systems are non-explosive, implying π yields the limiting distributions of the process.

In the case when the stochastic model does not have mass action kinetics, [8] also provides an extended result. In particular, [8] considers generalized intensity functions as mentioned in several past papers [30, 35],

$$\lambda_{y\to y'}(x) = \kappa_{y\to y'} \prod_{i=1}^{n} \theta_i(x_i)\cdots\theta_i(x_i - y_i + 1), \tag{3.2}$$

where $\kappa_{y\to y'}$ are positive rate constants, $\theta_i : \mathbb{Z} \to \mathbb{R}_{\geq 0}$, and $\theta_i(x) = 0$ if $x \leq 0$. The functions θ_i should be thought of as the "rate of association" of the i^{th} species [30]. For a system with intensity functions (3.2), the product form stationary distribution is quite similar to the one in Theorem 2.9.

Theorem 3.2. *Let $\{\mathcal{S}, \mathcal{R}, \mathcal{C}\}$ be a reaction network. Suppose that when modeled deterministically with mass action kinetics and rate constants $\{\kappa_{y\to y'}\}$ the system is complex balanced with a complex balanced equilibrium $c \in \mathbb{R}^n_{\geq 0}$. Then the stochastically modeled system with general intensity functions (3.2), with the same rate constants $\{\kappa_{y\to y'}\}$, admits the stationary measure*

$$\pi(x) = \prod_{i=1}^{n} \frac{c_i^{x_i}}{\theta_i(1)\cdots\theta_i(x_i)}, \quad x \in \mathbb{Z}^n_{\geq 0}. \tag{3.3}$$

In the next sections, we will show that π in (3.3) is summable under some mild growth condition on θ_i, and thus it can be normalized to a stationary distribution.

Interestingly, it has been proven in [16] that the converse is also true.

Theorem 3.3. *Let $\{\mathcal{S}, \mathcal{R}, \mathcal{C}\}$ be a reaction network and consider the stochastically modeled system with rate constants $\{\kappa_{y\to y'}\}$ and mass action kinetics (2.1). Suppose that for some $c \in \mathbb{R}^n_{\geq 0}$ the stationary distribution for the stochastic model is (2.7). Then c is a complex balanced equilibrium for the associated deterministic model with mass action kinetics and rate constants $\{\kappa_{y\to y'}\}$.*

Another follow-up result comes from the scaling behavior of the stationary distribution for complex balanced system. We first provide a key definition.

Definition 3.4. *Let π be a probability distribution on a countable set Γ such that $\pi(x) > 0$ for all $x \in \Gamma$. The non-equilibrium potential of the distribution π is the function $\phi_\pi : \Gamma \to \mathbb{R}$, defined by*

$$\phi_\pi(x) = -\ln(\pi(x)).$$

In [7] it was shown that under an appropriate scaling, the limit of the non-equilibrium potential of the stationary distribution of a complex balanced system converges to a certain well-known Lyapunov function.

Definition 3.5. *Let $E \subset \mathbb{R}^n_{\geq 0}$ be an open subset of $\mathbb{R}^n_{\geq 0}$ and let $f : \mathbb{R}^n_{\geq 0} \to \mathbb{R}$. A function $\mathcal{V} : E \to \mathbb{R}$ is called a Lyapunov function for the system $\dot{x} = f(x)$ at $x_0 \in E$ if x_0 is an equilibrium point for f, that is $f(x_0) = 0$, and*

1. *$\mathcal{V}(x) > 0$ for all $x \neq x_0, x \in E$ and $\mathcal{V}(x_0) = 0$.*

2. *$\nabla\mathcal{V}(x) \cdot f(x) \leq 0$, for all $x \in E$, with equality if and only if $x = x_0$, where $\nabla\mathcal{V}$ denotes the gradient of $\mathcal{V}$.*

In particular, the non-equilibrium potential of the stationary distribution in (2.7) converges to the usual Lyapunov function of Chemical Reaction Network Theory

$$\mathcal{V}(x) = \sum_{i=1}^{n} x_i(\ln(x_i) - \ln(c_i) - 1) + c_i. \tag{3.4}$$

Next, we briefly discuss the scaling in which the convergence happens. It is called the classical scaling in the literature. For more detailed discussions, see [10, 11, 31].

Let $|y| = \sum_i y_i$ and let V be the volume of the system times Avogadro's number. Suppose $\{\kappa_{y\to y'}\}$ are the rate constants for the stochastic model. We defined the scaled rate constants as follows

$$\kappa^V_{y\to y'} = \frac{\kappa_{y\to y'}}{V^{|y|-1}} \tag{3.5}$$

and denote the scaled intensity function for the stochastic model by

$$\lambda^V_{y\to y'}(x) = \frac{\kappa_{y\to y'}}{V^{|y|-1}} \prod_{i=1}^n \frac{x_i!}{(x_i - y_i)!}. \tag{3.6}$$

Note that if $x \in \mathbb{Z}^n_{\geq 0}$ gives the counts of the different species, then $\tilde{x} := V^{-1}x$ gives the concentrations in moles per unit volume. Then, by standard arguments

$$\lambda^V_{y\to y'}(x) \approx V\kappa_{y\to y'} \prod_{i=1}^n \tilde{x}_i^{y_i} = V\lambda_{y\to y'}(\tilde{x})$$

where the final equality defines $\lambda_{y\to y'}$ and justifies the definition of deterministic mass action kinetics.

Denote the stochastic process determining the counts by $X^V(t)$, then normalizing the original process X^V by V and defining $\bar{X}^V := \frac{X^V}{V}$ gives us

$$\bar{X}^V(t) \approx \bar{X}^V(0) + \sum_{y\to y'\in\mathcal{R}} \frac{1}{V} Y_{y\to y'}\left(V\int_0^t \lambda_{y\to y'}(\bar{X}^V(s))ds\right)(y' - y),$$

where we are utilizing the representation (2.2). Since the law of large numbers for the Poisson process implies $V^{-1}Y(Vu) \approx u$, we may conclude that a good approximation to the process $\bar{X}^V$ is the function $x = x(t)$ defined as the solution to the ODE

$$\dot{x} = \sum_{y\to y'\in\mathcal{R}} \kappa_{y\to y'} x^y (y' - y),$$

which is exactly (2.5).

A corollary of Theorem 2.9 gives us the stationary distribution for the classically scaled system.

Theorem 3.6. *Let $\{\mathcal{S}, \mathcal{R}, \mathcal{C}\}$ be a reaction network. Suppose that when modeled deterministically with mass action kinetics and rate constants $\{\kappa_{y\to y'}\}$ the system is complex balanced with a complex balanced equilibrium $c \in \mathbb{R}^n_{\geq 0}$. For $V > 0$, let $\{\kappa^V_{y\to y'}\}$ satisfy*

(3.5). *Then the stochastically modeled system on* $\mathbb{Z}^n_{\geq 0}$ *with rate constants* $\{\kappa^V_{y\to y'}\}$ *and intensity functions* (3.6) *admits the stationary distribution*

$$\pi^V(x) = \prod_{i=1}^{n} \frac{(Vc_i)^{x_i}}{x_i!} e^{-Vc_i}, \quad x \in \mathbb{Z}^n_{\geq 0}. \tag{3.7}$$

An immediate implication of Theorem 3.6 is that a stationary distribution for the scaled model $\bar{X}^V$ is

$$\tilde{\pi}^V(\tilde{x}^V) = \pi^V(V\tilde{x}^V), \quad \text{for} \quad \tilde{x}^V \in \frac{1}{V}\mathbb{Z}^n_{\geq 0}. \tag{3.8}$$

The main finding in [7] is concerned with the scaling limit of the stationary distribution $\tilde{\pi}^V$ of (3.8).

Theorem 3.7. *Let* $\{S, C, R\}$ *be a reaction network and let* $\{\kappa_{y\to y'}\}$ *be a choice of rate constants. Suppose that, modeled deterministically, the system is complex balanced. For* $V > 0$, *let* $\{\kappa^V_{y\to y'}\}$ *be related to* $\{\kappa_{y\to y'}\}$ *via* (3.5). *Fix a sequence of points* $\tilde{x}^V \in \frac{1}{V}\mathbb{Z}^n_{\geq 0}$ *for which* $\lim_{V\to\infty} \tilde{x}^V = \tilde{x} \in \mathbb{R}^n_{>0}$. *Further let* c *be the unique complex balanced equilibrium within the positive stoichiometric compatibility class of* $\tilde{x}$.

Let π^V *be given by* (3.7) *and let* $\tilde{\pi}^V$ *be as in* (3.8), *then*

$$\lim_{V\to\infty} \left[-\frac{1}{V} \ln(\tilde{\pi}^V(\tilde{x}^V)) \right] = \mathcal{V}(\tilde{x}),$$

where $\mathcal{V}$ *is the Lyapunov function for the ODE model satisfying* (3.4).

3.2 Main results for complex balanced networks with non-mass action kinetics

In this section, we first show that for stochastically modeled reaction networks with non-mass action kinetics defined via (3.2) whose associated mass action system is complex balanced, the stationary measure (3.3) can be normalized to yield a stationary distribution. We further show that these stochastic models are non-explosive. We then extend Theorems 3.3 and 3.7 to the non-mass action case.

3.2.1 Existence of a stationary distribution and non-explosivity

We begin with a theorem proving that the stochastic models considered in Theorem 3.2 are positive recurrent when only mild growth conditions are placed on the functions θ_i.

Theorem 3.8. *Let $\{S, C, R\}$ be a reaction network with rate constants $\{\kappa_{y\to y'}\}$. Suppose that when modeled deterministically, the associated mass action system is complex balanced with equilibrium $c \in \mathbb{R}^n_{>0}$. Suppose that θ_i and $\lambda_{y\to y'}$ satisfy the conditions in and around (3.2). Moreover, suppose that for each i we have $\lim_{x\to\infty} \theta_i(x) = \infty$. Then,*

1. *the measure π given in (3.3) is summable over $\mathbb{Z}^n_{\geq 0}$, and a stationary distribution exists for the stochastically modeled process, and moreover*

2. *the stochastically modeled process is non-explosive.*

Proof. We first show that π is summable over $\mathbb{Z}^n_{\geq 0}$. We have

$$\sum_{x\in\mathbb{Z}^n_{\geq 0}} \pi(x) = \sum_{x\in\mathbb{Z}^n_{\geq 0}} \prod_{i=1}^n \frac{c_i^{x_i}}{\theta_i(1)\cdots\theta_i(x_i)} = \prod_{i=1}^n \left(\sum_{x_i\in\mathbb{Z}_{\geq 0}} \frac{c_i^{x_i}}{\theta_i(1)\cdots\theta_i(x_i)} \right)$$

so long as each sum in the final expression is finite. Thus it is sufficient to prove that $\sum_{x\in\mathbb{Z}_{\geq 0}} \frac{c_i^x}{\theta_i(1)\cdots\theta_i(x)}$ is finite for each i. By the ratio test

$$\lim_{x\to\infty} \frac{c_i^{x+1}}{\theta_i(1)\cdots\theta_i(x+1)} \cdot \left(\frac{c_i^x}{\theta_i(1)\cdots\theta_i(x)} \right)^{-1} = \lim_{x\to\infty} \frac{c_i}{\theta_i(x+1)} = 0 < 1$$

where the last equality is due to the assumption that $\lim_{x\to\infty} \theta_i(x) = \infty$. Hence the sum is convergent.

We turn to showing that the process is non-explosive. From [6], to show that the process is non-explosive, it is sufficient to show

$$\sum_{x\in\mathbb{Z}^n_{\geq 0}}\left(\pi(x)\sum_{y\to y'\in\mathcal{R}}\lambda_{y\to y'}(x)\right)<\infty.$$

From (3.2) and (3.2), we need to show

$$\sum_{x\in\mathbb{Z}^n_{\geq 0}}\left(\prod_{i=1}^n\frac{c_i^{x_i}}{\theta_i(1)\cdots\theta_i(x_i)}\sum_{y\to y'\in\mathcal{R}}\kappa_{y\to y'}\prod_{i=1}^n\theta_i(x_i)\cdots\theta_i(x_i-y_i+1)\right)<\infty.$$

Let $s_i=\max_{y\to y'}\{y_i\}$ and $\kappa=\max_{y\to y'}\kappa_{y\to y'}$, where the max is over all source complexes, and let R be the number of reactions. Let $n_i>s_i$ be such that $\theta_i(x_i)>1,\cdots,\theta_i(x_i-s_i+1)>1$ for all $x_i>n_i$. Then

$$\begin{aligned}
&\sum_{x\in\mathbb{Z}^n_{\geq 0};x_i>n_i}\prod_{i=1}^n\frac{c_i^{x_i}}{\theta_i(1)\cdots\theta_i(x_i)}\sum_{y\to y'\in\mathcal{R}}\kappa_{y\to y'}\prod_{i=1}^n\theta_i(x_i)\cdots\theta_i(x_i-y_i+1)\\
&\qquad<\sum_{x\in\mathbb{Z}^n_{\geq 0};x_i>n_i}\prod_{i=1}^n\frac{c_i^{x_i}}{\theta_i(1)\cdots\theta_i(x_i)}R\kappa\prod_{i=1}^n\theta_i(x_i)\cdots\theta_i(x_i-s_i+1)\\
&\qquad=\sum_{x\in\mathbb{Z}^n_{\geq 0};x_i>n_i}\prod_{i=1}^n\frac{R\kappa c_i^{x_i}}{\theta_i(1)\cdots\theta_i(x_i-s_i)}\\
&\qquad<C\sum_{x\in\mathbb{Z}^n_{\geq 0};x_i>n_i}\prod_{i=1}^n\frac{c_i^{x_i-s_i}}{\theta_i(1)\cdots\theta_i(x_i-s_i)}<\infty
\end{aligned}$$

where $C=R\kappa\max_{i=1}^n\{c_i^{s_i}\}$, and the last inequality follows from part 1. Thus the process is non-explosive. $\square$

3.2.2 Generalization of Theorem 3.3

We are set to provide the next theorem, which is the converse statement of Theorem 3.2 and generalizes Theorem 3.3 from [16]. In the theorem below, we assume $\lim_{x\to\infty}\theta_i(x)=\infty$ for each i. In Corollary 3.11, we generalize the result to allow $\lim_{x\to\infty}\theta_i(x)\in\{0,\infty\}$ for each i.

Theorem 3.9. *Let $\{\mathcal{S}, \mathcal{R}, \mathcal{C}\}$ be a reaction network and consider the stochastically modeled system with rate constants $\{\kappa_{y\to y'}\}$ and intensity functions (3.2). Suppose that $\lim_{x\to\infty}\theta_i(x) = \infty$ for each $i = 1, \ldots n$ and that for some $c \in \mathbb{R}^n_{\geq 0}$ a stationary measure for the stochastic model satisfies (3.3). Then c is a complex balanced equilibrium for the associated deterministic model with mass action kinetics and rate constants $\{\kappa_{y\to y'}\}$.*

Proof. By assumption, we have that π satisfies

$$\sum_{y\to y'\in\mathcal{R}} \pi(x+y-y')\lambda_{y\to y'}(x+y-y') = \pi(x)\sum_{y\to y'\in\mathcal{R}}\lambda_{y\to y'}(x).$$

Plugging (3.2) and (3.3) into this equation yields

$$\sum_{y\to y'\in\mathcal{R}} \frac{c^{x+y-y'}}{\prod_{i=1}^n[\theta_i(1)\cdots\theta_i(x_i+y_i-y_i')]}\kappa_{y\to y'}\prod_{i=1}^n\theta_i(x_i+y_i-y_i')\cdots\theta_i(x_i-y_i'+1)$$
$$= \frac{c^x}{\prod_{i=1}^n[\theta_i(1)\cdots\theta_i(x_i)]}\sum_{y\to y'\in\mathcal{R}}\kappa_{y\to y'}\prod_{i=1}^n\theta_i(x_i)\cdots\theta_i(x_i-y_i+1).$$

Canceling and moving terms when necessary, we have

$$\sum_{y\to y'\in\mathcal{R}} c^{y-y'}\kappa_{y\to y'}\prod_{i=1}^n\theta_i(x_i)\cdots\theta_i(x_i-y_i'+1) = \sum_{y\to y'\in\mathcal{R}}\kappa_{y\to y'}\prod_{i=1}^n\theta_i(x_i)\cdots\theta_i(x_i-y_i+1).$$

Enumerating the reaction on the right by their product complexes, and the reactions on the left by their source complexes, the equation above becomes

$$\sum_{z\in\mathcal{C}}\prod_{i=1}^n\theta_i(x_i)\cdots\theta_i(x_i-z_i+1)\sum_{y\to y':y'=z}c^{y_k-y_k'}\kappa_{y\to y'} = \sum_{z\in\mathcal{C}}\prod_{i=1}^n\theta_i(x_i)\cdots\theta_i(x_i-z_i+1)\sum_{y\to y':y=z}\kappa_{y\to y'}.$$

Since the above holds for all $x \in \mathbb{Z}^n_{\geq 0}$, the two sides are equal as functions. Hence, if the functions in the set

$$\left\{\prod_{i=1}^n\theta_i(x_i)\cdots\theta_i(x_i-z_i+1)\right\}_{z\in\mathcal{C}} \tag{3.9}$$

are linearly independent, then we must have

$$\sum_{y\to y':y'=z}c^{y-y'}\kappa_{y\to y'} = \sum_{y\to y':y=z}\kappa_{y\to y'},$$

which is the condition for the associated mass action system to be complex balanced.

Thus it remains to show that the functions in the set (3.9) are linearly independent. We will prove that the functions are linearly independent by induction on the number of species in the lemma below.

$\square$

Lemma 3.10. *For all $n \in \mathbb{N}_{>0}$ the functions in the set* (3.9) *are linearly independent.*

Proof. We start with the case when there is one species, or when $n = 1$. Let $\mathcal{C} = \{z_1, \ldots, z_R\}$ ordered so that $z_i < z_{i+1}$ for each $i = 1, \ldots, R-1$. Suppose, in order to find a contradiction, the functions in the set (3.9) are linearly dependent. Then there exist $\alpha_1, \cdots, \alpha_r \in \mathbb{R}$ with $r \leq R$ and $\alpha_r \neq 0$, such that

$$\alpha_1 \theta(x) \cdots \theta(x - z_1 + 1) + \cdots + \alpha_r \theta(x) \cdots \theta(x - z_r + 1) = 0, \quad \text{for all} \quad x \in \mathbb{R}. \tag{3.10}$$

Let $M = \frac{|\alpha_1|}{|\alpha_r|} + \cdots + \frac{|\alpha_{r-1}|}{|\alpha_r|}$. Since $\theta(x) \to \infty$, as $x \to \infty$, we can find an $N > 0$ such that $\forall x > N$, we have $\theta(x - z_r + 1) > M$ and $\theta(x), \ldots, \theta(x - z_r + 1) \geq 1$. In this case,

$$\begin{aligned}
|\alpha_r \theta(x) \cdots \theta(x - z_r + 1)| &> M |\alpha_r| \theta(x) \cdots \theta(x - z_r + 2) \\
&= \left(\frac{|\alpha_1|}{|\alpha_r|} + \cdots + \frac{|\alpha_{r-1}|}{|\alpha_r|} \right) |\alpha_r| \theta(x) \cdots \theta(x - z_r + 2) \\
&= |\alpha_1| \theta(x) \cdots \theta(x - z_r + 2) + \cdots + |\alpha_{r-1}| \theta(x) \cdots \theta(x - z_r + 2) \\
&\geq |\alpha_1| \theta(x) \cdots \theta(x - z_1 + 1) + \cdots + |\alpha_{r-1}| \theta(x) \cdots \theta(x - z_{r-1} + 1).
\end{aligned}$$

This contradicts (3.10). Therefore, (3.9) must be linearly independent.

We turn to the inductive step. Thus, we now assume that functions of the form (3.9) for distinct complexes z are linearly independent when there are $n-1$ species. We must show that this implies linear independence when there are n species.

Enumerate the complexes as $\mathcal{C} = \{z^1, z^2, \ldots, z^R\}$. Suppose that there are $\alpha_1, \ldots, \alpha_R$ for which

$$\alpha_1 \prod_{i=1}^{n} \theta_i(x_i) \cdots \theta_i(x_i - z_i^1 + 1) + \ldots + \alpha_R \prod_{i=1}^{n} \theta_i(x_i) \cdots \theta_i(x_i - z_i^R + 1) = 0, \quad \text{for all} \quad x \in \mathbb{R}^n. \tag{3.11}$$

We will show that each $\alpha_i = 0$.

First note that we can not have $z_i^1 = z_i^2 = \ldots = z_i^R$ for each $i = 1, \ldots, n$, for otherwise all the complexes are the same. Thus, and without loss of generality, we assume that not all of the z_1^k are equal. In particular, we will assume that $z_1^1, \ldots, z_1^R$ consists of p distinct values with $2 \le p \le R$. We will also assume that the complexes are ordered so that the first r_1 terms of z_1^k are the same, the second r_2 terms are the same, etc. That is,

$$z_1^1 = \ldots = z_1^{r_1}, \quad z_1^{r_1+1} = \ldots = z_1^{r_1+r_2}, \ \ldots, \quad z_1^{n-r_p+1} = \ldots = z_1^R. \tag{3.12}$$

We now consider the left hand side of (3.11) as a function of x_1 alone. For $j = 1, \ldots, p$, we define

$$f_j(x_1) = \theta_1(x_1) \cdots \theta_1(x_1 - z_1^{r_1+\ldots+r_j} + 1).$$

By Lemma 3.10, the functions f_j, for $j = 1, \ldots, p$, are linearly independent. Combining similar terms in (3.11) we have

$$\begin{aligned}
&f_1(x_1)[\alpha_1 \prod_{i=2}^{n} \theta_i(x_i) \cdots \theta_i(x_i - z_i^1 + 1) + \ldots + \alpha_{r_1} \prod_{i=2}^{n} \theta_i(x_i) \cdots \theta_i(x_i - z_i^{r_1} + 1)] + \qquad (3.13)\\
&f_2(x_1)[\alpha_{r_1+1} \prod_{i=2}^{n} \theta_i(x_i) \cdots \theta_i(x_i - z_i^{r_1+1} + 1) + \ldots + \alpha_{r_2} \prod_{i=2}^{n} \theta_i(x_i) \cdots \theta_i(x_i - z_i^{r_1+r_2} + 1)] + \\
&\vdots \\
&+ f_p(x_1)[\alpha_{n-r_p+1} \prod_{i=2}^{n} \theta_i(x_i) \cdots \theta_i(x_i - z_i^{n-r_p+1} + 1) + \ldots + \alpha_R \prod_{i=2}^{n} \theta_i(x_i) \cdots \theta_i(x_i - z_i^R + 1)] = 0.
\end{aligned}$$

From the independence of the f_j, it must be the case that each bracketed term above is zero.

Without loss of generality, we just consider the first bracketed term in (3.13):

$$\alpha_1 \prod_{i=2}^{n} \theta_i(x_i) \cdots \theta_i(x_i - z_i^1 + 1) + \ldots + \alpha_{r_1} \prod_{i=2}^{n} \theta_i(x_i) \cdots \theta_i(x_i - z_i^{r_1} + 1), \tag{3.14}$$

which we know is equal to zero. The goal now is to apply our inductive hypothesis to conclude that each of $\alpha_1, \ldots, \alpha_{r_1}$ is equal to zero.

For each of $k = 1, \ldots, r_1$, we let $\tilde{z}^k = (z_2^k, \ldots, z_m^k)$. Then each term in the sum (3.14) is a function on $\mathbb{R}^{m-1}$ of the general form (3.9) with new complexes $\tilde{z}^k \in \mathbb{R}^{m-1}$. To use the inductive hypothesis, we must argue that the $\tilde{z}^k$ are distinct. Consider, for example, the first two terms: $\tilde{z}^1$ and $\tilde{z}^2$. By (3.12), we know that $z_1^1 = z_1^2$; that is, the coefficient of species 1 for the two complexes are the same. If we also had $\tilde{z}^1 = \tilde{z}^2$, then all the coefficients of the species would be the same for the two complexes, contradicting the fact that they are distinct complexes (i.e. $z^1 \neq z^2$). Hence, it must be that $\tilde{z}^1 \neq \tilde{z}^2$. Thus, by the inductive hypothesis, all the terms of the sum (3.14) are linearly independent, and $\alpha_1 = \cdots = \alpha_{r_1} = 0$. Repeating this argument for the other bracketed terms completes the proof.

We have proven the independence of (3.9) in all cases, which completes the proof of the lemma. □

The following relaxes the condition in Theorem 3.9 that the limit of the functions θ_i must be infinity.

Corollary 3.11. *Let $\{\mathcal{S}, \mathcal{R}, \mathcal{C}\}$ be a reaction network and consider the stochastically modeled system with rate constants $\{\kappa_{y \to y'}\}$ and intensity functions (3.2). Suppose that*

$\lim_{x\to\infty}\theta_i(x) \in \{0,\infty\}$ for each $i = 1,\dots,m$ and that for some $c \in \mathbb{R}^n_{\geq 0}$ a stationary measure for the stochastic model satisfies (3.3). Suppose further that $\theta_i(x) > 0$ for x large enough. Then c is a complex balanced equilibrium for the associated deterministic model with mass action kinetics and rate constants $\{\kappa_{y\to y'}\}$.

Proof. Without loss of generality, assume $\lim_{x\to\infty}\theta_i(x) = 0$ for $i \leq \ell$ and $\lim_{x\to\infty}\theta_i(x) = \infty$ for $i \geq \ell+1$. The proof is the same as that of Theorem 3.9 in that we must prove the linear independence of the functions in (3.9). Let

$$\alpha_1 \prod_{i=1}^{n} \theta_i(x_i)\cdots\theta_i(x_i - z_i^1 + 1) + \ldots + \alpha_R \prod_{i=1}^{n} \theta_i(x_i)\cdots\theta_i(x_i - z_i^R + 1) = 0. \tag{3.15}$$

For x large enough that $\theta_i(x) > 0$, let $\phi_i(x) = \frac{1}{\theta_i(x)}$ for each $i \leq \ell$. Then we have $\lim_{x\to\infty}\phi_i(x) = \infty$. Now (3.15) becomes

$$\frac{\alpha_1 \prod_{i=\ell+1}^{n} \theta_i(x_i)\cdots\theta_i(x_i - z_i^1 + 1)}{\prod_{i=1}^{\ell} \phi_i(x_i)\cdots\phi_i(x_i - z_i^1 + 1)} + \ldots + \frac{\alpha_R \prod_{i=\ell+1}^{n} \theta_i(x_i)\cdots\theta_i(x_i - z_i^R + 1)}{\prod_{i=1}^{\ell} \phi_i(x_i)\cdots\phi_i(x_i - z_i^R + 1)} = 0. \tag{3.16}$$

Let $w_k = \max_{1\leq j\leq n}\{z_k^j\}$. Then from (3.16) we have

$$\begin{aligned}
&\alpha_1 \prod_{i=1}^{\ell} \phi_i(x_i - z_i^1)\cdots\phi_i(x_i - w_i) \prod_{i=\ell+1}^{n} \theta_i(x_i)\cdots\theta_i(x_i - z_i^1 + 1) + \\
&\ldots + \alpha_R \prod_{i=1}^{\ell} \phi_i(x_i - z_i^n)\cdots\phi_i(x_i - w_i) \prod_{i=\ell+1}^{n} \theta_i(x_i)\cdots\theta_i(x_i - z_i^R + 1) = 0.
\end{aligned}$$

This is similar to the set-up of Theorem 3.9 (since each $\phi_i(x) \to \infty$, as $x \to \infty$) and we can conclude $\alpha_1 = \ldots = \alpha_n = 0$ and complete the proof. □

3.2.3 Generalization of Theorem 3.7

This section is concerned with the convergence of the non-equilibrium potential of the stationary distribution of systems with general kinetics, under some appropriate scaling.

In particular, we would like to have a similar result as Theorem 3.7 for the case of general kinetics. One difficulty that arises is that the classical scaling is not, in general, appropriate for our purposes. This is illustrated by the example below.

Example 3. Consider the reaction network with one species A and reactions given by

$$\emptyset \leftrightharpoons A$$

with the intensity function given by (3.2), where the rate constants are $\kappa_{\emptyset\to A} = \kappa_{A\to\emptyset} = 1$ and $\theta(x) = x^2$. Consider the process under the classical scaling. We obtain a stationary distribution for the scaled model $\tilde{X}^V$ from Theorem 3.6 and (3.8):

$$\tilde{\pi}^V(\tilde{x}^V) = \pi^V(V\tilde{x}^V) = \frac{1}{M}\frac{(Vc)^{V\tilde{x}^V}}{\theta(1)\cdots\theta(V\tilde{x}^V)} = \frac{1}{M}\frac{(Vc)^{V\tilde{x}^V}}{((V\tilde{x}^V)!)^2}, \qquad \tilde{x}^V \in \frac{1}{V}\mathbb{Z}_{\geq 0}.$$

We consider the limiting behavior of the non-equilibrium potential $-\frac{1}{V}\ln(\tilde{\pi}^V(\tilde{x}^V))$, as $V \to \infty$. Using Stirling's approximation,

$$\begin{aligned}
-\frac{1}{V}\ln(\tilde{\pi}^V(\tilde{x}^V)) &= -\frac{1}{V}\ln\left(\frac{1}{M}\frac{(Vc)^{V\tilde{x}^V}}{((V\tilde{x}^V)!)^2}\right)\\
&= -\frac{1}{V}(-\ln M + V\tilde{x}^V\ln V + V\tilde{x}^V\ln c - 2\ln((V\tilde{x}^V)!))\\
&\approx -\frac{1}{V}(-\ln M + V\tilde{x}^V\ln V + V\tilde{x}^V\ln c - 2V\tilde{x}^V\ln V\tilde{x}^V + 2V\tilde{x}^V)\\
&= -\frac{1}{V}(-\ln M + V\tilde{x}^V\ln c - 2V\tilde{x}^V\ln\tilde{x}^V - V\tilde{x}^V\ln V + 2V\tilde{x}^V).
\end{aligned}$$

We need to estimate M when $V \to \infty$. From Lemma A.1 in the Appendix, we have

$$\begin{aligned}
\ln M &= \ln\sum_{x\in\mathbb{Z}_{\geq 0}}\frac{(Vc)^x}{(x!)^2}\\
&\approx 2(Vc)^{1/2} + a\ln(Vc) + b,
\end{aligned}$$

for some constants $a, b \in \mathbb{R}$. Thus

$$
\begin{aligned}
-\frac{1}{V}\ln(\tilde{\pi}^V(\tilde{x}^V)) &\approx -\frac{1}{V}(-2(Vc)^{1/2} - a\ln(Vc) - b + V\tilde{x}^V \ln c - 2V\tilde{x}^V \ln \tilde{x}^V - V\tilde{x}^V \ln V + 2V\tilde{x}^V) \\
&= 2\frac{c^{1/2}}{V^{1/2}} + \frac{a\ln(Vc)}{V} + \frac{b}{V} - \tilde{x}^V \ln c + 2\tilde{x}^V \ln \tilde{x}^V + \tilde{x}^V \ln V - 2\tilde{x}^V.
\end{aligned}
$$

Clearly, $\lim_{V\to\infty} -\frac{1}{V}\ln(\tilde{\pi}^V(\tilde{x}^V)) = \infty$, and we do not have convergence of the non-equilibrium potential under the classical scaling. $\square$

With the above example in mind, we provide an alternative scaling.

The modified scaling

Define $|y| = \sum_i y_i$ and let V be a scaling parameter. For each reaction $y \to y'$ let $\kappa_{y\to y'}$ be a positive parameter. We now define the rate constant for $y \to y'$ as

$$
\kappa^V_{y\to y'} = \frac{\kappa_{y\to y'}}{V^{d\cdot y - 1}} \tag{3.17}
$$

where the parameter d is a vector to be chosen (they will depend upon the limiting values $\lim_{x\to\infty}\theta_i(x)$). Note that the classical scaling is the case when $d = (1, 1, \ldots, 1)$. Then we define the scaled intensity function

$$
\lambda^V_{y\to y'}(x) = \frac{\kappa_{y\to y'}}{V^{d\cdot y - 1}} \prod_{i=1}^{n} \theta_i(x_i)\cdots\theta_i(x_i - y_i + 1), \tag{3.18}
$$

where, as usual, $\theta_i : \mathbb{Z} \to \mathbb{R}_{\geq 0}$, and $\theta_i(x) = 0$ if $x \leq 0$.

Theorem 3.12. *Let $\{S, C, R\}$ be a reaction network with rate constants $\{\kappa_{y\to y'}\}$. Suppose that when modeled deterministically, the associated mass action system is complex balance with equilibrium $c \in \mathbb{R}^n_{>0}$. For some V, let $\{\kappa^V_{y\to y'}\}$ be related to the $\{\kappa_{y\to y'}\}$ via* (3.17)*. Then the stochastically modeled system with scaled intensity function* (3.18) *has* *stationary measure*

$$\pi_*^V(x) = \prod_{i=1}^n \frac{(V^{d_i} c_i)^{x_i}}{\theta_i(1)\cdots\theta_i(x_i)}, \qquad \textit{where} \qquad x \in \mathbb{Z}^n_{\geq 0}. \tag{3.19}$$

If (3.19) *is summable, then a normalizing constant* M *can be found so that*

$$\pi^V(x) = \frac{1}{M}\prod_{i=1}^n \frac{(V^{d_i} c_i)^{x_i}}{\theta_i(1)\cdots\theta_i(x_i)}, \qquad \textit{where} \qquad x \in \mathbb{Z}^n_{\geq 0}, \tag{3.20}$$

is a stationary distribution.

Proof. The proof is similar to that of Theorem 2.9, as found in [8], except care must be taken to ensure that the terms associated with the scaling parameter V cancel appropriately. □

Let X^V be the process associated with the intensities (3.18) and let $\tilde{X}^V = V^{-1}X^V$ be the scaled process. By Theorem 3.12, we have that for $x^V \in \frac{1}{V}\mathbb{Z}^n_{\geq 0}$

$$\tilde{\pi}_*^V(\tilde{x}^V) = \pi_*^V(V\tilde{x}^V), \tag{3.21}$$

is a stationary measure for the scaled process. If (3.21) is summable, which is ensured by Theorem 3.8 so long as $\theta_i(x) \to \infty$ as $x \to \infty$, then

$$\tilde{\pi}^V(\tilde{x}^V) = \pi^V(V\tilde{x}^V), \tag{3.22}$$

is a stationary distribution. In the next section, we consider the the limiting behavior of $-\frac{1}{V}\ln(\tilde{\pi}^V(\tilde{x}^V))$ as $V \to \infty$ for a class of θ_i.

Limiting behavior of $-\frac{1}{V}\ln(\tilde{\pi}^V(\tilde{x}^V))$

We make the following assumption on the functions θ_i.

Assumption 1. *We assume that (i) $\theta_i : \mathbb{Z} \to \mathbb{R}_{\geq 0}$, (ii) $\theta_i(x) = 0$ if $x \leq 0$, and (iii) there exists $d, A \in \mathbb{R}^n_{>0}$ such that $\lim_{x_i \to \infty} \frac{\theta_i(x_i)}{x_i^{d_i}} = A_i$ for each i.*

Roughly speaking, this class of functions act like power functions when x is large. We will utilize functions satisfying Assumption 1 to build intensity functions as in (3.18). We will show that if the deterministic mass action system is complex balanced, then the limiting behavior of the scaled non-equilibrium potential of the stochastically modeled system with intensities (3.18) is a Lyapunov function for the ODE system

$$\dot{x} = \sum_{y \to y'} \kappa_{y \to y'} (Ax^d)^y (y' - y), \qquad \text{for} \qquad x \in \mathbb{R}^n_{\geq 0}. \tag{3.23}$$

where we recall that Ax^d is the vector with ith component $A_i x_i^{d_i}$. This result therefore generalizes Theorem 3.7 (which is Theorem 8 in [7]).

Lemma 3.13. *Let $\{S, C, R\}$ be a reaction network with rate constants $\{\kappa_{y \to y'}\}$. Suppose that when modeled deterministically, the associated mass action system is complex balanced with equilibrium $c \in \mathbb{R}^n_{>0}$. Let $d, A \in \mathbb{R}^n_{>0}$. Then the system (3.23) is complex balanced with equilibrium vector $\tilde{c}$ satisfying*

$$\tilde{c}_i = \left(\frac{c_i}{A_i} \right)^{1/d_i}.$$

Proof. The proof consists of verifying that for each $z \in \mathcal{C}$,

$$\sum_{y \to y' : y = z} \kappa_{y \to y'} (A\tilde{c}^d)^y = \sum_{y \to y' : y' = z} \kappa_{y \to y'} (A\tilde{c}^d)^y,$$

where the sum on the left consists of those reactions with source complex z and the sum on the right consists of those with product complex z. This is immediate from the definition of $\tilde{c}$. □

We now turn to the scaled models, and prove that the properly scaled non-equilibrium potential converges to a Lyapunov function for the ODE system (3.23).

Theorem 3.14. *Let $\{S, C, R\}$ be a reaction network with rate constants $\{\kappa_{y\to y'}\}$. Suppose that when modeled deterministically, the associated mass action system is complex balanced with equilibrium $c \in \mathbb{R}^n_{>0}$.*

Fix $d, A \in \mathbb{R}^n_{>0}$ and let θ_i be a choice of functions satisfying Assumption 1. For $V > 0$ and the $d > 0$ already selected, let $\{\kappa^V_{y\to y'}\}$ be related to $\{\kappa_{y\to y'}\}$ as in (3.17) and let the intensity functions for the stochastically modeled system be (3.18).

Let $\tilde{\pi}^V$ be the stationary distribution for the scaled process guaranteed to exist by Theorems 3.12 and 3.8 and given by (3.20).

Fix a sequence of points $\tilde{x}^V \in \frac{1}{V}\mathbb{Z}^n_{\geq 0}$ for which $\lim_{V\to\infty} \tilde{x}^V = \tilde{x} \in \mathbb{Z}^n_{>0}$. Then

$$\lim_{V\to\infty}\left[-\frac{1}{V}\ln(\tilde{\pi}^V(\tilde{x}^V))\right] = \mathcal{V}(\tilde{x}) = \sum_{i=1}^n [\tilde{x}_i(d_i \ln(\tilde{x}_i) - \ln(c_i) - d_i + \ln(A_i))] + \sum_{i=1}^n d_i (c_i/A_i)^{1/d_i}, \tag{3.24}$$

where $\mathcal{V}$ is defined by the final equality, and moreover $\mathcal{V}$ is a Lyapunov function for the ODE system (3.23).

Note that by taking $d = (1, .., 1)$ and $A = (1, .., 1)$, the limit of the θ_i in Assumption 1 is simply mass action kinetics. Hence, the main result in [7] is contained within the above theorem.

Proof. Using (3.20) and (3.22) we have

$$
\begin{aligned}
-\frac{1}{V}\ln(\tilde{\pi}^V(\tilde{x}^V)) &= -\frac{1}{V}\ln\left(\frac{1}{M}\prod_{i=1}^{n}\frac{(V^{d_i}c_i)^{V\tilde{x}_i^V}}{\theta_i(1)\cdots\theta_i(V\tilde{x}_i^V)}\right)\\
&= -\frac{1}{V}\Big(-\ln M + \sum_{i=1}^{n} V\tilde{x}_i^V\ln(V^{d_i}c_i) - \sum_{i=1}^{n}\ln(\theta_i(1)\cdots\theta_i(V\tilde{x}_i^V))\Big)\\
&= -\frac{1}{V}\Big(-\ln M + \sum_{i=1}^{n} Vd_i x_i^V\ln(V) + \sum_{i=1}^{n} V\tilde{x}_i^V\ln(c_i) - \sum_{i=1}^{n}\ln(\theta_i(1)\cdots\theta_i(V\tilde{x}_i^V))\Big)\\
&= -\frac{1}{V}\Big(-\ln M + \sum_{i=1}^{n} Vd_i\tilde{x}_i^V\ln(V) + \sum_{i=1}^{n} V\tilde{x}_i^V\ln(c_i) - \sum_{i=1}^{n}\ln((V\tilde{x}_i^V!)^{d_i})\\
&\quad + \sum_{i=1}^{n}\ln((V\tilde{x}_i^V!)^{d_i}) - \sum_{i=1}^{n}\ln(\theta_i(1)\cdots\theta_i(V\tilde{x}_i^V))\Big)
\end{aligned}
$$

We analyze the limiting behavior of the different pieces of the last expression.

1. We begin with the first term

$$\frac{1}{V}\ln M = \frac{1}{V}\ln\left(\sum_{x\in\mathbb{Z}^n}\frac{(V^d c)^x}{\prod_{i=1}^n\theta_i(1)\cdots\theta_i(x_i)}\right),$$

where M is defined using (3.20). In Lemma A.2 in the appendix we show that as $V\to\infty$

$$
\begin{aligned}
\frac{1}{V}\ln\left(\sum_{x\in\mathbb{Z}^n}\frac{(V^d c)^x}{\prod_{i=1}^n\theta_i(1)\cdots\theta_i(x_i)}\right) &\sim \frac{1}{V}\ln\left(\sum_{x\in\mathbb{Z}^n}\frac{(V^d c)^x}{\prod_{i=1}^n A_i^{x_i}(x_i!)^{d_i}}\right)\\
&= \frac{1}{V}\ln\left(\sum_{x\in\mathbb{Z}^n}\frac{(V^d cA^{-1})^x}{\prod_{i=1}^n(x_i!)^{d_i}}\right), \qquad (3.25)
\end{aligned}
$$

where by $a_V\sim b_V$, as $V\to\infty$, we mean $\lim_{V\to\infty}(a_V-b_V)=0$. We may then apply Lemma A.1 to (3.25) to conclude there are constants a, b such that

$$\frac{1}{V}\ln\left(\sum_{x\in\mathbb{Z}^n}\frac{(V^d cA^{-1})^x}{\prod_{i=1}^n(x_i!)^{d_i}}\right) \sim \frac{1}{V}\sum_{i=1}^{n}(d_i(V^{d_i}c_iA_i^{-1})^{1/d_i} + a\ln(V^{d_i}c_iA_i^{-1}) + b).$$

Taking the limit $V \to \infty$, we see that only the first term remains, which yields

$$\lim_{V\to\infty} \frac{1}{V} \ln M = \sum_{i=1}^{n} d_i (c_i/A_i)^{1/d_i}.$$

2. We use Stirling's approximation with the middle terms

$$\begin{aligned}
-\frac{1}{V}\Bigg(\sum_{i=1}^{n} (V d_i \tilde{x}_i^V \ln(V) + V\tilde{x}_i^V \ln(c_i)) - \sum_{i=1}^{n} \ln((V\tilde{x}_i^V!)^{d_i}) \Bigg) \\
\sim -\frac{1}{V}\Bigg(\sum_{i=1}^{n} V d_i \tilde{x}_i^V \ln(V) + V\tilde{x}_i^V \ln(c_i) - \sum_{i=1}^{n} d_i((V\tilde{x}_i^V)\ln(V\tilde{x}_i^V) - V\tilde{x}_i^V) \Bigg) \\
= -\frac{1}{V}\Bigg(\sum_{i=1}^{n} V\tilde{x}_i^V \ln(c_i) - \sum_{i=1}^{n} d_i(V\tilde{x}_i^V)\ln(\tilde{x}_i^V) + \sum_{i=1}^{n} d_i V\tilde{x}_i^V \Bigg) \\
= \sum_{i=1}^{n} d_i\tilde{x}_i^V \ln(\tilde{x}_i^V) - \sum_{i=1}^{n} \tilde{x}_i^V \ln(c_i) - \sum_{i=1}^{n} d_i\tilde{x}_i^V.
\end{aligned}$$

Taking the limit $V \to \infty$, and noting that $\tilde{x}^V \to \tilde{x}$, we have

$$\begin{aligned}
\lim_{V\to\infty} -\frac{1}{V}\Bigg(\sum_{i=1}^{n} (V d_i \tilde{x}_i^V \ln(V) + V\tilde{x}_i^V \ln(c_i)) - \sum_{i=1}^{n} \ln((V\tilde{x}_i^V!)^{d_i}) \Bigg) \\
= \sum_{i=1}^{n} d_i\tilde{x}_i \ln(\tilde{x}_i) - \sum_{i=1}^{n} \tilde{x}_i \ln(c_i) - \sum_{i=1}^{n} d_i\tilde{x}_i.
\end{aligned}$$

3. We turn to the final term. By using an argument similar to (3.25), there is a constant $C > 0$ for which

$$\begin{aligned}
-\sum_{i=1}^{n} \frac{1}{V}[\ln((V\tilde{x}_i^V!)^{d_i}) - \ln(\theta_i(1)\cdots\theta_i(V\tilde{x}_i^V))] &= -\frac{1}{V}\ln\left(\frac{(V\tilde{x}^V!)^d}{\prod_{i=1}^n \theta_i(1)\cdots\theta_i(V\tilde{x}_i^V)} \right) \\
&\sim -\frac{1}{V}\ln\left(\frac{C}{(A)^{V\tilde{x}^V}} \right) \qquad (3.26) \\
&= -\frac{1}{V}\left(\ln C - \sum_{i=1}^{n} V\tilde{x}_i^V \ln(A_i) \right).
\end{aligned}$$

Taking the limit $V \to \infty$, and noting that $\tilde{x}^V \to \tilde{x}$, we have

$$\lim_{V\to\infty} -\sum_{i=1}^{n} \frac{1}{V}[\ln((V\tilde{x}_i^V!)^{d_i}) - \ln(\theta_i(1)\cdots\theta_i(V\tilde{x}_i^V))] = \sum_{i=1}^{n} \tilde{x}_i \ln(A_i).$$

Combining the three parts, we conclude (3.24) holds. The fact that the limit is a Lyapunov function is proven in Lemma 3.15 below. □

Lemma 3.15. *The function given by* (3.24),

$$\mathcal{V}(x) = \sum_{i=1}^{n} [x_i(d_i \ln(x_i) - \ln(c_i) - d_i + \ln(A_i)) + d_i(c_i/A_i)^{1/d_i}], \quad x \in \mathbb{Z}^n_{\geq 0},$$

is a Lyapunov function for the system (3.23).

Proof. We have

$$\nabla\mathcal{V}(x) = (d_1 \ln(x_1) - \ln(c_1) + \ln(A_1), \ldots, d_n \ln(x_n) - \ln(c_n) + \ln(A_n)).$$

Let f be the right-hand side of (3.23) and recall that c is a complex balanced equilibrium of the mass action model. We have

$$\begin{aligned}
\nabla\mathcal{V}(x) \cdot f(x) &= \sum_{y \to y' \in \mathcal{R}} \kappa_{y \to y'} (Ax^d)^y \left(\ln(x^d) - \ln\left(\frac{c}{A}\right)\right) \cdot (y' - y) \\
&= \sum_{y \to y' \in \mathcal{R}} \kappa_{y \to y'} c^y \frac{(Ax^d)^y}{c^y} \left(\ln\left(\frac{Ax^d}{c}\right)^{y'} - \ln\left(\frac{Ax^d}{c}\right)^{y}\right) \\
&\leq \sum_{y \to y' \in \mathcal{R}} \kappa_{y \to y'} c^{y_k} \left(\left(\frac{Ax^d}{c}\right)^{y'} - \left(\frac{Ax^d}{c}\right)^{y}\right) \\
&= \sum_{y \to y' \in \mathcal{R}} (Ax^d)^{y'} \kappa_{y \to y'} c^{y - y'} - (Ax^d)^y \kappa_{y \to y'} \\
&= \sum_{z \in \mathcal{C}} \left[\sum_{y \to y' : y' = z} (Ax^d)^{y'} \kappa_{y \to y'} c^{y - y'} - \sum_{y \to y' : y = z} (Ax^d)^y \kappa_{y \to y'}\right] \\
&= \sum_{z \in \mathcal{C}} (Ax^d)^z \left[\sum_{y \to y' : y' = z} \kappa_{y \to y'} c^{y_k - y'_k} - \sum_{y \to y' : y = z} \kappa_{y \to y'}\right] = 0,
\end{aligned}$$

where we used the inequality $a(\ln b - \ln a) \leq b - a$ and the last equation holds because c is the complex balanced equilibrium for the mass-action system. □

Chapter 4

Strongly endotactic reaction networks and tier structure

As introduced in Chapter 1, a central question in reaction network theory concerns the connection between the graph or network structure and dynamical properties: given a network with a certain graph structure, what can one say about the qualitative behavior of the underlying dynamical system? Among many special network structures, *strong endotacticity* has gained attention in recent studies (see [1, 2, 18, 24]). A strongly endotactic network is essentially "inward pointing" in the sense that all reactions point inside of the convex hull formed by the *source complexes* (the complexes being consumed in reactions). Intuitively, this topological feature ensures that whenever the trajectory of the species' abundances tries to escape to infinity, there is a reaction that pulls it back to a compact region.

In order to examine the dynamical properties of strongly endotactic reaction networks, we will introduce an analytical tool called *tier structure*, which originated from [3] and [4]. The main idea around the concept of tiers is that along a trajectory towards infinity, it is possible to infer a hierarchy of reactions: the most dominant (or most likely to happen) reactions belong to tier 1, the second most dominant reactions belong to tier 2, and so on. We will also introduce the notion of a *tier descending network* to describe

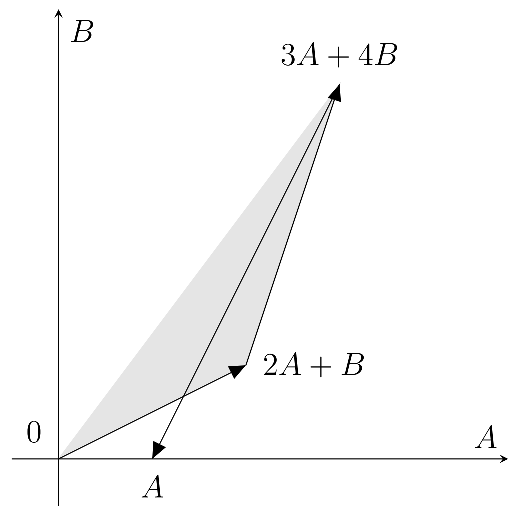

Figure 2: A strongly endotactic reaction network with two species A and B. All three reactions point inside the shaded region, which is the convex hull formed by the source complexes.

the case where along all trajectories towards infinity, there is always a reaction from tier 1 to a less dominant tier. Then we draw an important connection between tier structure and strong endotacticity: a reaction network is strongly endotactic if and only if it is tier descending.

Using this characterization, we can provide the proofs for many qualitative behaviors of strongly endotactic networks. In particular, strongly endotactic networks are *persistent* (trajectories do not touch the boundary) and *permanent* (trajectories go towards to a compact set) in the deterministic case, and they satisfy a Large Deviation Principle (LDP) and are positively recurrent with some additional assumptions in the stochastic case. While some of these results were proven before (see [1, 2, 24]) in a geometric and algebraic manner, the alternative proofs we provide, which are analytical in nature, are more streamlined and straightforward.

This section will focus on my main contribution, the LDP result. Specifically, strongly endotactic and asiphonic (an additional assumption to avoid extinction) networks satisfy a Lyapunov-like condition which in turns ensures that the networks satisfy a LDP. The main challenge in the proof came from the cases where trajectories are on the boundary. To deal with this challenge, I construct perturbed trajectories that are slightly off the boundary. I then prove the Lyapunov-like condition for these perturbed trajectories and used that result to handle the boundary cases.

Note to the readers: since this chapter focuses on sequence of concentration or species counts, we will use d to denote the number of species (instead of n in other chapters).

4.1 Strongly endotactic reaction networks

We give here the formal definition of strongly endotactic networks, that was first introduced in [24].

Definition 4.1. *Consider a reaction network $\mathcal{G}$, and a vector $w \in \mathbb{R}^d$ that is not orthogonal to the stoichiometric subspace S. We say that a complex $y \in \mathcal{C}$ is $w-$*maximal *if y is a source complex and for any other source complex y' we have $\langle w, y' - y \rangle \leq 0$.*

Definition 4.2. *A reaction network $\mathcal{G}$ is* strongly endotactic *if for all vectors $w \in \mathbb{R}^d$ that are not orthogonal to the stoichiometric subspace S the following holds:*

1. *if y is a $w-$maximal complex, then for all reactions of the form $y \to y'$ we have $\langle w, y' - y \rangle \leq 0$;*
2. *there exists a $w-$maximal complex y and a reaction $y \to y' \in \mathcal{R}$ with $\langle w, y' - y \rangle < 0$.*

Strongly endotactic networks are a generalization of weakly reversible single linkage class networks studied in [4]: the following proposition, which was proved in [24], makes the statement precise.

Proposition 4.1. *Assume $\mathcal{G}$ is a reaction network such that for any two complexes y, y' there exists a sequence of ℓ complexes, $y = y_1, y_2, \ldots, y_\ell = y'$, such that $y_j \to y_{j+1} \in \mathbb{R}$ for all $1 \leq j \leq \ell - 1$ (this condition is equivalent to saying that $\mathcal{G}$ is* weakly reversible and consists of a single linkage class*). Then, $\mathcal{G}$ is strongly endotactic.*

Strongly endotactic network are not necessarily weakly reversible single linkage class networks, an example is provided in Examples 4.

Example 4. Consider the reaction network

$$0 \to 2A + B \to 4A + 4B \to A.$$

The reaction network is strongly endotactic: to check that this statement is true, it is convenient to draw the complexes considered as vectors on a Cartesian plane, and depict the reactions as arrows among them. This is done in Figure 3. Now consider the shaded regions of Figure 4: it can be checked that

- If $w \in R_1$, then the $w-$maximal complex is $4A + 4B$. The only reaction with source complex $4A + 4B$ is $4A + 4B \to A$, and we have $\langle w, (-3, -4) \rangle < 0$.

- If $w \in R_2$, then the $w-$maximal complex is 0. The only reaction with source complex 0 is $0 \to 2A + B$, and we have $\langle w, (2, 1) \rangle < 0$.

- If $w \in R_3$, then the $w-$maximal complex is $2A + B$. The only reaction with source complex $2A + B$ is $2A + B \to 4A + 4B$, and we have $\langle w, (2, 3) \rangle < 0$.

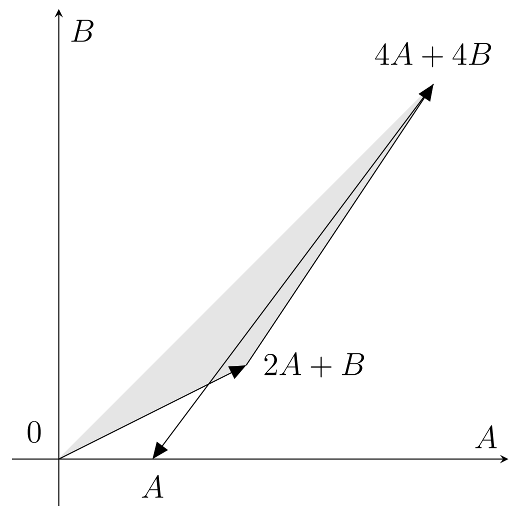

Figure 3: The reaction network of Example 4

- If w is a positive multiple of $(-1, 1)$, then the $w-$maximal complexes are 0 and $4A + 4B$, which are source complexes of $0 \to 2A + B$ and $4A + 4B \to A$. In this case, we have $\langle w, (2, 1)\rangle < 0$ and $\langle w, (-3, -4)\rangle < 0$.

- If w is a positive multiple of $(1, -2)$, then the $w-$maximal complexes are 0 and $2A + B$, which are source complexes of $0 \to 2A + B$ and $2A + B \to 4A + 4B$. In this case, we have $\langle w, (2, 1)\rangle = 0$ and $\langle w, (2, 3)\rangle < 0$.

- If w is a positive multiple of $(1, -2/3)$, then the $w-$maximal complexes are $2A+B$ and $4A+4B$, which are source complexes of $2A+B \to 4A+4B$ and $4A+4B \to A$. In this case, we have $\langle w, (2, 3)\rangle = 0$ and $\langle w, (-3, -4)\rangle < 0$.

Hence, the network is strongly endotactic. A general strategy to recognize strongly endotactic network, called the *sweep test*, and which we essentially carried out here in detail, is discussed in [24].

Next, we provide an example that is not strongly endotactic.

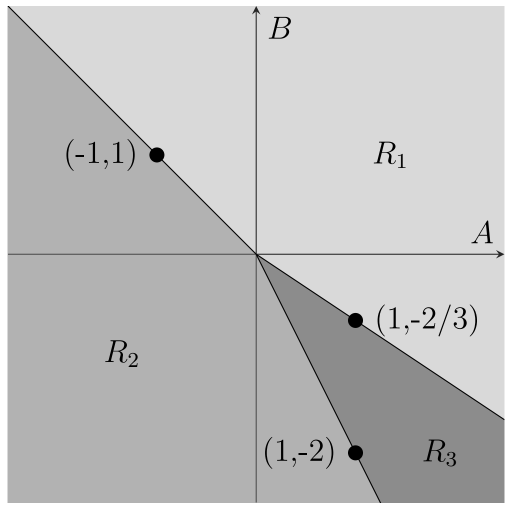

Figure 4: The space is divided into the open regions R_1, R_2, and R_3, which correspond to the loci of vectors w with different $w-$maximal complexes, and into the rays separating them. The vectors w laying on the separating lines have two $w-$maximal complexes.

Example 5. The reaction network

$$A \rightleftharpoons 2B, \quad A + C \rightleftharpoons B + C$$

is not strongly endotactic. Indeed, consider the vector $w = (1, 1, 10)$: it is not orthogonal to the stoichiometric subspace since $\langle w, (-1, 2, 0)\rangle \neq 0$, $(-1, 2, 0)$ being the reaction vector of $A \to 2B$. It can be checked that the $w-$maximal complexes are $A + C$ and $B+C$, but there is no reaction $y \to y' \in \mathcal{R}$ with $y \in \{A+C, B+C\}$ and $\langle w, y'-y\rangle < 0$.

It is interesting to note that within every stoichiometric compatibility class the amount of molecules of C is kept constant, hence the above network equipped with mass-action kinetics is equivalent to

$$B \rightleftharpoons A \rightleftharpoons 2B,$$

for a suitable choice of rate constants. Somewhat surprisingly, the latter is strongly endotactic by Proposition 4.1.

4.2 Tier structure-the main analytic tool

This section is broken into 3 subsections. In subsection 4.2.1, we introduce the relevant definitions related to tiers. We also provide a few results related to these definitions. In subsection 4.2.2, we provide Theorem 4.10, which is our main technical result and characterizes strongly endotactic networks in terms of their tier structures. Finally, in subsection 4.2.3, we collect results relating tier sequences with a commonly used Lyapunov function that plays a role in each of the subsequent results of the present paper.

4.2.1 Definitions

Definition 4.3. *A sequence $(x_n)_{n=0}^{\infty}$ of positive vectors of $\mathbb{R}^d_{>0}$ is called a* tier sequence *if*

$$\lim_{n\to\infty} \|\ln(x_n)\|_\infty = \infty$$

and for all pairs of complexes $y, y' \in \mathcal{C}$ the limit

$$\lim_{n\to\infty} x_n^{y'-y}$$

exists (it could be infinity). Moreover, we say that a tier sequence is transversal *if there exists at least one reaction $y \to y' \in \mathcal{R}$ such that*

$$\lim_{n\to\infty} |\ln(x_n^{y'-y})| = \infty.$$

Finally, a tier sequence is proper *if for all $n, m \in \mathbb{Z}_{\geq 0}$ we have $x_n - x_m \in S$.*

Remark 4.4. *Note that, given a sequence* $(x_n)_{n=0}^{\infty}$ *of positive vectors with* $\lim_{n\to\infty} \|\ln(x_n)\|_\infty = \infty$, *it is always possible to extract a subsequence that is a tier sequence. This follows from the fact that there are finitely many complexes.*

Remark 4.5. *The definition of tier sequence is tied to the choice of mass action kinetics for the reaction network. Indeed,* x_n^y *is proportional to the deterministic mass action rate function associated with a reaction whose source is* y, *and* $x_n^{y-y'}$ *is nothing but the ratio* $x_n^y/x_n^{y'}$. *Hence, a sequence is a tier sequence if a ranking of the reaction rates* λ^D *along* x_n *can be made, in the sense specified by the next definition.*

Definition 4.6. *Given a tier sequence* $(x_n)_{n=0}^{\infty}$, *we define tiers as subsets of* $\mathcal{C}$ *in the following recursive manner:*

1. *we say that a complex* y *is in* tier 1 *(and write* $y \in T^1_{(x_n)}$*) if for all complexes* $y' \in \mathcal{C}$

$$\lim_{n\to\infty} x_n^{y-y'} > 0;$$

2. *we say that a complex* y *is in* tier i *(and write* $y \in T^i_{(x_n)}$*) if there exists* $y' \in T^{i-1}_{(x_n)}$ *with*

$$\lim_{n\to\infty} x_n^{y-y'} = 0$$

and for all complexes $y' \notin \bigcup_{j=1}^{i-1} T^j_{(x_n)}$ *we have*

$$\lim_{n\to\infty} x_n^{y-y'} > 0.$$

Given a tier sequence, tiers describe a partition of $\mathcal{C}$. We further define an order relation on $\mathcal{C}$ in the following way: we write $y \precsim_{(x_n)} y'$ if $y \in T^i_{(x_n)}$, $y' \in T^j_{(x_n)}$ and $i \geq j$. Similarly, we write $y \prec_{(x_n)} y'$ if $y \in T^i_{(x_n)}$, $y' \in T^j_{(x_n)}$ and $i > j$. Note that the inequality on the indexes of the tiers is reversed, and $y \prec_{(x_n)} y'$ if and only if the ratio $x_n^y/x_n^{y'}$

converges to 0 as n tends to infinity, meaning that x_n^y is much smaller than $x_n^{y'}$ for large n. Finally, we write $y \sim_{(x_n)} y'$ if y and y' are in the same tier. Note that by definition for all complexes $y \in \mathcal{C}$ we have $y \sim_{(x_n)} y$.

Example 6. Consider the reaction network

$$A \rightleftharpoons B \rightleftharpoons 2C$$

and the sequences $(x_n)_{n=0}^{\infty}$ and $(\hat{x}_n)_{n=0}^{\infty}$ defined by

$$x_n = \left(\frac{1}{n}, 5 - \frac{1}{n} - \frac{1}{2\sqrt{n}}, \frac{1}{\sqrt{n}}\right) \quad \text{and} \quad \hat{x}_n = \left(e^n, 2e^n, \frac{1}{n}\right).$$

Then, $(x_n)_{n=0}^{\infty}$ is a proper tier sequence, which we demonstrate now. The entries $x_{n,1}$ and $x_{n,3}$ go to zero as n goes to infinity, which implies $\lim_{n\to\infty} \|\ln(x_n)\|_\infty = \infty$. Moreover,

$$\lim_{n\to\infty} x_n^{(-1,0,2)} = 1 \quad \text{and} \quad \lim_{n\to\infty} x_n^{(-1,1,0)} = \infty,$$

which implies that $(x_n)_{n=0}^{\infty}$ is a tier sequence and $A \sim_{(x_n)} 2C$ and $A \prec_{(x_n)} B$. Finally, $(x_n)_{n=0}^{\infty}$ is proper because for any $n \geq 1$

$$x_{n+1} - x_n = \left(\frac{1}{n+1} - \frac{1}{n}\right)(1, -1, 0) + \left(\frac{1}{2\sqrt{n+1}} - \frac{1}{2\sqrt{n}}\right)(0, -1, 2) \in S.$$

For what concerns $(\hat{x}_n)_{n=0}^{\infty}$, we still have $\lim_{n\to\infty} \|\ln(\hat{x}_n)\|_\infty = \infty$. Moreover,

$$\lim_{n\to\infty} \hat{x}_n^{(0,-1,2)} = 0 \quad \text{and} \quad \lim_{n\to\infty} x_n^{(-1,1,0)} = 2,$$

so $(\hat{x}_n)_{n=0}^{\infty}$ is a tier sequence and $A \sim_{(\hat{x}_n)} B$ and $2C \prec_{(\hat{x}_n)} A$. Finally, $(\hat{x}_n)_{n=0}^{\infty}$ is transversal but not proper, indeed

$$\lim_{n\to\infty} |\ln(\hat{x}_n^{(0,-1,2)})| = \infty$$

but for any $n \geq 1$

$$\langle \hat{x}_{n+1} - \hat{x}_n, (2, -2, 1)\rangle = -2(e^{n+1} - e^n) + \frac{1}{n+1} - \frac{1}{n} \neq 0,$$

and $(2, -2, 1)$ is orthogonal to S (hence $\hat{x}_{n+1} - \hat{x}_n \notin S$).

The following result connects proper and transversal tier sequences. As illustrated in Example 6, the converse does not hold.

Lemma 4.7. *A proper tier sequence is transversal.*

Proof. Consider a proper tier sequence $(x_n)_{n=0}^{\infty}$. By definition,

$$\lim_{n\to\infty} \| \ln(x_n)\|_\infty = \infty$$

and

$$\lim_{n\to\infty} | \ln(x_n^{y'-y})|$$

exists for any $y \to y' \in \mathcal{R}$. After potentially considering a subsequence, we may assume that for any $n \geq 0$

$$\begin{aligned} x_{n+1,i} &\geq x_{n,i} \quad \text{if } \limsup_{n\to\infty} \ln(x_{n,i}) = \infty; \\ x_{n+1,i} &\leq x_{n,i} \quad \text{if } \liminf_{n\to\infty} \ln(x_{n,i}) = -\infty, \end{aligned}$$

which implies that the above lim sup and lim inf are limits. It also follows that

$$\lim_{n\to\infty} | \ln(x_{n,i})| = \infty$$

for at least one index $1 \leq i \leq d$. Hence, by [3, Theorem 3.9] there exists a vector $w \in \mathbb{R}^d$ such that

$$\begin{aligned} w_i > 0 &\quad \text{if and only if } \lim_{n\to\infty} \ln(x_{n,i}) = \infty; \\ w_i < 0 &\quad \text{if and only if } \lim_{n\to\infty} \ln(x_{n,i}) = -\infty; \\ \langle w, y' - y\rangle = 0 &\quad \text{if } y \sim_{(x_n)} y'. \end{aligned}$$

In particular, it follows that

$$\lim_{n\to\infty} \langle w, x_n \rangle = \begin{cases} \infty & \text{if } \lim_{n\to\infty} \|x_n\|_\infty = \infty; \\ 0 & \text{otherwise} \end{cases}.$$

We will show that there must be an $\hat{n} \geq 1$ for which $\langle w, x_{\hat{n}} \rangle \neq 0$. First, if $\lim_{n\to\infty} \langle w, x_n \rangle = \infty$, the assertion is clear. If, on the other hand, $\lim_{n\to\infty} \langle w, x_n \rangle = 0$, then none of the $x_{n,i}$ converge to infinity. Since all the vectors $\{x_n\}_{n=0}^\infty$ are positive, and at least one of $x_{n,i}$ converges to zero, we may conclude that $\langle w, x_n \rangle < 0$ for all n.

If $(x_n)_{n=0}^\infty$ were not transversal, then we would have

$$\lim_{n\to\infty} |\ln(x_n^{y'-y})| < \infty$$

for any reaction $y \to y' \in \mathcal{R}$, which would imply that $y \sim_{(x_n)} y'$ for any $y \to y' \in \mathcal{R}$. It would follow that $\langle w, y' - y \rangle = 0$ for any $y \to y' \in \mathcal{R}$, which means $w \in S^\perp$. Let $\hat{n} \geq 1$ be such that $\langle w, x_{\hat{n}} \rangle \neq 0$. Since $(x_n)_{n=0}^\infty$ is proper, we have

$$\lim_{n\to\infty} \langle w, x_n \rangle = \langle w, x_{\hat{n}} \rangle + \lim_{n\to\infty} \langle w, x_n - x_{\hat{n}} \rangle = \langle w, x_{\hat{n}} \rangle \notin \{0, \infty\}.$$

This is a contradiction, and the proof is concluded. $\square$

For notational convenience, we give the following definition.

Definition 4.8. *Define $\mathcal{C}^S \subseteq \mathcal{C}$ to be the set of source complexes. Given a tier sequence $(x_n)_{n=0}^\infty$, we define* source tier 1 *to be the set*

$$T_{(x_n)}^{1,S} = \{y \in \mathcal{C}^S \,:\, y' \precsim_{(x_n)} y \textit{ for all } y' \in \mathcal{C}^S\}.$$

The following is a key concept of this paper, and will provide a characterization of strongly endotactic networks.

Definition 4.9. *We say that a tier sequence* $(x_n)_{n=0}^{\infty}$ *is* tier descending *if both the following statements hold:*

1. *for all* $y \in T_{(x_n)}^{1,S}$ *and all* $y \to y' \in \mathcal{R}$ *we have* $y' \precsim_{(x_n)} y$;

2. *there exist* $y \in T_{(x_n)}^{1,S}$ *and* $y \to y' \in \mathcal{R}$ *with* $y' \prec_{(x_n)} y$.

Moreover, we say that a reaction network $\mathcal{G}$ *is* tier descending *if all transversal tier sequences are tier descending.*

4.2.2 Relation between strongly endotactic networks and its tiers

We now state our first main result, which provides a characterization of strongly endotactic networks in terms of tiers.

Theorem 4.10. *A reaction network is strongly endotactic if and only if it is tier descending.*

Before proceeding with the proof of Theorem 4.10, we present an immediate corollary.

Corollary 4.11. *If a reaction network is strongly endotactic, then every proper tier sequence is tier descending. Moreover, if* $S = \mathbb{R}^d$ *then a reaction network is strongly endotactic if and only if every proper tier sequence is tier descending.*

Proof. The first part of the result follows from Lemma 4.7 and Theorem 4.10. Moreover, if $S = \mathbb{R}^d$ then any transversal tier sequence is proper (since all sequences are proper in this case), and the proof follows from Theorem 4.10. $\square$

Remark 4.12. *It is tempting to believe that if every proper tier sequence of a reaction network is tier descending, then the network is strongly endotactic. By Corollary 4.11 we see that this is true in the case when* $S = \mathbb{R}^d$*. However, this statement is false, in general. As an example, consider the reaction network*

$$A \rightleftharpoons 2B, \quad A + C \rightleftharpoons B + C.$$

The network is not strongly endotactic, as shown in Example 5. Nevertheless, every proper tier sequence is tier descending: since no reaction changes the amount of molecules of the species C*, every proper tier sequence* $(x_n)_{n=0}^{\infty}$ *is of the form*

$$x_n = (x_{n,1}, x_{n,2}, c)$$

for a constant $c \in \mathbb{R}_{>0}$*. It is then easy to check that* $(x_n)_{n=0}^{\infty}$ *is tier descending if and only if* $(\hat{x}_n)_{n=0}^{\infty}$ *defined by*

$$\hat{x}_n = (x_{n,1}, x_{n,2})$$

is tier descending for

$$B \rightleftharpoons A \rightleftharpoons 2B.$$

The latter is strongly endotactic by Proposition 4.1. Hence, each proper tier sequence (such as $(\hat{x}_n)_{n=0}^{\infty}$*) is tier descending by Corollary 4.11, thus proving our claim.*

We now proceed by providing a key lemma that will be used in the proof of Theorem 4.10.

Lemma 4.13. *If* $(x_n)_{n=0}^{\infty}$ *is a tier sequence, then there exist* $\ell \in \mathbb{Z}$ *with* $0 < \ell \leq d$*, sequences of positive real numbers* $(m_n^1)_{n=0}^{\infty}$*,* $(m_n^2)_{n=0}^{\infty}$*, ...,* $(m_n^\ell)_{n=0}^{\infty}$*, a sequence of real vectors* $(C_n)_{n=0}^{\infty}$*, vectors* $\alpha_1, \alpha_2, \dots, \alpha_\ell \in \mathbb{R}^d$ *and a subsequence* $(x_{n_k})_{k=0}^{\infty}$ *such that:*

1. $\ln(x_{n_k}) = \sum_{i=1}^{\ell} m_{n_k}^i \alpha_i + C_{n_k}$;

2. $\limsup_{k\to\infty} \|C_{n_k}\|_\infty < \infty$;

3. *For all* $1 \le i \le \ell$ *we have* $\lim_{k\to\infty} m_{n_k}^i = \infty$, *and if* $1 \le j < i \le \ell$ *then* $\lim_{k\to\infty} m_{n_k}^i / m_{n_k}^j = 0$;

4. *if* $y' \sim_{(x_n)} y$ *then* $\langle y' - y, \alpha_i \rangle = 0$ *for all* $1 \le i \le \ell$;

5. *if* $y' \prec_{(x_n)} y$ *then*

$$i_{y,y'} = \min\{1 \le i \le \ell : \langle \alpha_i, y' - y \rangle \neq 0\} \tag{4.1}$$

exists and $\langle \alpha_{i_{y,y'}}, y' - y \rangle < 0$.

Remark 4.14. *Parts 1 and 2 of the lemma show that the logarithm of a tier sequence can be substantially decomposed into fixed vectors,* α_i*, apart from a bounded error term,* C_{n_k}*. Part 3 then shows that if* $i < j$*, then the influence of the vector* α_i *is greater than the influence of the vector* α_j*. Finally, by parts 4 and 5 we see that the* α_i*'s separate complexes in a natural manner among the tiers.*

As an example, consider the reaction network

$$A \rightleftharpoons B \rightleftharpoons 2C$$

and the tier sequence

$$x_n = \left(\frac{1}{n}, 5 - \frac{1}{n} - \frac{1}{2\sqrt{n}}, \frac{1}{\sqrt{n}}\right),$$

introduced in Example 6. We have

$$\ln(x_{n_k}) = \ln(n)\left(-1, 0, -\frac{1}{2}\right) + C_n,$$

where

$$C_n = \left(0, \ln\left(5 - \frac{1}{n} - \frac{1}{2\sqrt{n}}\right), 0\right).$$

Note that $\|C_n\|_\infty < \ln(5)$ *for all* $n > 1$. *Moreover, recall that* $A \sim_{(x_n)} 2C$ *and* $A \prec_{(x_n)} B$, *which is implied also by parts 4 and 5 of the lemma, since*

$$\left\langle (1,0,-2), \left(-1,0,-\frac{1}{2}\right) \right\rangle = 0 \quad \textit{and} \quad \left\langle (1,-1,0), \left(-1,0,-\frac{1}{2}\right) \right\rangle < 0.$$

Proof of Lemma 4.13. Define $m_n^1 = \|\ln(x_n)\|_\infty$. Note that for any $n \geq 0$ we have $\|\ln(x_n)/m_n^1\|_\infty = 1$. Hence, we can consider a subsequence of $(x_n)_{n=0}^\infty$ such that

$$\alpha_1 = \lim_{k\to\infty} \frac{\ln(x_{n_k})}{m_{n_k}^1} \tag{4.2}$$

exists. We further note that α_1 cannot be zero since it is the limit of a sequence of points in the ball of radius 1 with respect to $\|\cdot\|_\infty$ in $\mathbb{R}^d$.

Since the dimension of the vectors x_n is $d < \infty$, we can further choose a subsequence such that the maximal absolute values of the entries of $\ln(x_{n_k})$ are always obtained in the same position. This implies that at least one entry of $\ln(x_{n_k})$ has absolute value constantly equal to $m_{n_k}^1$. Moreover, by (4.2) the sign of such entries will stabilize for k large enough. Hence, the vectors

$$\ln(x_{n_k}) - m_{n_k}^1 \alpha_1$$

have at least one component constantly equal to zero for k large enough.

We define $m_{n_k}^i$ and α_i iteratively in the following way: for each $j \geq 2$, if

$$\limsup_{k\to\infty} \|\ln(x_{n_k}) - \sum_{i=1}^{j-1} m_{n_k}^i \alpha_i\|_\infty = \infty,$$

then define $m^j_{n_k} = \| \ln(x_{n_k}) - \sum_{i=1}^{j-1} m^i_{n_k}\alpha_i\|_\infty$. By potentially considering a subsequence of $(x_{n_k})_{k=0}^\infty$, we can assume that

$$\alpha_j = \lim_{k\to\infty} \frac{\ln(x_{n_k}) - \sum_{i=1}^{j-1} m^i_{n_k}\alpha_i}{m^j_{n_k}}$$

exists. As before, note that α_j cannot be zero. Moreover, we can choose a subsequence such that the maximal absolute values of the entries of $\ln(x_{n_k}) - \sum_{i=1}^{j-1} m^i_{n_k}\alpha_i$ are always obtained in the same position, so by induction it follows that at least $j-1$ components of $\ln(x_{n_k}) - \sum_{i=1}^{j-1} m^i_{n_k}\alpha_i$ are equal to zero for k large enough (the argument is the same as for $j = 1$, which serves as base case). In particular, it follows that there exists a number $\ell \le d$ such that

$$\limsup_{k\to\infty} \left\| \ln(x_{n_k}) - \sum_{i=1}^{\ell} m^i_{n_k}\alpha_i \right\|_\infty < \infty.$$

We define

$$C_n = \ln(x_n) - \sum_{i=1}^{\ell} m^i_n \alpha_i.$$

Parts (1) and (2) trivially hold by the definition of C_n. For part (3), note that for all $2 \le j \le \ell$

$$\lim_{k\to\infty} \frac{m^j_{n_k}}{m^{j-1}_{n_k}} = \lim_{k\to\infty} \left\| \frac{\ln(x_{n_k}) - \sum_{i=1}^{j-1} m^i_{n_k}\alpha_i}{m^{j-1}_{n_k}} \right\|_\infty = \|\alpha_{j-1} - \alpha_{j-1}\|_\infty = 0.$$

For part (4), consider $y \sim_{(x_n)} y'$. Then,

$$0 < \lim_{k\to\infty} x_{n_k}^{y'-y} < \infty.$$

By taking the logarithm, it follows that

$$-\infty < \lim_{k\to\infty} \ln(x_{n_k}^{y'-y}) < \infty.$$

Hence, since $m^1_{n_k}$ tends to infinity as k tends to infinity, we have

$$0 = \lim_{k\to\infty} \frac{\ln(x_{n_k}^{y'-y})}{m^1_{n_k}} = \langle \alpha_1, y'-y\rangle.$$

We complete the proof of part (4) by induction: consider $1 < j \le \ell$ and assume that the statement holds for any $1 \le i \le j-1$. Then, by part (1) and since $m^j_{n_k}$ tends to infinity as k tends to infinity, we have

$$\begin{aligned} 0 = \lim_{k\to\infty} \frac{\ln(x_{n_k}^{y'-y})}{m^j_{n_k}} &= \lim_{k\to\infty} \frac{\langle \sum_{i=1}^{\ell} m^i_{n_k}\alpha_i + C_{n_k}, y'-y\rangle}{m^j_{n_k}} \\ &= \lim_{k\to\infty} \frac{\langle \sum_{i=j}^{\ell} m^i_{n_k}\alpha_i + C_{n_k}, y'-y\rangle}{m^j_{n_k}} = \langle \alpha_j, y'-y\rangle. \end{aligned}$$

Finally, for part (5) consider $y' \prec_{(x_n)} y$. Then, we have

$$\lim_{k\to\infty} x_{n_k}^{y'-y} = 0,$$

which implies

$$-\infty = \lim_{k\to\infty} \ln(x_{n_k}^{y'-y}) = \lim_{k\to\infty} \left(\sum_{i=1}^{\ell} m^i_{n_k} \langle \alpha_i, y'-y\rangle + \langle C_{n_k}, y'-y\rangle \right). \tag{4.3}$$

Since the values $\|C_{n_k}\|_\infty$ are bounded uniformly in k, we have

$$\lim_{k\to\infty} \sum_{i=1}^{\ell} m^i_{n_k} \langle \alpha_i, y'-y\rangle = -\infty,$$

which implies that

$$i_{y,y'} = \min\{1 \le i \le \ell \,:\, \langle \alpha_i, y'-y\rangle \neq 0\}$$

exists. Moreover, by part 3 we have

$$\langle \alpha_{i_{y,y'}}, y'-y\rangle = \lim_{k\to\infty} \frac{\ln(x_{n_k}^{y'-y})}{m_{n_k}^{i_{y,y'}}}.$$

By construction, the term on the left is non-zero. Further, by (4.3) the right-hand size is non-positive. Hence, $\langle \alpha_{i_{y,y'}}, y'-y\rangle < 0$, which concludes the proof. $\square$

Now we are able to prove Theorem 4.10.

Proof of Theorem 4.10. Assume that the network is tier descending. Consider a vector w that is not orthogonal to the stoichiometric subspace S. Consider the sequence $(x_n)_{n=0}^{\infty}$ defined by

$$x_n = e^{nw}.$$

We have

$$\lim_{n\to\infty} \|\ln(x_n)\|_\infty = \lim_{n\to\infty} n\|w\|_\infty = \infty$$

and for any two complexes $y, y' \in \mathcal{C}$

$$\lim_{n\to\infty} \ln(x_n^{y'-y}) = \lim_{n\to\infty} n\langle w, y'-y\rangle = \begin{cases} -\infty & \text{if } \langle w, y'-y\rangle < 0 \\ 0 & \text{if } \langle w, y'-y\rangle = 0 \\ \infty & \text{if } \langle w, y'-y\rangle > 0 \end{cases} \cdot \tag{4.4}$$

Hence, $(x_n)_{n=0}^{\infty}$ is a tier sequence. Moreover, it is transversal: since w is not orthogonal to S, there exists a reaction $y \to y'$ with $\langle w, y'-y\rangle \neq 0$, which implies $\lim_{n\to\infty} |\ln(x_n^{y'-y})| = \infty$. It follows that $(x_n)_{n=0}^{\infty}$ is tier descending, which together with equation 4.4 concludes the proof of one direction of the result.

For the other direction, we suppose that the network is strongly endotactic. Let $(x_n)_{n=0}^{\infty}$ be a transversal tier sequence. In order to prove the result, it is sufficient to construct a vector w such that

1. $w \notin S^\perp$;

2. $\langle w, y'-y\rangle = 0$ if and only if $y' \sim_{(x_n)} y$, and $\langle w, y'-y\rangle < 0$ if and only if $y' \prec_{(x_n)} y$.

Indeed, if such a vector is constructed, then it follows that the set of $w-$maximal complexes coincides with $y \in T^{1,S}_{(x_n)}$, and by Definition 4.2 the sequence $(x_n)_{n=0}^{\infty}$ is tier descending.

Consider a subsequence $(x_{n_k})_{k=0}^{\infty}$ as in Lemma 4.13, such that there exist $\ell \in \mathbb{Z}$ with $0 < \ell \leq d$, sequences of positive real numbers $(m^1_{n_k})_{k=0}^{\infty}$, $(m^2_{n_k})_{k=0}^{\infty}$, $\ldots$, $(m^{\ell}_{n_k})_{k=0}^{\infty}$, $(C_{n_k})_{k=0}^{\infty}$, and vectors $\alpha_1, \alpha_2, \ldots, \alpha_\ell \in \mathbb{R}^d$ such that

$$\ln(x_{n_k}) = \sum_{i=1}^{\ell} m^i_{n_k} \alpha_i + C_{n_k}.$$

Note that $(x_{n_k})_{k=0}^{\infty}$ is still a transversal tier sequence, and the tier structures of $(x_n)_{n=0}^{\infty}$ and of its subsequence $(x_{n_k})_{k=0}^{\infty}$ are identical, meaning that for any $i \geq 1$ we have $T^i_{(x_n)} = T^i_{(x_{n_k})}$. Let

$$w = \sum_{i=1}^{\ell} v_i \alpha_i,$$

with the positive constants v_i defined recursively as follows: $v_\ell = 1$ and

$$v_i = 1 + \max_{\substack{y \to y' \in \mathcal{R} \\ \langle \alpha_i, y'-y \rangle \neq 0}} \left| \frac{\sum_{j=i+1}^{\ell} v_j \langle \alpha_j, y'-y \rangle}{\langle \alpha_i, y'-y \rangle} \right| \quad \text{for } 1 \leq i \leq \ell - 1.$$

We have the following:

1. Since $(x_{n_k})_{k=0}^{\infty}$ is transversal and since $\|C_{n_k}\|_\infty$ are bounded, there must exist a reaction $y \to y'$ and a vector α_i such that $\langle \alpha_i, y'-y \rangle \neq 0$. Let

   $$\hat{\imath} = \min_{1 \leq i \leq \ell \,:\, \langle \alpha_i, y'-y \rangle \neq 0}.$$

   By definition of the constants v_i, we have

   $$|v_{\hat{\imath}} \langle \alpha_{\hat{\imath}}, y'-y \rangle| > \left| \sum_{j=\hat{\imath}+1}^{\ell} v_j \langle \alpha_j, y'-y \rangle \right|,$$

hence

$$\langle w, y' - y\rangle = \sum_{j=\hat{\imath}}^{\ell} v_j \langle \alpha_j, y' - y\rangle \neq 0,$$

which is equivalent to say that $w \notin S^\perp$.

2. By Lemma 4.13(4)(5), $y' \sim_{(x_n)} y$ if and only if $\langle \alpha_i, y' - y\rangle = 0$ for all $1 \leq i \leq \ell$. By the definition of w the latter is in turn equivalent to $\langle w, y' - y\rangle = 0$. Moreover, $y' \prec_{(x_n)} y$ if and only if $\langle \alpha_{i_{y,y'}}, y' - y\rangle < 0$, where $i_{y,y'}$ is defined in (4.1), which by definition of the constants v_i is equivalent to

$$\langle w, y' - y\rangle = \sum_{j=i_{y,y'}}^{\ell} v_j \langle \alpha_j, y' - y\rangle < 0.$$

The proof is then concluded. □

4.2.3 Tier sequences and Lyapunov functions

Let $u(x) : \mathbb{R} \to \mathbb{R}_{\geq 0}$ be the function

$$u(x) = \begin{cases} x(\ln x - 1) + 1 & \text{if} \quad x > 0, \\ 1 & \text{otherwise.} \end{cases} \tag{4.5}$$

Then we define

$$U(x) = 1 + \sum_{i=1}^{d} u(x_i). \tag{4.6}$$

This function has been utilized often as a Lyapunov function in the context of reaction network theory. In particular, it was utilized in the foundational papers of the field in order prove local asympotic stability of complex balanced deterministic mass action systems [21, 26]. Moreover, it (or slight modifications thereof) has notably been used to derive the results of [4, 3, 24, 2], which are of direct interest for the present paper.

More discussion on the role of Lyapunov functions for stochastic reaction networks can be found in [7] and [9].

In the present section, we will unveil some important connections between tier sequences and the Lyapunov function (4.6) by extending the techniques of [4] to the setting of tier descending networks. We will then use these connections to develop the results presented in sections 4.3.

Lemma 4.15. *Consider a tier descending reaction network $\mathcal{G}$ and let $(x_n)_{n=0}^{\infty}$ be a transversal tier sequence. Then, for any $y \to y' \in \mathcal{R}$ with $y \precsim_{(x_n)} y'$ there exists $y^\star \in \mathcal{C}$ and $y^\star \to y^{\star\star} \in \mathcal{R}$ such that $y \precsim_{(x_n)} y^\star$, $y^{\star\star} \prec_{(x_n)} y^\star$ and for any choice of $c_1, c_2 \in \mathbb{R}_{>0}$ and $c_3, c_4 \in \mathbb{R}$ there exists $N < \infty$ with*

$$c_1 x_n^{y^\star}\left(\ln(x_n^{y^{\star\star}-y^\star}) + c_3\right) + c_2 x_n^{y}\left(\ln(x_n^{y'-y}) + c_4\right) < 0 \quad \textit{for all } n \geq N. \tag{4.7}$$

Moreover, if $x_n^{y^\star} \geq c > 0$ for all n, then for any choice of $c_1, c_2 \in \mathbb{R}_{>0}$ and $c_3, c_4 \in \mathbb{R}$ we have

$$\lim_{n\to\infty}\left(c_1 x_n^{y^\star}\left(\ln(x_n^{y^{\star\star}-y^\star}) + c_3\right) + c_2 x_n^{y}\left(\ln(x_n^{y'-y}) + c_4\right)\right) = -\infty. \tag{4.8}$$

Proof. Fix $y \to y' \in \mathcal{R}$. We consider two cases separately: $y \sim_{(x_n)} y'$ and $y \prec_{(x_n)} y'$.

Case 1. Assume that $y \sim_{(x_n)} y'$. Then

$$\lim_{n\to\infty} |\ln(x_n^{y'-y})| < \infty.$$

By the definition of a descending reaction network there must be at least one reaction $y^\star \to y^{\star\star}$ with $y^\star \in T_{(x_n)}^{1,S}$ (implying $y \precsim_{(x_n)} y^\star$) and $y^{\star\star} \prec_{(x_n)} y^\star$. Hence, we have

$$\lim_{n\to\infty} x_n^{y-y^\star} < \infty$$

and

$$\lim_{n\to\infty} \ln(x_n^{y^{\star\star}-y^\star}) = -\infty.$$

It follows that

$$c_1 x_n^{y^\star}\left(\ln(x_n^{y^{\star\star}-y^\star})+c_3\right)+c_2 x_n^{y}\left(\ln(x_n^{y'-y})+c_4\right) = x_n^{y^\star}\left(c_1\left(\ln(x_n^{y^{\star\star}-y^\star})+c_3\right)+c_2 x_n^{y-y^\star}\left(\ln(x_n^{y'-y})+c_4\right)\right)$$

is negative for n large enough, which proves (4.7). Moreover, if $x_n^{y^\star} \geq c > 0$, then (4.8) follows.

Case 2. Assume that $y \prec_{(x_n)} y'$. If (4.7) did not hold, then there would exist a subsequence $(x_{n_k})_{k=0}^{\infty}$ such that for any $y^\star \to y^{\star\star} \in \mathcal{R}$ with $y \precsim_{(x_n)} y^\star$ and $y^{\star\star} \prec_{(x_n)} y^\star$, there exist $c_1, c_2 \in \mathbb{R}_{>0}$ and $c_3, c_4 \in \mathbb{R}$ with

$$c_1 x_n^{y^\star}\left(\ln(x_n^{y^{\star\star}-y^\star})+c_3\right)+c_2 x_n^{y}\left(\ln(x_n^{y'-y})+c_4\right) \geq 0 \quad \text{for all } k \in \mathbb{Z}_{\geq 0}. \tag{4.9}$$

Our aim is to prove that such a subsequence does not exist.

Every subsequence of a descending tier sequence is still a descending tier sequence. Hence, by potentially considering a further subsequence, we can assume that $(x_{n_k})_{k=0}^{\infty}$ is as in Lemma 4.13.

Consider the sequence $(\tilde{x}_{n_k})_{k=0}^{\infty}$ defined by

$$\ln(\tilde{x}_{n_k}) = \sum_{i=1}^{i_{y',y}} m_{n_k}^i \alpha_i \tag{4.10}$$

where $i_{y',y}$ is as defined in (4.1), and exists by Lemma 4.13(5). We will first show that $(\tilde{x}_{n_k})_{k=0}^{\infty}$ is also a transversal tier sequence, and is therefore tier descending. By Lemma 4.13(3), we have

$$\lim_{k\to\infty} \frac{\|\ln(\tilde{x}_{n_k})\|_\infty}{m_{n_k}^1 \|\alpha_1\|_\infty} = 1,$$

and so $\lim_{k\to\infty} \| \ln(\tilde{x}_{n_k})\|_\infty = \infty$. Furthermore, for any two complexes $\tilde{y}, \tilde{y}' \in \mathcal{C}$ the limit

$$\lim_{k\to\infty} \tilde{x}_{n_k}^{\tilde{y}'-\tilde{y}} = \lim_{k\to\infty} e^{\sum_{i=1}^{i_{y',y}} m_{n_k}^i \langle \alpha_i, \tilde{y}'-\tilde{y}\rangle}$$

exists (it can potentially be infinity). Hence, $(\tilde{x}_{n_k})_{k=0}^\infty$ is a tier sequence. Moreover,

$$\lim_{k\to\infty} |\ln(\tilde{x}_{n_k}^{y'-y})| = \lim_{k\to\infty} \left| \sum_{i=1}^{i_{y',y}} m_{n_k}^i \langle \alpha_i, y'-y\rangle \right| = \lim_{k\to\infty} m_{n_k}^{i_{y',y}} |\langle \alpha_{i_{y',y}}, y'-y\rangle| = \infty.$$

Hence, $(\tilde{x}_{n_k})_{k=0}^\infty$ is a transversal tier sequence. Combining this with the fact that $\mathcal{G}$ is a tier descending reaction network, we may conclude that $(\tilde{x}_{n_k})_{k=0}^\infty$ is tier descending. Since $(\tilde{x}_{n_k})_{k=0}^\infty$ is tier sequence, Lemma 4.13 guarantees that it can be decomposed as detailed therein. It is straightforward to prove that the vectors and coefficients as constructed in the proof of the lemma coincide with the $m_{n_k}^i$ and α_i in (4.10), for $1 \le i \le i_{y',y}$.

By Lemma 4.13(3)(5) we have

$$\lim_{k\to\infty} \ln(\tilde{x}_{n_k}^{y'-y}) = \lim_{k\to\infty} m_{n_k}^{i_{y',y}} \langle \alpha_{i_{y',y}}, y'-y\rangle = -\infty,$$

allowing us to conclude that $\lim_{k\to\infty} \tilde{x}_{n_k}^{y'-y} = 0$. Thus, $y \prec_{(\tilde{x}_{n_k})} y'$. Since $(\tilde{x}_{n_k})_{k=0}^\infty$ is tier descending, y cannot be in $T^{1,S}_{(\tilde{x}_{n_k})}$. Hence, there must exist a complex $y^\star$ with $y \prec_{(\tilde{x}_{n_k})} y^\star$ and a reaction $y^\star \to y^{\star\star} \in \mathbb{R}$ with $y^{\star\star} \prec_{(\tilde{x}_{n_k})} y^\star$. Combining $y \prec_{(\tilde{x}_{n_k})} y^\star$ with Lemma 4.13(5), it follows that $i_{y^\star,y} \le i_{y',y}$. Hence, by Lemma 4.13(3) we may conclude

$$\lim_{k\to\infty} \ln(\tilde{x}_{n_k}^{y-y^\star}) = \lim_{k\to\infty} \ln(x_{n_k}^{y-y^\star}),$$

as they are both asymptotically equivalent to the same term. Therefore, the latter is negative infinity and $y \prec_{(x_{n_k})} y^\star$.

Similarly as above, since $y^{\star\star} \prec_{(\tilde{x}_{n_k})} y^\star$ we may conclude that $i_{y^{\star\star},y^\star} \le i_{y',y}$ and $y^{\star\star} \prec_{(x_{n_k})} y^\star$. Hence, combining $x_{n_k}^{y^\star} > 0$ and $y^{\star\star} \prec_{(x_{n_k})} y^\star$ we know that for k large enough

$$x_{n_k}^{y^\star}\left(\ln(x_{n_k}^{y^{\star\star}-y^\star})+c_3\right)<0. \tag{4.11}$$

Moreover, combining $y \prec_{(x_{n_k})} y^\star$, $i_{y^\star,y^{\star\star}} \le i_{y',y}$, and Lemma 4.13(3)(5) we have

$$\lim_{k\to\infty}\frac{x_{n_k}^{y^\star}\left(\ln(x_{n_k}^{y^{\star\star}-y^\star})+c_3\right)}{x_{n_k}^{y}\left(\ln(x_{n_k}^{y'-y})+c_4\right)}=\lim_{k\to\infty}x_{n_k}^{y^\star-y}\frac{m_{n_k}^{i_{y^{\star\star},y^\star}}\langle\alpha_{i_{y^{\star\star},y^\star}},y^{\star\star}-y^\star\rangle}{m_{n_k}^{i_{y',y}}\langle\alpha_{i_{y',y}},y'-y\rangle}=-\infty, \tag{4.12}$$

where we use that $\langle\alpha_{i_{y^{\star\star},y^\star}},y^{\star\star}-y^\star\rangle<0$ and $\langle\alpha_{i_{y',y}},y'-y\rangle>0$. By (4.11) and (4.12), for any positive constants c_1,c_2 we have

$$\limsup_{k\to\infty}\left(c_1x_n^{y^\star}\left(\ln(x_n^{y^{\star\star}-y^\star})+c_3\right)+c_2x_n^{y}\left(\ln(x_n^{y'-y})+c_4\right)\right)<0,$$

which is a contradiction of (4.9), hence (4.7) holds.

Now assume also that $x_n^{y^\star}\ge c>0$. Let $d_1,d_2\in\mathbb{R}_{>0}$ and $d_3,d_4\in\mathbb{R}$. We must show that for the particular choice of sequence $(x_n)_{n=0}^\infty$, and the particular choice of $y^\star$ and $y^{\star\star}$ we have that

$$\lim_{n\to\infty}\left(d_1x_n^{y^\star}\left(\ln(x_n^{y^{\star\star}-y^\star})+d_3\right)+d_2x_n^{y}\left(\ln(x_n^{y'-y})+d_4\right)\right)=-\infty. \tag{4.13}$$

We may apply (4.7) with $c_1=d_1/2$, $c_2=d_2$, $c_3=d_3$ and $c_4=d_4$ to conclude that for n large enough we have

$$\begin{aligned}&d_1x_n^{y^\star}\left(\ln(x_n^{y^{\star\star}-y^\star})+d_3\right)+d_2x_n^{y}\left(\ln(x_n^{y'-y})+d_4\right)\\&\qquad<d_1x_n^{y^\star}\left(\ln(x_n^{y^{\star\star}-y^\star})+d_3\right)+d_2\left(\frac{d_1/2}{d_2}\left|x_n^{y^\star}\left(\ln(x_n^{y^{\star\star}-y^\star})+d_3\right)\right|\right)\\&\qquad=\frac{d_1}{2}x_n^{y^\star}\left(\ln(x_n^{y^{\star\star}-y^\star})+d_3\right),\end{aligned} \tag{4.14}$$

where we are using that $x_n^y\ln(x_n^{y'-y})>0$ and $x_n^{y^\star}\ln(x_n^{y^{\star\star}-y^\star})<0$. Then, since $y^{\star\star}\prec_{(x_n)} y^\star$, by Lemma 4.13(3)(5) we have

$$\lim_{n\to\infty}\ln(x_n^{y^{\star\star}-y^\star})=\lim_{n\to\infty}\sum_{i=i_{y^\star,y^{\star\star}}}^{\ell}m_n^i\langle\alpha_i,y^{\star\star}-y^\star\rangle=-\infty.$$

It follows that

$$\lim_{n\to\infty} \frac{d_1}{2} x_n^{y^\star} \left(\ln(x_n^{y^{\star\star}-y^\star}) + d_3\right) \le \lim_{n\to\infty} \frac{d_1}{2} c \left(\ln(x_n^{y^{\star\star}-y^\star}) + d_3\right) = -\infty. \tag{4.15}$$

Combining (4.15) and (4.14) yields (4.13), and completes the proof.

□

Proposition 4.2. *Consider a tier descending reaction network $\mathcal{G}$. Then, for any transversal tier sequence $(x_n)_{n=0}^\infty$ and any choice of positive constants $\kappa_{y\to y'}$, there exists $N < \infty$ such that*

$$\sum_{y\to y'\in\mathcal{R}} \kappa_{y\to y'} x_n^y \ln(x_n^{y'-y}) < 0 \quad \text{for all } n \ge N. \tag{4.16}$$

Moreover, if the complex 0 is a source complex, then

$$\lim_{n\to\infty} \sum_{y\to y'\in\mathcal{R}} \kappa_{y\to y'} x_n^y \ln(x_n^{y'-y}) = -\infty. \tag{4.17}$$

Proof. The result follows from noting that for any reaction $y \to y' \in \mathcal{R}$ either $y' \prec_{(x_n)} y$ and

$$x_n^y \ln(x_{n_k}^{y'-y}) < 0,$$

or $y \precsim_{(x_n)} y'$ and Lemma 4.15 holds. Hence, since there are finitely many reactions, for any choice of positive constants $\kappa_{y\to y'}$ there exists $N < \infty$ such that (4.16) holds.

For the second part of the statement, assume that 0 is a source complex. Then, by definition of $T^{1,S}_{(x_n)}$ we have $0 \precsim_{(x_n)} y$ for all $y \in T^{1,S}_{(x_n)}$, which implies that for all $y \in T^{1,S}_{(x_n)}$

$$\lim_{n\to\infty} x_n^y = \lim_{n\to\infty} x_n^{y-0} > 0.$$

Since $(x_n)_{n=0}^\infty$ is transversal and $\mathcal{G}$ is tier descending, $(x_n)_{n=0}^\infty$ is tier descending. Hence, there is a reaction $y \to y' \in \mathcal{R}$ with $y \in T^{1,S}_{(x_n)}$ and $y' \prec_{(x_n)} y$. Hence

$$\lim_{n\to\infty} x_n^y \ln(x_n^{y'-y}) = -\infty,$$

and similarly as before (4.17) follows from Lemma 4.15.

□

4.3 Asiphonic strongly endotactic reaction networks and large deviation principle

In this section, we consider large deviations of classically scaled reaction networks. In particular, we utilize the findings of section 4.2 to recover the main results of [2, 1] in a straightforward manner.

Following [31] we introduce the family of classically scaled process indexed by a real number $V > 0$. In particular, we assume the process associated with V is a stochastic mass action system with rate constant $\kappa_{y\to y'}/V^{\|y\|_1-1}$, where $\kappa_{y\to y'}$ is a fixed positive constant. Hence, for a particular choice of $V > 0$, the intensity function for $y \to y' \in \mathcal{R}$ is

$$\lambda^V_{y\to y'}(x) = \frac{\kappa_{y\to y'}}{V^{\|y\|_1-1}} 1_{\{x\geq y\}} \frac{x!}{(x-y)!}, \qquad \text{for } x \in \mathbb{Z}^d_{\geq 0}.$$

We then denote the resulting stochastic process by X^V. Next, we consider the scaled process

$$\overline{X}^V(t) = V^{-1}X^V(t) \in V^{-1}\mathbb{Z}^d_{\geq 0}. \tag{4.18}$$

The associated transition intensities for the process $\overline{X}^V$ are

$$\lambda^{S,V}_{y\to y'}(x) = \lambda^V_{y\to y'}(Vx) = \frac{\kappa_{y\to y'}}{V^{\|y\|_1-1}} \frac{(Vx)!}{(Vx-y)!}, \quad x \in V^{-1}\mathbb{Z}^d_{\geq 0}, \tag{4.19}$$

and the generator is

$$(\mathcal{L}_V f)(x) = \sum_{y\to y'\in\mathcal{R}} \lambda^{S,V}_{y\to y'}(x) \left(f\left(x + \frac{y'-y}{V}\right) - f(x) \right), \quad x \in V^{-1}\mathbb{Z}^d_{\geq 0}. \tag{4.20}$$

Following [2, 1], we are interested in finding conditions for a reaction network to satisfy a large deviation principle (LDP). By standard arguments, we see that for a fixed $x \in \mathbb{R}^d_{>0}$ and V large

$$\lambda^{S,V}_{y\to y'}\left(\frac{\lfloor Vx\rfloor}{V}\right) = \frac{\kappa_{y\to y'}}{V^{\|y\|_1-1}}\frac{(\lfloor Vx\rfloor)!}{(\lfloor Vx\rfloor - y)!} \approx \frac{\kappa_{y\to y'}}{V^{\|y\|_1-1}}V^{\|y\|_1}x^y = V\kappa_{y\to y'}x^y.$$

Hence, we also define the analogous "deterministic" intensity function

$$\lambda^{D,V}_{y\to y'}(x) = V\kappa_{y\to y'}x^y, \qquad \text{for } x \in \mathbb{R}^d_{\geq 0}. \tag{4.21}$$

For completeness, we provide the following definition for a LDP in the setting of reaction networks.

Definition 4.16. *Fix a positive $T < \infty$ and a lower semi-continuous mapping $I : D_{0,T}(\mathbb{R}^d_{>0}) \to [0,\infty]$ such that for any $\alpha \in \mathbb{R}_{>0}$, the level set $\{z : I(z) \leq \alpha\}$ is a compact subset of $D_{0,T}(\mathbb{R}^d_{>0})$. The probability distribution of sample paths of the processes $\{\overline{X}^V\}_{V>0}$ with fixed initial condition $\overline{X}^V(0) = x \in \mathbb{R}^d_{>0}$ obeys a LDP with good rate function $I(\cdot)$ if for any measurable $\Gamma \subset D_{0,T}(\mathbb{R}^d_{>0})$ we have*

$$\begin{aligned} -\inf_{z\in\Gamma^o} I(z) &\leq \liminf_{V\to\infty}\frac{1}{V}\ln\left(P\left(\overline{X}^V(t)\in\Gamma \mid \overline{X}^V(0)=x\right)\right) \\ &\leq \limsup_{V\to\infty}\frac{1}{V}\ln\left(P\left(\overline{X}^V(t)\in\Gamma \mid \overline{X}^V(0)=x\right)\right) \leq -\inf_{z\in\bar{\Gamma}} I(z) \end{aligned}$$

where Γ^o and $\bar{\Gamma}$ denote the interior and closure of Γ respectively.

In [1], it is shown that under Assumption (2) below, the process $\overline{X}^V$ satisfies a sample path LDP in the supremum norm.

Assumption 2. *Let $\overline{X}^V$ be the process (4.18). We assume*

1. *There exists $b < \infty$ and a continuous, positive function $U(\cdot)$ with compact sublevel sets, such that for some non-decreasing function $v' : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$,*

$$(\mathcal{L}_V U^V)(x) \leq e^{bV} \qquad \forall V > v'(\|x\|_1), \qquad x \in V^{-1}\mathbb{Z}^d_{\geq 0} \tag{4.22}$$

where $U^V(\cdot)$ denotes the V^{th} power of $U(\cdot)$, and $\mathcal{L}_V$ is defined as in (4.20).

2. *With positive probability, starting at $\overline{X}^V(0) = 0$, the Markov process $\overline{X}^V$ reaches in finite time some state x_+ in the strictly positive orthant $V^{-1}\mathbb{Z}^d_{>0}$.*

Moreover, [1] and [2] show that Assumption 2 holds for reaction networks with a certain structure. We require the following definition before stating their result.

Definition 4.17. *A non-empty subset $\mathcal{P} \subset \mathcal{S} = \{S_1, \ldots, S_d\}$ is called a* siphon *if for every reaction $y \to y' \in \mathbb{R}$ the following condition holds: if $y'_i > 0$ for some $S_i \in \mathcal{P}$, then $y_j > 0$ for some $S_j \in \mathcal{P}$. A reaction network is called* asiphonic *if no such $\mathcal{P}$ exists.*

In words, $\mathcal{P}$ is a siphon if every reaction whose product complex contains an element of $\mathcal{P}$ also has an element of $\mathcal{P}$ in its source complex. Note that if a network is asiphonic, then $0 \in \mathcal{C}^S$ (the set of source complexes) for otherwise $\mathcal{S}$ would be a siphon.

Theorem 4.18. *If the network is asiphonic and strongly endotactic (ASE), then the Markov process $\overline{X}^V$ satisfies Assumption (2) with U defined as in (4.6) (which is the usual Lyapunov function) and the function $v'(x) = e^x$.*

Note that there is a simple argument showing that asiphonic reaction networks automatically satisfy the second part of Assumption 2 (see Remark 1.11 in [1]). It is significantly harder to show ASE reaction networks satisfy the first condition in Assumption 2. Here we will provide a proof showing that ASE reaction networks satisfy the first condition of Assumption 2, and will do so using a tier structure argument. Specifically, we will prove Theorem 4.19 below, which implies Theorem 4.18, and is the main result of this section.

Theorem 4.19. *Suppose the reaction network $(\mathcal{S}, \mathcal{C}, \mathcal{R})$ is ASE. Furthermore, let U be defined as in (4.6) and let $v'(x) = e^x$. Then there exists a compact set $B \subset \mathbb{R}^d$ such that for all pairs (V, x) satisfying $V > v'(\|x\|_1) = e^{\|x\|_1}$, $x \in V^{-1}\mathbb{Z}^d_{\geq 0}$, and $x \in B^c$, we have*

$$(\mathcal{L}_V U^V)(x) < 0. \tag{4.23}$$

Before getting to the proof of the Theorem, we need a preliminary technical result which we prove using the tier sequence technique.

Lemma 4.20. *Suppose that there is a sequence $(x_n, V_n)_{n=0}^{\infty}$ such that:*

- $(x_n)_{n=0}^{\infty}$ *is a tier sequence* (4.24)
- $\lim_{n\to\infty} \|x_n\|_1 = \infty$ (4.25)
- $V_n > e^{\|x_n\|_1}$ *and* $x_n \in V_n^{-1}\mathbb{Z}^d_{>0}$. (4.26)

Let $c_1 \in \mathbb{R}$ and $c_2 \in \mathbb{R}_{>0}$ and let

$$H(x_n, V_n) = \sum_{y\to y' \in \mathcal{R}} \kappa_{y\to y'} x_n^y U(x_n)\left(\exp\left(\frac{\ln(x_n^{y'-y}) + c_1}{c_2 U(x_n)}\right) - 1\right). \tag{4.27}$$

Then

$$\liminf_{n\to\infty} H(x_n, V_n) = -\infty. \tag{4.28}$$

Proof. Note that $U(x_n)$ grows like $\|x_n\|_1 \ln(\|x_n\|_1)$, as $n \to \infty$, which itself converges to ∞ by (4.25). Thus it must be that $\limsup_{n\to\infty} \frac{\ln(x_{n,i})}{U(x_n)} \leq 0$ for each $i \in \{1, \dots, d\}$. Let us consider the set of indices

$$E = \left\{ i : \liminf_{n\to\infty} \frac{\ln(x_{n,i})}{U(x_n)} < 0 \right\}.$$

The set E can be non-empty, and consists of the indices of those species which are relatively small. For example, we could have a two-dimensional system with $x_n = (e^{-n^2}, n)$ and $V_n = e^{n^2}$. In this case, $\ln(x_{n,1}) = -n^2$ whereas $U(x_n)$ grows like $n\ln(n)$ as $n \to \infty$. Thus, $\lim_{n\to\infty} \frac{\ln(x_{n,1})}{U(x_n)} = -\infty$ and $1 \in E$.

By potentially considering another subsequence, we may replace all the lim inf and lim sup by lim in the above. Using E, we can partition the set of reactions $\mathcal{R}$ into 3 mutually exclusive groups

1. $\mathcal{R}_1 = \{y \to y' : y_i \neq 0 \text{ for some } i \in E\}$.
2. $\mathcal{R}_2 = \{y \to y' : y_i = 0\ \forall i \in E \text{ and } y'_i \neq 0 \text{ for some } i \in E\}$.
3. $\mathcal{R}_3 = \{y \to y' : y_i = y'_i = 0\ \forall i \in E\}$.

Note that because the network is asiphonic, $0 \in \mathcal{C}^S$. Hence, $\mathcal{R}_1 \neq \mathcal{R}$. We then decompose H in the obvious manner as $H(x_n, V_n) = H_1(x_n, V_n) + H_2(x_n, V_n) + H_3(x_n, V_n)$, where

$$H_i(x_n, V_n) = \sum_{y\to y'\in\mathcal{R}_i} \kappa_{y\to y'} x_n^y U(x_n) \Bigg(\exp\left(\frac{\ln(x_n^{y'-y}) + c_1}{c_2 U(x_n)} \right) - 1 \Bigg).$$

We will show that (i) $\lim_{n\to\infty} H_1(x_n, V_n) = 0$, (ii) the terms in H_2 are negative, and (iii) the negative terms in H_2 and H_3 are sufficient to guarantee that (4.28) holds.

We turn to $H_1(x_n, V_n)$. First note that for $y \to y' \in \mathcal{R}_1$, we have that

$$\ln(x_n^{y'-y}) = \langle y', \ln(x_n)\rangle - \langle y, \ln(x_n)\rangle \le c_3 \sum_{i\in E} |\ln(x_{n,i})| = -c_3 \sum_{i\in E} \ln(x_{n,i}),$$

for some positive constant c_3. Hence, there is a $c_4 > 0$ so that for n large enough

$$
\begin{aligned}
x_n^y U(x_n) \exp\left(\frac{\ln(x_n^{y'-y}) + c_1}{c_2 U(x_n)}\right) &\leq x_n^y U(x_n) \exp\left(-\frac{c_4 \sum_{i\in E} \ln(x_{n,i})}{U(x_n)}\right) \\
&= \exp\left(\sum_{i=1}^{d} y_i \ln(x_{n,i}) + \ln(U(x_n)) - \frac{\sum_{i\in E} c_4 \ln(x_{n,i})}{U(x_n)}\right) \\
&= \exp\left(\sum_{i\in E} \ln(x_{n,i})\left(y_i - \frac{c_4}{U(x_n)}\right) + \sum_{j\notin E} y_j \ln(x_{n,j}) + \ln(U(x_n))\right).
\end{aligned}
\tag{4.29}
$$

Note that from the construction of E, for $i \in E$ and $j \notin E$, we must have $|\ln(x_{n,i})| \gg \ln(U(x_n))$ and $|\ln(x_{n,i})| \gg \ln(x_{n,j})$. Since $y_i \geq 1$ for some $i \in E$, we must have

$$
\lim_{n\to\infty} \sum_{i\in E} \ln(x_{n,i})\left(y_i - \frac{c_4}{U(x_n)}\right) + \sum_{j\notin E} y_j \ln(x_{n,j}) + \ln(U(x_n)) = -\infty.
$$

Moreover, by a similar argument we see that for $y \to y' \in \mathcal{R}_1$

$$
\lim_{n\to\infty} x_n^y U(x_n) = \lim_{n\to\infty} \exp\left(\sum_{i\in E} y_i \ln(x_{n,i}) + \sum_{j\notin E} y_j \ln(x_{n,j}) + \ln(U(x_n))\right) = 0. \tag{4.30}
$$

Thus for each $y \to y' \in \mathcal{R}_1$

$$
\lim_{n\to\infty} x_n^y U(x_n)\left(\exp\left(\frac{\ln(x_n^{y'-y}) + c_1}{c_2 U(x_n)}\right) - 1\right) = 0
$$

and $\lim_{n\to\infty} H_1(x_n, V_n) = 0$.

Next, we consider $H_2(x_n, V_n)$. Let $y \to y' \in \mathcal{R}_2$. We know that $y_j = 0$ for all $j \in E$ and that there exist an $i \in E$ with $y_i' > 0$. Hence, using that $\lim_{n\to\infty} U(x_n) = \infty$ and the definition of E, we have

$$
\begin{aligned}
\exp\left(\frac{\ln(x_n^{y'-y}) + c_1}{c_2 U(x_n)}\right) - 1 &= \exp\left(\frac{\sum_{i\in E} y_i' \ln(x_{n,i}) + \sum_{j\notin E}(y_j' - y_j)\ln(x_{n,j}) + c_1}{c_2 U(x_n)}\right) - 1 \\
&< e^{-c_5} - 1 < -c_6 < 0
\end{aligned}
$$

for some positive constants c_5 and c_6 and n large enough. Thus

$$H_2(x_n, V_n) < -c_6 \sum_{y \to y' \in \mathcal{R}_2} \kappa_{y \to y'} x_n^y U(x_n). \tag{4.31}$$

We turn to $H_3(x_n, V_n)$. Let $y \to y' \in \mathcal{R}_3$. Since $y_i = y_i' = 0$ for all $i \in E$, we have by the definition of E that

$$\lim_{n \to \infty} \frac{\ln(x_n^{y'-y}) + c_1}{c_2 U(x_n)} = 0.$$

Note that we can choose a subsequence for which each term on the left above is either non-negative or non-positive for each n and each $y \to y' \in \mathcal{R}_3$. If the terms are non-positive, we may use that $e^\rho - 1 \leq \frac{1}{2}\rho$ for small $\rho \leq 0$ to conclude that

$$\kappa_{y \to y'} x_n^y U(x_n) \left(\exp\left(\frac{\ln(x_n^{y'-y}) + c_1}{c_2 U(x_n)} \right) - 1 \right) \leq \frac{1}{2c_2} \kappa_{y \to y'} x_n^y (\ln(x_n^{y'-y}) + c_1). \tag{4.32}$$

Moreover, if the terms are non-negative, we use that $e^\rho - 1 \leq 2\rho$ for small $\rho \geq 0$ to conclude that

$$\kappa_{y \to y'} x_n^y U(x_n) \left(\exp\left(\frac{\ln(x_n^{y'-y}) + c_1}{c_2 U(x_n)} \right) - 1 \right) \leq \frac{2}{c_2} \kappa_{y \to y'} x_n^y (\ln(x_n^{y'-y}) + c_1). \tag{4.33}$$

Thus, there are positive constants $c_{y \to y'}$ for which

$$H_3(x_n, V_n) \leq \sum_{y \to y' \in \mathcal{R}_3} c_{y \to y'} \kappa_{y \to y'} x_n^y (\ln(x_n^{y'-y}) + c_1). \tag{4.34}$$

Finally, we return to $H(x_n, V_n) = H_1(x_n, V_n) + H_2(x_n, V_n) + H_3(x_n, V_n)$. To conclude that (4.28) holds, it is now sufficient to show two things. First, we will prove that there is always a term in either (4.31) or (4.34) (i.e., terms associated with reactions in $\mathcal{R}_2$ or $\mathcal{R}_3$) that goes to $-\infty$, as $n \to \infty$. Second, we will prove that any positive term in the sum (4.27) is dominated, in the sense of Lemma 4.15, by a negative term.

Since the network is asiphonic, there must be a reaction for which 0 is the source complex. By definition of $T^{1,S}$ we have $0 \precsim_{(x_n)} y$ for all $y \in T^{1,S}$, which implies that for all $y \in T^{1,S}$

$$\lim_{n\to\infty} x_n^y > 0. \tag{4.35}$$

Since the network is strongly endotactic it must be tier descending by Theorem 4.10. Hence there exists a reaction $y \to y' \in \mathbb{R}$ with $y \in T^{1,S}$ and $y' \prec_{(x_n)} y$. Recall that (4.30) showed that $x_n^y U(x_n) \to 0$, as $n \to \infty$, if $y \to y' \in \mathcal{R}_1$. Hence, (4.35) shows that $y \to y' \notin \mathcal{R}_1$. If $y \to y' \in \mathcal{R}_2$, we consider the relevant term in (4.31) and conclude

$$\lim_{n\to\infty} -c_6 \kappa_{y\to y'} x_n^y U(x_n) = -\infty$$

due to the fact that $\lim_{n\to\infty} U(x_n) = \infty$. Finally, if $y \to y' \in \mathcal{R}_3$, we have

$$\lim_{n\to\infty} c_{y\to y'} \kappa_{y\to y'} x_n^y (\ln(x_n^{y'-y}) + c_1) = -\infty$$

since $y' \prec_{(x_n)} y$. Thus, in either case, we have a term which converges to $-\infty$ as $n \to \infty$.

Next, we will show that a positive term is necessarily dominated by a negative term. Specifically, note that the only terms that could be positive and not tend to zero come from the sum (4.34) and are associated with reactions $y \to y' \in \mathcal{R}_3$ with $y \precsim_{(x_n)} y'$. Fix such a reaction $y \to y' \in \mathcal{R}_3$. We will now show that there is necessarily a term either in the sum (4.31) or the sum (4.34) that is negative and dominates it.

Suppose first that there is a reaction $\tilde{y} \to \tilde{y}' \in \mathcal{R}_2$ for which $y \precsim_{(x_n)} \tilde{y}$. Because $y \to y' \in \mathcal{R}_3$, we know

$$U(x_n) \gg \ln(x_n^{y'-y}).$$

Hence, the term in (4.31) associated with $\tilde{y} \to \tilde{y}'$ dominates the positive term.

Now assume there is no such reaction $\tilde{y} \to \tilde{y}' \in \mathcal{R}_2$ with $y \precsim_{(x_n)} \tilde{y}$. Because our network is strongly endotactic, we may apply Lemma 4.15 to conclude that there exists $y^\star \in \mathcal{C}$ and $y^\star \to y^{\star\star} \in \mathbb{R}$ such that $y \precsim_{(x_n)} y^\star$, $y^{\star\star} \prec_{(x_n)} y^\star$ and for any choice of constants $c_1', c_2' \in \mathbb{R}_{>0}$ and $c_3', c_4' \in \mathbb{R}$, the inequality (4.7) holds for n large enough. Thus, if we can show that $y^\star \to y^{\star\star} \in \mathcal{R}_3$, then the term in (4.34) associated with $y^\star \to y^{\star\star}$ dominates the positive term.

Since $y \precsim_{(x_n)} y^\star$, we know from our assumption that $y^\star \to y^{\star\star} \notin \mathcal{R}_2$. Moreover, since $y \precsim_{(x_n)} y^\star$, the reaction $y^\star \to y^{\star\star}$ cannot be in $\mathcal{R}_1$ (for otherwise the definition of E and the fact that $y \to y' \in \mathcal{R}_3$ would imply $y^\star \ln(x_n) - y \ln(x_n) \to -\infty$, as $n \to \infty$). Thus, we must have $y^\star \to y^{\star\star} \in \mathcal{R}_3$, and this concludes the proof of the Lemma 4.20. $\square$

We now turn to the proof of Theorem 4.18

Proof of Theorem 4.18. We will prove the theorem by contradiction. We therefore suppose that there is a sequence $(x_n, V_n)_{n=0}^\infty$ such that:

- $$\lim_{n\to\infty} \|x_n\|_1 = \infty \tag{4.36}$$
- $$V_n > e^{\|x_n\|_1} \quad \text{and} \quad x_n \in V_n^{-1}\mathbb{Z}^d_{\geq 0} \tag{4.37}$$
- $$(\mathcal{L}_{V_n} U^{V_n})(x_n) \geq 0. \tag{4.38}$$

Note that, after potentially considering a subsequence, we may assume the following

(i) $(x_n)_{n=0}^\infty$ is a tier sequence (this follows from Remark 4.4),

(ii) there is an $\ell \in \{0, \ldots, d\}$ for which $x_{n,1} = \cdots = x_{n,\ell} = 0$ and $x_{n,j} > 0$ for all $j \geq \ell + 1$ and all n (note that ℓ can be zero), and

(iii) there is a subset of the reactions, $\mathcal{P} \subseteq \mathcal{R}$, for which

$$\lambda^{S,V_n}_{y\to y'}(x_n) \begin{cases} > 0 & \text{if } y \to y' \in \mathcal{P} \\ = 0 & \text{if } y \to y' \in \mathcal{R} \setminus \mathcal{P} \end{cases} \tag{4.39}$$

for every n.

(iv) the sign of the terms $U^{V_n}(x_n) - U^{V_n}(x_n + \frac{y'-y}{V_n})$ are constant in n, for each $y \to y' \in \mathcal{P}$.

We will prove that $\liminf_{n\to\infty}(\mathcal{L}_{V_n}U^{V_n})(x_n) = -\infty$, leading to a contradiction.

First, note that for any reaction $y \to y' \in \mathcal{P}$ we have

$$\lambda^{S,V_n}_{y\to y'}(x_n) = V_n\kappa_{y\to y'}\prod_{i=1}^{d} x_{n,i}\left(x_{n,i} - \frac{1}{V_n}\right)\cdots\left(x_{n,i} - \frac{y_i - 1}{V_n}\right),$$

which is positive by assumption. Hence, $x_{n,i} \geq \frac{y_i}{V_n}$. Thus, for any $1 \leq j \leq y_i - 1$,

$$x_{n,i} - \frac{j}{V_n} = x_{n,i} - \frac{j}{y_i}\frac{y_i}{V_n} \geq x_{n,i}\left(1 - \frac{j}{y_i}\right).$$

Thus, letting $c_y = \prod_{i=1}^{d}\prod_{j=1}^{y_i-1}\left(1 - \frac{j}{y_i}\right) > 0$, we have

$$V_n\kappa_{y\to y'}x_n^y \geq \lambda^{S,V_n}_{y\to y'}(x_n) \geq c_y V_n \kappa_{y\to y'}x_n^y. \tag{4.40}$$

Combining (4.40) with the fact that the signs of the terms $U^{V_n}(x_n) - U^{V_n}(x_n + \frac{y'-y}{V_n})$ are constant over n, we may conclude that

$$(\mathcal{L}_{V_n}U^{V_n})(x_n) \leq \sum_{y\to y'\in\mathcal{P}} V_n\tilde{\kappa}_{y\to y'}x_n^y\left(U^{V_n}\left(x + \frac{y'-y}{V_n}\right) - U^{V_n}(x)\right) \tag{4.41}$$

for all n and for some positive constants $\tilde{\kappa}_{y\to y'}$, with $y \to y' \in \mathcal{P}$. For notational convenience, we define the operator

$$(\widetilde{\mathcal{L}}_V f)(x) = \sum_{y\to y'\in\mathcal{P}} V_n\tilde{\kappa}_{y\to y'}x_n^y\left(f\left(x + \frac{y'-y}{V}\right) - f(x)\right), \quad x \in V^{-1}\mathbb{Z}^d_{\geq 0},$$

and we point out that this operator is similar to the generator of the process $\overline{X}^V$ for the modified reaction rates $\tilde{\kappa}_{y\to y'}$. In fact, we are simply exchanging the stochastic intensities for the "deterministic" intensities for the reactions in $\mathcal{P}$. By (4.41), it suffices to show that

$$\liminf_{n\to\infty}(\widetilde{\mathcal{L}}_{V_n}U^{V_n})(x_n) = -\infty. \tag{4.42}$$

We consider the terms of $(\widetilde{\mathcal{L}}_{V_n}U^{V_n})(x_n)$ individually. Let $y \to y' \in \mathcal{P}$ and note that we must have $y_i = 0$ for each $i \leq \ell$. Let

$$C_{y\to y'}(V_n) = \sum_{i=1}^{\ell} y'_i \left(\ln\left(\frac{y'_i}{V_n}\right) - 1\right). \tag{4.43}$$

Note that $|C_{y\to y'}(V_n)|$ grows at most logarithmically in V_n, as $n \to \infty$. Utilizing a Taylor expansion of the logarithm yields

$$U\left(x_n + \frac{y'-y}{V_n}\right) = d + 1 + V_n^{-1}C_{y\to y'}(V_n) + \sum_{i=\ell+1}^{d}\left(x_{n,i} + \frac{y'_i - y_i}{V_n}\right)\left(\ln\left(x_{n,i} + \frac{y'_i - y_i}{V_n}\right) - 1\right)$$

$$= d + 1 + V_n^{-1}C_{y\to y'}(V_n) + \sum_{i=\ell+1}^{d}\left(x_{n,i} + \frac{y'_i - y_i}{V_n}\right)\left(\ln(x_{n,i}) + \frac{y'_i - y_i}{x_{n,i}V_n} + r_i(x_{n,i}, V_n) - 1\right)$$

$$= U(x_n) +$$

$$\frac{1}{V_n}\left(C_{y\to y'}(V_n) + \sum_{i=\ell+1}^{d}(y'_i - y_i)\ln(x_{n,i}) + \sum_{i=\ell+1}^{d}\left(\frac{(y'_i - y_i)^2}{x_{n,i}V_n} + (x_{n,i}V_n + y'_i - y_i)r_i(x_{n,i}, V_n)\right)\right),$$

where

$$|r_i(x_{n,i}, V_n)| \leq \frac{c_1}{x_{n,i}^2 V_n^2},$$

for some $c_1 > 0$. We denote

$$R_i(x_{n,i}, V_n) = \frac{(y'_i - y_i)^2}{x_{n,i}V_n} + (x_{n,i}V_n + y'_i - y_i)r_i(x_{n,i}, V_n).$$

We have $x_{n,i}V_n \geq 1$ for all $i \geq \ell+1$, thus

$$|R_i(x_{n,i}, V_n)| \leq \frac{(y_i' - y_i)^2}{x_{n,i}V_n} + \frac{c_1}{x_{n,i}V_n} + \frac{c_1|y_i' - y_i|}{x_{n,i}^2 V_n^2} \leq \frac{c_2}{x_{n,i}V_n} \leq c_2, \tag{4.44}$$

for some positive constant c_2. Combining the above, and utilizing the inequality

$$(1+\varepsilon)^n \leq e^{\varepsilon n},$$

which holds for all integers n when $|\varepsilon| < 1$, it follows that for n large enough

$$\begin{aligned}
&(\widetilde{\mathcal{L}}_{V_n} U^{V_n})(x_n) \\
&= \sum_{y \to y' \in \mathcal{P}} V_n \tilde{\kappa}_{y \to y'} x_n^y U(x_n)^{V_n} \\
&\left(\left(1 + \frac{1}{V_n} \frac{C_{y \to y'}(V_n) + \sum_{i=\ell+1}^d (y_i' - y_i) \ln(x_{n,i}) + \sum_{i=\ell+1}^d R_i(x_{n,i}, V_n)}{U(x_n)} \right)^{V_n} - 1 \right) \\
&\leq V_n U(x_n)^{V_n - 1} H_{\mathcal{P}}(x_n, V_n)
\end{aligned} \tag{4.45}$$

where

$$H_{\mathcal{P}}(x_n, V_n) =$$

$$\sum_{y \to y' \in \mathcal{P}} \tilde{\kappa}_{y \to y'} x_n^y U(x_n) \left(\exp\left(\frac{C_{y \to y'}(V_n) + \sum_{i=\ell+1}^d (y_i' - y_i) \ln(x_{n,i}) + \sum_{i=\ell+1}^d R_i(x_{n,i}, V_n)}{U(x_n)} \right) - 1 \right).$$

In order to justify the inequality above, we use that (i) $\lim_{n\to\infty} U(x_n) = \infty$, (ii) the terms $R_i(x_{n,i}, V_n)$ are uniformly bounded by (4.44), and (iii) $\ln(x_n^{y'-y})$ is at most of order $\ln(V_n)$ because of (4.37) and since $x_{n,i} \geq V_n^{-1}$ for $i \geq \ell+1$.

We will now show that $\liminf_{n\to\infty} H_{\mathcal{P}}(x_n, V_n) = -\infty$. To do so, we consider a new sequence $\tilde{x}_n$, where

$$\tilde{x}_{n,1} = \cdots = \tilde{x}_{n,\ell} = \frac{\alpha}{V_n} \tag{4.46}$$

with $\alpha = \max_{z \in \mathcal{C}, i \in \{1,\ldots,d\}} z_i$, and

$$\tilde{x}_{n,i} = x_{n,i} \quad \text{for} \quad i > \ell.$$

Because of (4.46) and since u defined in (4.5) is a decreasing function in a positive neighborhood of zero, we have that $U(\tilde{x}_n) < U(x_n)$ for all n. Also, since $\lim_{n\to\infty} \tilde{x}_{n,i} = 0$ for $i \le \ell$, we have $\lim_{n\to\infty} \frac{U(\tilde{x}_n)}{U(x_n)} = 1$. Recalling that $y \to y' \in \mathcal{P}$ implies $y_i = 0$ for $i \le \ell$, we have

$$x_n^y = \tilde{x}_n^y. \tag{4.47}$$

From (4.43), and because in (4.46) we chose $\alpha \ge y_i'$ for all i,

$$C_{y\to y'}(V_n) < \sum_{i=1}^{\ell} y_i' \ln(\tilde{x}_{n,i}). \tag{4.48}$$

Combining (4.48), $\lim_{n\to\infty} \frac{U(\tilde{x}_n)}{U(x_n)} = 1$, and the bound on R_i, we may conclude there exists $c_3 \in \mathbb{R}$ and $c_4 \in \mathbb{R}_{>0}$ such that

$$\frac{C_{y\to y'}(V_n) + \sum_{i=\ell+1}^{d}(y_i' - y_i)\ln(x_{n,i}) + \sum_{i=\ell+1}^{d} R_i(x_{n,i}, V_n)}{U(x_n)} < \frac{\ln(\tilde{x}_n^{y'-y}) + c_3}{U(x_n)} < \frac{\ln(\tilde{x}_n^{y'-y}) + c_3}{c_4 U(\tilde{x}_n)}$$

for n large enough. Therefore, utilizing (4.47) and the above yields

$$H_{\mathcal{P}}(x_n, V_n) < \frac{U(x_n)}{U(\tilde{x}_n)} \sum_{y\to y'\in\mathcal{P}} \tilde{\kappa}_{y\to y'} \tilde{x}_n^y U(\tilde{x}_n)\Bigg(\exp\left(\frac{\ln(\tilde{x}_n^{y'-y}) + c_3}{c_4 U(\tilde{x}_n)}\right) - 1\Bigg). \tag{4.49}$$

By Lemma 4.20 we have

$$\liminf_{n\to\infty} \sum_{y\to y'\in\mathcal{R}} \tilde{\kappa}_{y\to y'} \tilde{x}_n^y U(\tilde{x}_n)\Bigg(\exp\left(\frac{\ln(\tilde{x}_n^{y'-y}) + c_3}{c_4 U(\tilde{x}_n)}\right) - 1\Bigg) = -\infty. \tag{4.50}$$

Therefore, in order to conclude that $\liminf_{n\to\infty} H_{\mathcal{P}}(x_n, V_n) = -\infty$, it is sufficient to show that

$$\lim_{n\to\infty} \sum_{y\to y'\in\mathcal{R}\setminus\mathcal{P}} \tilde{\kappa}_{y\to y'} \tilde{x}_n^y U(\tilde{x}_n)\Bigg(\exp\left(\frac{\ln(\tilde{x}_n^{y'-y}) + c_3}{c_4 U(\tilde{x}_n)}\right) - 1\Bigg) = 0. \tag{4.51}$$

Let $y \to y' \in \mathcal{R} \setminus \mathcal{P}$. At least one of the following must be true

1. there is a k with $k > \ell$ such that $y_k > 0$ and $x_{n,k} < \frac{y_k}{V_n}$. In this case we also have $\tilde{x}_{n,k} = x_{n,k} < \frac{y_k}{V_n}$.

2. there is a k with $k \leq \ell$ such that $y_k > 0$. In this case we have $\tilde{x}_{n,k} = \frac{\alpha}{V_n}$.

In either case we have $\frac{1}{V_n} \leq \tilde{x}_{n,k} \leq \frac{\alpha}{V_n}$. Using this, together with the fact that $\ln(\|x_n\|_1) < \ln(\ln(V_n))$, implies there is a $c_5 > 0$ for which

$$\exp\left(\frac{\ln(\tilde{x}_n^{y'-y}) + c_3}{c_4 U(\tilde{x}_n)}\right) \leq \exp\left(\frac{c_5 \ln V_n}{U(\tilde{x}_n)}\right) = V_n^{c_5/U(\tilde{x}_n)}.$$

Thus

$$\begin{aligned}
&\left|\tilde{x}_n^y U(\tilde{x}_n)\Bigg(\exp\left(\frac{\ln(\tilde{x}_n^{y'-y}) + c_3}{c_4 U(\tilde{x}_n)}\right) - 1\Bigg)\right| \\
&\qquad \leq U(\tilde{x}_n)\Bigg(\prod_{i \neq k} \tilde{x}_{n,i}^{y_i}\Bigg)\frac{\alpha^{y_k}}{V_n^{y_k}} V_n^{c/U(\tilde{x}_n)} + U(\tilde{x}_n)\Bigg(\prod_{i \neq k} \tilde{x}_{n,i}^{y_i}\Bigg)\frac{\alpha^{y_k}}{V_n^{y_k}} \qquad (4.52) \\
&\qquad = U(\tilde{x}_n)\Bigg(\prod_{i \neq k} \tilde{x}_{n,i}^{y_i}\Bigg)\frac{\alpha^{y_k}}{V_n^{y_k - c/U(\tilde{x}_n)}} + U(\tilde{x}_n)\Bigg(\prod_{i \neq k} \tilde{x}_{n,i}^{y_i}\Bigg)\frac{\alpha^{y_k}}{V_n^{y_k}}.
\end{aligned}$$

Since $V_n \geq e^{\|\tilde{x}_n\|_1}$ and $U(\tilde{x}_n)$ grows like $\|\tilde{x}_n\|_1 \ln \|\tilde{x}_n\|_1$, as $n \to \infty$, both terms go to 0, showing (4.51). Combining (4.45), (4.49), (4.50), and (4.51), allows us to conclude that (4.42) holds. Thus, the proof of the theorem is complete. $\square$

Chapter 5

Deficiency zero reaction networks and their prevalence

As introduced in Chapter 2, deficiency zero and weak reversibility ensure that a network is complex balanced for any rate constants, which in turns leads to many stability properties of the network in both the deterministic and the stochastic models. Given the significant role of deficiency zero in reaction network theory, one may ask: are such networks common? The earliest attempt to answer this question can be traced back to some work by Horn in 1973 [27]. In that paper, Horn considered all reaction networks with exactly 3 binary complexes, but no condition on the number of species. Horn found 43 isomorphism classes of such networks, and among these, 41 have deficiency zero.

We choose a different tack by considering networks with a fixed number of species, say n, and then quantifying the prevalence of the deficiency zero property via limit theorems (as $n \to \infty$) in two random graph frameworks that we utilize to generate random reaction networks. In the first two sections, we consider random reaction networks generated via an Erdős-Rényi framework, while in the next two sections, we consider random reaction networks generated via a stochastic block model.

However, we are immediately confronted with a modeling problem: for any finite number of species there are an infinite number of possible graphs that can be constructed from them. For example, with just the single species S_1, possible vertices include $S_1, 2S_1, 3S_1, \ldots$. Hence, we must restrict ourselves in some manner so that for a given number of species, only a finite number of vertices are possible. In this chapter, we restrict ourselves to study binary reaction networks (see Definition 2.7), which are by far the most common in the literature.

5.1 An Erdős-Rényi framework for random reaction networks

In this section we will set up an Erdős-Rényi framework for generating random reaction networks. Let the set of species be $\mathcal{S} = \{S_1, S_2, \ldots, S_n\}$. We consider binary reaction networks with species in $\mathcal{S}$. The set of all possible vertices is then

$$\mathcal{C}_n^0 = \{\emptyset, S_i, S_i + S_j : \text{for } 1 \leq i \leq n \text{ and } 1 \leq j \leq n.\}$$

For a given n, we denote $N_n = |\mathcal{C}_n^0|$, the cardinality of $\mathcal{C}_n^0$. Thus, N_n is the total number of possible unary, binary, and zeroth order vertices that can be generated from n distinct species. A straightforward calculation gives

$$N_n = 1 + n + n + \frac{n(n-1)}{2} = \frac{n^2 + 3n + 2}{2},$$

and so

$$n \sim \sqrt{2N_n}.$$

Here we use the notation $\sim$ in the standard way: for any two sequences of real numbers $\{a_n\}$ and $\{b_n\}$, we write $a_n \sim b_n$ if $\lim_{n\to\infty} \frac{a_n}{b_n} = c$ for some constant $c \in \mathbb{R}$.

We consider an Erdős-Rényi random graph $G(N_n, p_n)$, which we will simply denote G_n throughout, where the set of vertices is the set of all possible binary vertices $\mathcal{C}_n^0$, and the probability that there is an edge between any 2 particular vertices is p_n, independently of all other edges. Each such random graph now corresponds to an associated graph from a reaction network in the following way,

1. each vertex with positive degree in the random graph represents a vertex in the reaction network graph, and
2. each edge in the random graph represents a reaction in the reaction network graph (we can assume all reactions are reversible, i.e., that $y \to y' \in \mathcal{R} \implies y' \to y \in \mathcal{R}$, since we do not need to worry about direction–see Lemma 2.13(a)).

We will denote the reaction network associated with the graph $G(N_n, p_n)$ by $R(N_n, p_n)$, which we will often simplify to R_n. We will denote the deficiency of R_n by δ_{R_n}.

Next, in order to build intuition for the calculations to come, we provide two simple examples when the number of species is small, and thus we are able to explicitly compute the probability that the reaction network associated with the randomly generated network has a deficiency of zero.

Example 7 (The case with $n = 1$ species)**.** Denote the only species by A. The set of vertices, or equivalently the set of all possible complexes, is $\mathcal{C}_1^0 = \{\emptyset, A, 2A\}$. Figure 5 shows one possible realization of the random graph when $p \in (0, 1)$. The corresponding reaction network for the particular graph shown in Figure 5 is $\emptyset \leftrightarrows A \leftrightarrows 2A$.

Returning to the general case when $n = 1$, let $R_1 = \{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ be the reaction network corresponding with the random graph, and let ℓ and s be defined as usual. Recall that

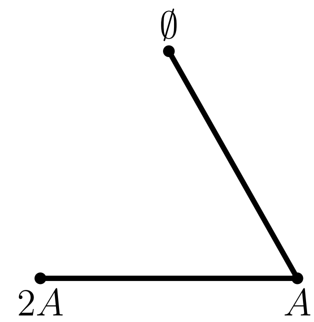

Figure 5: A realization of a random graph when $n = 1$ and $p \in (0, 1)$.

the deficiency is given by $\delta_{R_1} = |\mathcal{C}| - \ell - s$. Since $|\mathcal{C}| \in \{0, 2, 3\}$, there are three cases to consider.

- If $|\mathcal{C}| = 0$, then the reaction network is the empty network (recall Definition 2.1) and has a deficiency of zero.
- If $|\mathcal{C}| = 2$, then $\ell = 1$ and $s = 1$, and the deficiency is zero.
- If $|\mathcal{C}| = 3$, then $\ell = 1$ and $s = 2$ and the deficiency is one.

Since having $|\mathcal{C}| = 2$ corresponds to the case of having precisely one edge,

$$P(\delta_{R_1} = 0) = P(|\mathcal{C}| = 0) + P(|\mathcal{C}| = 2) = (1 - p)^3 + 3p(1 - p)^2.$$

Example 8 (The case with $n = 2$ species)**.** Denote the set of species by $\mathcal{S} = \{A, B\}$. The set of vertices is $\mathcal{C}_2^0 = \{\emptyset, A, B, 2A, 2B, A + B\}$. Figure 6 illustrates a possible realization of the random graph when $p \in (0, 1)$. The corresponding reaction network for the particular graph shown in Figure 6 is

$$\emptyset \leftrightarrows 2B$$

$$B \leftrightarrows A + B.$$

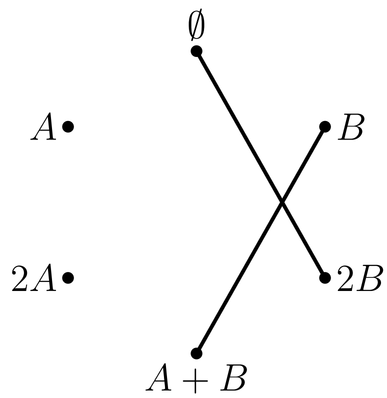

Figure 6: A realization of a random graph when $n = 2$ and $p \in (0, 1)$.

Returning to the general case when $n = 2$, we again let $R_2 = \{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ be the reaction network corresponding with the random graph, with ℓ and s defined as usual, and $\delta_{R_2} = |\mathcal{C}| - \ell - s$. Since $s \in \{0, 1, 2\}$, there are three cases to consider.

- Case 1: $\delta_{R_2} = 0$ and $s = 0$. In this case, the reaction network is the empty network and its deficiency is zero. Note that

$$\mathbb{P}(\delta_{R_2} = 0, s = 0) = \mathbb{P}(s = 0) = \mathbb{P}(\text{no edges}) = (1 - p)^{15},$$

since we have a total of $\binom{6}{2} = 15$ possible edges.

- Case 2: $\delta_{R_2} = 0$ and $s = 1$. In this case, we must have $|\mathcal{C}| = \ell + 1$, and by Remark 2.3 we have $\ell \leq \frac{|\mathcal{C}|}{2}$. Thus, we may conclude that $|\mathcal{C}| \leq 2$. As the network cannot be empty with $s = 1$, we have $|\mathcal{C}| = 2$. As in the previous example, this corresponds to a graph with only one edge. Thus

$$\mathbb{P}(\delta_{R_2} = 0, s = 1) = 15p(1 - p)^{14}.$$

- Case 3: $\delta_{R_2} = 0$ and $s = 2$. In this case we have $|\mathcal{C}| = \ell + 2$ and, again by Remark 2.3, $\ell \leq \frac{|\mathcal{C}|}{2}$. Combining these two facts yields $|\mathcal{C}| \leq 4$. In addition, the fact that

$s = 2$ ensures $|\mathcal{C}| \geq 3$. If $|\mathcal{C}| = 3$, then $\ell = 1$ and the corresponding graph must have either 2 or 3 edges. If $|\mathcal{C}| = 4$, then $\ell = 2$ and the corresponding graph must have 2 edges. Thus

$$\mathbb{P}(\delta_{R_2} = 0, s = 2) = \mathbb{P}(\delta_{R_2} = 0, s = 2, 2 \text{ edges}) + \mathbb{P}(\delta_{R_2} = 0, s = 2, 3 \text{ edges}).$$

We handle each term separately.

Suppose there are 2 edges in the graph. There are $\binom{15}{2}$ such configurations. Among these, the associated reaction network has a positive deficiency if and only if the 2 reaction vectors are linearly dependent. This can only happen if the 2 edges are from one of the three groups: $\{\emptyset \leftrightarrows A, A \leftrightarrows 2A, \emptyset \leftrightarrows 2A, B \leftrightarrows A + B\}$, $\{\emptyset \leftrightarrows B, B \leftrightarrows 2B, \emptyset \leftrightarrows 2B, A \leftrightarrows A + B\}$, and $\{A \leftrightarrows B, 2A \leftrightarrows 2B, A + B \leftrightarrows 2A, A + B \leftrightarrows 2B\}$. Excluding the configurations with positive deficiency, we have

$$\mathbb{P}(\delta_{R_2} = 0, s = 2, 2 \text{ edges}) = p^2(1-p)^{13}\left(\binom{15}{2} - 3\binom{4}{2}\right).$$

Suppose there are 3 edges in the graph. As argued above, we must additionally have $|\mathcal{C}| = 3$ and $\ell = 1$, which implies the graph only contains a single triangle formed by 3 vertices. There are $\binom{6}{3}$ such configurations. Among these, the associated reaction network has a positive deficiency if and only if all three reaction vectors span only 1 dimension. Therefore, the only 3 configurations with positive deficiency in this case are $\{\emptyset \leftrightarrows A, A \leftrightarrows 2A, \emptyset \leftrightarrows 2A\}$, $\{\emptyset \leftrightarrows B, B \leftrightarrows 2B, \emptyset \leftrightarrows 2B\}$, and $\{A+B \leftrightarrows 2A, A+B \leftrightarrows 2B, 2A \leftrightarrows 2B\}$. Excluding the configurations with positive deficiency, we have

$$\mathbb{P}(\delta_{R_2} = 0, s = 2, 3 \text{ edges}) = p^3(1-p)^{12}\left(\binom{6}{3} - 3\right).$$

Combining the three cases above yields

$$\mathbb{P}(\delta_{R_2}=0)=(1-p)^{15}+15p(1-p)^{14}+p^2(1-p)^{13}\left(\binom{15}{2}-3\binom{4}{2}\right)+p^3(1-p)^{12}\left(\binom{6}{3}-3\right).$$

As implied by the two previous examples, the computation of $P(\delta_{R_n}=0)$ gets more complicated as more species are added to the model. As a result of this fact, when we let the number of species go to infinity, it is more practical to consider the two extremes: when the probability of being deficiency zero converges to 0 and when it converges to 1. In particular, we want to find a threshold function $r(n)$ such that

$$\lim_{n\to\infty}\mathbb{P}(\delta_{R_n}=0)=\begin{cases}0, & \text{if } \lim_{n\to\infty}\frac{p_n}{r(n)}=\infty\\ 1 & \text{if } \lim_{n\to\infty}\frac{p_n}{r(n)}=0.\end{cases} \tag{5.1}$$

In the next section, we show $r(n)=\frac{1}{n^3}$.

5.2 Prevalence of deficiency zero reaction networks under an Erdős-Rényi framework

In Sections 5.2.1 and 5.2.2, we will show that the limits in (5.1) hold. Throughout this section, we will make use of the standard notation of $a_n \ll b_n$ or $b_n \gg a_n$ to mean $\lim_{n\to\infty}\frac{a_n}{b_n}=0$, whenever $\{a_n\}$ and $\{b_n\}$ are sequences of non-negative real numbers. We also remind the reader that we write $a_n \sim b_n$ to mean $\lim_{n\to\infty}\frac{a_n}{b_n}=c$ for some constant $c\in\mathbb{R}_{>0}$.

5.2.1 The case $\lim_{n\to\infty}\frac{p_n}{r(n)}=\infty$

Lemma 2.13(e) provides an upper bound on $|\mathcal{C}|$ for deficiency zero reaction network. We will utilize this bound to show $\lim_{n\to\infty}\mathbb{P}(\delta_{R_n}=0)=0$ when $\lim_{n\to\infty}\frac{p_n}{r(n)}=\infty$. Since

$|\mathcal{C}|$ is the number of non-isolated vertices in G_n, we start with a lemma regarding the number of isolated vertices in G_n.

Lemma 5.1. *Suppose* $p_n = \frac{2n+\alpha_n}{N_n(N_n-1)}$ *with* $\alpha_n \gg n^{1/2}$. *Let* I *be the set of isolated vertices in* G_n, *that is* $I = \{v \in \mathcal{C}_n^0 : deg(v) = 0\}$. *Then we have*

$$\lim_{n\to\infty} \mathbb{P}(|I| \geq N_n - 2n) = 0.$$

Proof. We require both $\mathbb{E}(|I|)$ and $\mathrm{Var}(|I|)$. First, a straightforward calculation yields

$$\mathbb{E}(|I|) = \mathbb{E}\left[\sum_{v\in\mathcal{C}_n^0} 1_{\{\deg(v)=0\}}\right] = N_n\mathbb{P}(\deg(v)=0) = N_n(1-p_n)^{N_n-1}.$$

Turning to the variance, we have

$$|I|^2 = \sum_{v,w\in\mathcal{C}_n^0} 1_{\{\deg(v)=\deg(w)=0\}} = \sum_{v\in\mathcal{C}_n^0} 1_{\{\deg(v)=0\}} + \sum_{v,w\in\mathcal{C}_n^0 : v\neq w} 1_{\{\deg(v)=\deg(w)=0\}}.$$

Therefore, we have

$$\begin{aligned}
\mathrm{Var}(|I|) &= \mathbb{E}(|I|^2) - (\mathbb{E}(|I|))^2 \\
&= \mathbb{E}\left[\sum_{v\in\mathcal{C}_n^0} 1_{\{\deg(v)=0\}} + \sum_{v,w\in\mathcal{C}_n^0 : v\neq w} 1_{\{\deg(v)=\deg(w)=0\}}\right] - N_n^2(1-p_n)^{2N_n-2} \\
&= N_n(1-p_n)^{N_n-1} + N_n(N_n-1)(1-p_n)^{2N_n-3} - N_n^2(1-p_n)^{2N_n-2} \\
&= N_n(1-p_n)^{N_n-1}(1-(1-p_n)^{N_n-2}) + N_n^2(1-p_n)^{2N_n-3}p_n \\
&\leq N_n(1-p_n)^{N_n-1}(N_n-2)p_n + N_n^2(1-p_n)^{2N_n-3}p_n \\
&\leq N_n(N_n-2)p_n + N_n^2 p_n \leq 2N_n^2 p_n,
\end{aligned}$$

where the first inequality follows from Bernoulli's inequality.

We will utilize $\mathbb{E}(|I|)$ and $\mathrm{Var}(|I|)$ to show that

$$\lim_{n\to\infty} \mathbb{P}(|I| \geq N_n - 2n) = 0. \tag{5.2}$$

It suffices to prove (5.2) in the three cases below.

1. When $\alpha_n \gg N_n$, we have $p_n \gg \frac{1}{N_n}$. Applying Markov's inequality, we have

$$\mathbb{P}(|I| > N_n - 2n) \leq \frac{\mathbb{E}(|I|)}{N_n - 2n} = \frac{N_n}{N_n - 2n}(1 - p_n)^{N_n - 1}.$$

Since $\lim_{n\to\infty}(1-p_n)^{N_n-1} = \lim_{n\to\infty}(1-p_n)^{\frac{1}{p_n}p_n(N_n-1)} = \lim_{n\to\infty} e^{-p_n(N_n-1)} = 0$, we have

$$\lim_{n\to\infty} \mathbb{P}(|I| > N_n - 2n) = 0.$$

2. When $\alpha_n \sim N_n$, we have $p_n \sim \frac{1}{N_n}$, and thus $p_n > \frac{c}{N_n}$ for some constant $c > 0$ and n large enough. Therefore

$$\mathbb{E}(|I|) = N_n(1 - p_n)^{N_n - 1} \leq N_n \left(1 - \frac{c}{N_n}\right)^{N_n - 1} \leq N_n e^{-c}.$$

Applying Chebyshev's inequality yields

$$\mathbb{P}(|I| > N_n - 2n) \leq \frac{\mathrm{Var}(|I|)}{(N_n - 2n - E[|I|])^2} \leq \frac{2N_n^2 p_n}{(N_n - 2n - N_n e^{-c})^2} = \frac{2p_n}{(1 - 2n/N_n - e^{-c})^2}.$$

Since $p_n \sim \frac{1}{N_n}$ and $N_n \sim n^2$, we have

$$\lim_{n\to\infty} \mathbb{P}(|I| > N_n - 2n) = 0.$$

3. The last case is when $\alpha_n \ll N_n$, or $p_n \ll \frac{1}{N_n}$. Using Taylor's expansion, we have

$$\mathbb{E}(|I|) = N_n(1 - p_n)^{N_n - 1} \leq N_n \left(1 - p_n(N_n - 1) + p_n^2 \frac{(N_n - 1)(N_n - 2)}{2}\right).$$

Again, we apply Chebyshev's inequality:

$$\begin{aligned}
\mathbb{P}(|I| \geq N_n - 2n) &\leq \frac{\mathrm{Var}(|I|)}{(N_n - 2n - E[|I|])^2} \\
&\leq \frac{2N_n^2 p_n}{\left(N_n - 2n - N_n + N_n(N_n-1)p_n - \frac{N_n(N_n-1)(N_n-2)}{2} p_n^2\right)^2} \\
&= \frac{2N_n^2 p_n}{\left(-2n + N_n(N_n-1)p_n - \frac{N_n(N_n-1)(N_n-2)}{2} p_n^2\right)^2}
\end{aligned}$$

Now we plug in $p_n = \frac{2n+\alpha_n}{N_n(N_n-1)}$ and proceed:

$$\begin{aligned}
\mathbb{P}(|I| \geq N_n - 2n) &\leq \frac{\frac{2N_n}{N_n-1}(2n+\alpha_n)}{\left(-2n + 2n + \alpha_n - \frac{N_n-2}{2N_n(N_n-1)}(2n+\alpha_n)^2\right)^2} \\
&= \frac{2N_n}{N_n-1} \frac{2n+\alpha_n}{\left(\alpha_n - \frac{N_n-2}{2N_n(N_n-1)}(2n+\alpha_n)^2\right)^2}.
\end{aligned}$$

If $\alpha_n \ll n$ or $\alpha_n \sim n$, we have

$$\frac{2n+\alpha_n}{\left(\alpha_n - \frac{N_n-2}{2N_n(N_n-1)}(2n+\alpha_n)^2\right)^2} \sim \frac{n}{\alpha_n^2} \to 0,$$

as $n \to \infty$, since $\alpha_n \gg n^{1/2}$.

If $\alpha_n \gg n$, we have

$$\frac{2n+\alpha_n}{\left(\alpha_n - \frac{N_n-2}{2N_n(N_n-1)}(2n+\alpha_n)^2\right)^2} \sim \frac{\alpha_n}{\alpha_n^2} = \frac{1}{\alpha_n} \to 0,$$

as $n \to \infty$, since $\alpha_n \ll N_n$. Thus, either way we must have

$$\lim_{n\to\infty} \mathbb{P}(|I| > N_n - 2n) = 0.$$

In all cases above, we have $\lim_{n\to\infty} \mathbb{P}(|I| \geq N_n - 2n) = 0$. $\square$

We are now ready to provide the first main theorem.

Theorem 5.2. *For* $\lim_{n\to\infty} \frac{p_n}{r(n)} = \infty$*, the following holds*

$$\lim_{n\to\infty} \mathbb{P}(\delta_{R_n} = 0) = 0.$$

Proof. Note that the vertices of the reaction network R_n correspond to the vertices in G_n with positive degree. Thus, letting I denote the set of isolated vertices of G_n, Lemma 2.13(e) implies that if the network is deficiency zero, we must have

$$|I| \geq N_n - 2n. \tag{5.3}$$

From (5.3), we have

$$\mathbb{P}(\delta_{R_n} = 0) \leq \mathbb{P}(|I| \geq N_n - 2n). \tag{5.4}$$

Since $r(n) = \frac{1}{n^3} \sim \frac{n}{N_n^2}$ and $p_n \gg r(n)$, we have p_n satisfies the condition in Lemma 5.1. Hence, using Lemma 5.1 we have

$$\lim_{n\to\infty} \mathbb{P}(\delta_{R_n} = 0) = \lim_{n\to\infty} \mathbb{P}(|I| \geq N_n - 2n) = 0. \qquad \square$$

5.2.2 The case $\lim_{n\to\infty} \frac{p_n}{r(n)} = 0$

The previous section considered when $\lim_{n\to\infty} \frac{p_n}{r(n)} = \infty$. Here we focus on the latter case, where $\lim_{n\to\infty} \frac{p_n}{r(n)} = 0$.

We will show in Lemma 5.3 that as $n \to \infty$, a random reaction network with $\lim_{n\to\infty} \frac{p_n}{r(n)} = 0$ almost surely contains only connected components that consist of 2 vertices. Thus in the corresponding reaction network each connected component has exactly 2 vertices.

Lemma 5.3. *Suppose* $\lim_{n\to\infty} \frac{p_n}{r(n)} = 0$. *Then*

$$\lim_{n\to\infty} \mathbb{P}(R_n \textit{ is not paired}) = 0$$

Proof. We have

$$\begin{aligned}\mathbb{P}(R_n \text{ is not paired}) = \mathbb{P}(&R_n \text{ is not paired}, R_n \text{ contains only trees}) \\ &+ \mathbb{P}(R_n \text{ is not paired}, R_n \text{ contains a cycle}).\end{aligned}$$

It is a well-known fact in random graph theory (for example, see [22]) that for $p_n \ll \frac{1}{n^3} \ll \frac{1}{N_n}$ we have

$$\lim_{n\to\infty} \mathbb{P}(R_n \text{ contains a cycle}) = 0.$$

Thus it suffices to show

$$\lim_{n\to\infty} \mathbb{P}(R_n \text{ is not paired}, R_n \text{ contains only trees}) = 0.$$

We follow the notation in [15] and for $k \geq 2$ let $T_k(n)$ be the number of trees in R_n with k vertices. Using estimates similar to the ones in [15], we have

$$\begin{aligned}\mathbb{P}(R_n \text{ is not paired}, R_n \text{ contains only trees}) &\leq \sum_{k=3}^{N_n} \mathbb{P}(T_k(n) > 0) \\ &\leq \sum_{k=3}^{N_n} \binom{N_n}{k} k^{k-2} p_n^{k-1} \\ &\leq \sum_{k=3}^{N_n} \frac{N_n^k e^k}{\sqrt{2\pi} k^k} k^k p_n^{k-1} \\ &= \frac{1}{\sqrt{2\pi}} N_n^3 e^3 p_n^2 \sum_{k=0}^{N_n - 3} (N_n e p_n)^k,\end{aligned}$$

where the first inequality follows since $\{\text{not paired}, \text{only trees}\} \subset \cup_{k=3}^{N_n} \{T_k(n) > 0\}$, the second follows by choosing the k vertices from the N_n choices and noting there are k^{k-2}

possible trees from these vertices (each with $k-1$ edges), and the third follows from Stirling. Since $p_n \ll \frac{1}{n^3} \sim N_n^{-3/2}$, we have $\lim_{n\to\infty} N_n^3 e^3 p_n^2 = 0$ and $\sum_{k=0}^{N_n-3}(N_n e p_n)^k$ is bounded. Thus we have

$$\lim_{n\to\infty} \mathbb{P}(R_n \text{ is not paired}) = \lim_{n\to\infty} \mathbb{P}(R_n \text{ is not paired, } R_n \text{ contains only trees}) = 0,$$

and the proof is complete $\square$

Remark 5.4. *Note that for $p_n \ll \frac{1}{n^3}$, the expected number of edges is*

$$p_n \binom{N_n}{2} = p_n \frac{N_n(N_n-1)}{2} \ll n.$$

Thus for $p_n \ll \frac{1}{n^3}$, R_n is almost surely paired with the number of pairs $k_n \ll n$.

Recall that we only consider binary reaction networks, thus each reaction can contain at most 4 species (2 species in each vertex). The next Lemma shows that for our analysis later, it suffices to only consider reactions that contain exactly 4 species.

Note that in the construction we are using, random graphs with the same number of edges have the same probability. We use this fact heavily in the proofs of the next two lemmas, where we condition on R_n being k_n-paired and can therefore generate R_n uniformly from the set of all k_n-paired graphs.

Lemma 5.5. *Suppose that $k_n \ll n$. Let A_n be the event that all reactions in R_n have exactly 4 distinct species. Then we have*

$$\lim_{n\to\infty} \mathbb{P}(A_n | R_n \textit{ is } k_n\textit{-paired}) = 1.$$

Proof. Let R_n be a k_n-paired reaction network, where $k_n \ll n$. Denote the k_n reaction vectors by $\{v_n^i\}_{i=1}^{k_n} \in \mathbb{Z}^n$. We denote by A_n^i the event that the vector v_n^i has 4 non-zero elements, thus $A_n = \cap_{i=1}^{k_n} A_n^i$. The proof will proceed by using that

$$\mathbb{P}(A_n | R_n \text{ is } k_n\text{-paired}) = \prod_{j=0}^{k_n-1} \mathbb{P}(A_n^{j+1} | \cap_{i=1}^{j} A_n^i, R_n \text{ is } k_n\text{-paired}),$$

and showing the limit of the right-hand side, as $n \to \infty$, is 1.

First, note that the total number of vertices of the form $S_k + S_m$ where $k \neq m$ is $\binom{n}{2}$. Suppose we have already picked j pairs of reversible reactions where each pair has 4 species. Then the number of unpicked vertices of the form $S_k + S_m$ where $k \neq m$ is $\binom{n}{2} - 2j$. After picking one such $S_k + S_m$ for the $j + 1^{st}$ pair, we need to pick another vertex. The number of available vertices of the form $S_p + S_q$, where p, q, m, and k are all different is at least $\binom{n-2}{2} - 2j$, where the minus 2 comes from the fact that we remove the species S_k and S_m from the possibilities, and the $2j$ is the number of vertices we have already chosen. Thus for n large enough, we have

$$\begin{aligned}
&\mathbb{P}(A_n^{j+1} | \cap_{i=1}^{j} A_n^i, R_n \text{ is } k_n\text{-paired}) \\
&\qquad \geq \frac{\frac{1}{2}(\binom{n}{2} - 2j)(\binom{n-2}{2} - 2j)}{\binom{N_n - 2j}{2}} && \text{(by considering our choices as detailed above)} \\
&\qquad \geq \frac{\frac{1}{2}(\binom{n}{2} - 2n)(\binom{n-2}{2} - 2n)}{\binom{N_n}{2}} && \text{(since } j \leq n\text{)} \\
&\qquad = \frac{(n^2 - 5n)(n^2 - 9n + 6)}{(n^2 + 3n + 2)(n^2 + 3n)} \geq \frac{(n^2 - 5n)(n^2 - 9n)}{(n^2 + 4n)(n^2 + 3n)} \\
&\qquad = \frac{n^2 - 14n + 45}{n^2 + 7n + 12} = 1 - \frac{21n - 33}{n^2 + 7n + 12} \\
&\qquad \geq 1 - \frac{21}{n},
\end{aligned}$$

and where the 1/2 in the first term accounts for the symmetry between the selected vertices.

Therefore, for n large enough, we have

$$\mathbb{P}(A_n|R_n \text{ is } k_n\text{-paired}) = \prod_{j=0}^{k_n-1} \mathbb{P}(A_n^{j+1}| \cap_{i=1}^{j} A_n^i, R_n \text{ is } k_n\text{-paired}) \geq \left(1 - \frac{21}{n}\right)^{k_n} \geq 1 - \frac{21k_n}{n} \tag{5.5}$$

where the last inequality is due to Bernoulli's inequality. Using the assumption that $k_n \ll n$, we have

$$\lim_{n\to\infty} \mathbb{P}(A_n|R_n \text{ is } k_n\text{-paired}) = 1,$$

and the proof is complete. □

Lemma 5.5 showed that if $k_n \ll n$ and R_n is k_n-paired, then with high probability each reaction vector will have precisely 4 non-zero components. The following proposition, stated in terms of discrete random matrices, proves that with probability approaching one, as $n \to \infty$, this set of reaction vectors will be linearly independent.

Proposition 5.1. *For each $n \geq 1$, let $D_n \subset \mathbb{R}^n$ be a set of vectors for which (i) each vector in D_n has precisely four non-zero elements, and (ii) for each choice of four distinct indices from $\{1, \dots, n\}$ there is precisely one vector in D_n with those as its non-zero components. Let $k_n \ll n$ and let $\Gamma_n \in \mathbb{R}^{n\times k_n}$ be a matrix whose columns are distinct vectors chosen uniformly from D_n. Then, Γ_n will have full column rank with probability converging to one, as $n \to \infty$.*

Proof. Let I_n be the event that all column vectors of Γ_n are linearly independent. It suffices to show

$$\lim_{n\to\infty} \mathbb{P}(I_n^c) = 0.$$

We denote the k_n column vectors of Γ_n by $\{v_n^i\}_{i=1}^{k_n} \in \mathbb{R}^n$. We say a set of vectors is *minimally dependent* if any of its proper subsets are linearly independent. For any set of indices of vectors $T \subseteq \{1, 2, \dots, k_n\}$ we denote $V_n^T = \{v_n^i : i \in T\}$. By noting that

$$I_n^c = \bigcup_{\ell=2}^{k_n} \{\exists \text{ a minimally dependent set of size } \ell\},$$

we have

$$\mathbb{P}(I_n^c) \leq \sum_{\ell=2}^{k_n} \sum_{|T|=\ell} \mathbb{P}(V_n^T \text{ is minimaly dependent}) = \sum_{\ell=2}^{k_n} \binom{k_n}{l} \mathbb{P}(B_\ell) \tag{5.6}$$

where B_ℓ is the event that V_n^T is minimally dependent for a particular set T satisfying $|T| = \ell$.

Now fix a set T with $|T| = \ell$. Without loss of generality, let $T = \{1, 2, \dots, \ell\}$. Consider a matrix M_ℓ whose columns are the vectors in V_n^T. Note that the set V_n^T being minimally dependent implies that M_ℓ has no row with only one non-zero entry (for otherwise, the set of vectors without the column associated to that element would be linearly dependent). This implies further that each non-zero row of M_ℓ has at least 2 entries. Since each column of M_ℓ has exactly 4 entries, M_ℓ has exactly 4ℓ entries. Therefore, the number of non-zero rows in M_ℓ must be at most 2ℓ and the number of zero rows in M_ℓ must be at least $n - 2\ell$. Combining all of the arguments above, we must have

$$\mathbb{P}(B_\ell) \leq \mathbb{P}(M_\ell \text{ has at least } n - 2\ell \text{ zero rows}). \tag{5.7}$$

We denote the row vectors of M_ℓ by $\{w_n^i\}_{i=1}^n$. For a subset of indices of species $R \subseteq \{1, 2, \dots, n\}$ we denote $W_n^R = \{w_n^i : i \in R\}$. We say that $W_n^R = 0$ if all the vectors in the set are the zero vector. We have

$$\mathbb{P}(M_\ell \text{ has at least } n - 2\ell \text{ zero rows}) \leq \sum_{|R|=n-2\ell} \mathbb{P}(W_n^R = 0) = \binom{n}{n-2\ell} \mathbb{P}(C_\ell) \tag{5.8}$$

where C_ℓ is the event that $W_n^R = 0$ for a particular R satisfying $|R| = n - 2\ell$.

Now fix a set R with $|R| = n-2\ell$. Without loss of generality, let $R = \{2\ell+1, \dots, n\}$. Then the event C_ℓ involves picking ℓ column vectors: $V_n^T = \{v_n^1, \dots, v_n^\ell\}$ where the last $n - 2\ell$ elements of each column vector are zero. Recall that each column vector has exactly 4 non-zero elements. Suppose we have already picked j such column vectors. The number of ways we can pick the $j+1$-st vector is at least $(\binom{n}{2} - 2j)(\binom{n-2}{2} - 2j)$ (this follows from the same argument as in the proof of Lemma 5.5). Among these, the number of ways we can pick the $j+1$-st vector whose last $n-2\ell$ elements are zero is less than $\binom{2\ell}{2}\binom{2\ell-2}{2}$. Thus we have

$$\mathbb{P}(C_\ell) \le \prod_{j=0}^{\ell-1} \frac{\binom{2\ell}{2}\binom{2\ell-2}{2}}{(\binom{n}{2} - 2j)(\binom{n-2}{2} - 2j)} \le \prod_{j=0}^{\ell-1} \frac{\binom{2\ell}{2}\binom{2\ell-2}{2}}{\frac{1}{4}\binom{n}{2}\binom{n-2}{2}} \le 4\left(\frac{2\ell}{n}\right)^4.$$

Plugging the above into (5.8), we see

$$\begin{aligned}\mathbb{P}(M_\ell \text{ has at least } n-2\ell \text{ zero rows}) &\le \binom{n}{n-2\ell} 4\left(\frac{2\ell}{n}\right)^{4\ell} \le \frac{n^{2\ell}}{(2\ell)!} 4\left(\frac{2\ell}{n}\right)^{4\ell} \\ &\le \frac{4n^{2\ell}}{\sqrt{2\pi}(2\ell/e)^{2\ell}}\left(\frac{2\ell}{n}\right)^{4\ell} = \frac{4}{\sqrt{2\pi}}\left(\frac{2\ell e}{n}\right)^{2\ell}.\end{aligned} \tag{5.9}$$

Now combining (5.6), (5.7), and (5.9), we have

$$\begin{aligned}\mathbb{P}(I_n^c) &\le \sum_{\ell=2}^{k_n} \binom{k_n}{\ell} \frac{4}{\sqrt{2\pi}}\left(\frac{2\ell e}{n}\right)^{2\ell} \le \sum_{\ell=2}^{k_n} \frac{k_n^\ell}{\ell!} \frac{4}{\sqrt{2\pi}}\left(\frac{2\ell e}{n}\right)^{2\ell} \\ &\le \sum_{\ell=2}^{k_n} \frac{k_n^\ell}{\sqrt{2\pi}(\ell/e)^\ell} \frac{4}{\sqrt{2\pi}}\left(\frac{2\ell e}{n}\right)^{2\ell} = \sum_{\ell=2}^{k_n} \frac{2}{\pi}\left(\frac{4\ell e^3 k_n}{n^2}\right)^{\ell} \\ &\le \sum_{\ell=2}^{\infty} \frac{2}{\pi}\left(\frac{4e^3 k_n^2}{n^2}\right)^{\ell} \le c\frac{k_n^4}{n^4}.\end{aligned} \tag{5.10}$$

for some constant $c > 0$, since $k_n \ll n$. Thus we have

$$\lim_{n\to\infty} \mathbb{P}(I_n^c) = 0$$

which concludes the proof. $\square$

We return to the setting of reaction networks with our final key lemma.

Lemma 5.6. *Suppose that $k_n \ll n$. Then we have*

$$\lim_{n\to\infty} \mathbb{P}(\delta_{R_n} = 0 | R_n \textit{ is } k_n\textit{-paired}) = 1. \tag{5.11}$$

Proof. Let R_n be a k_n-paired reaction network, where $k_n \ll n$. From Lemma 2.13, R_n has deficiency zero if and only if all k_n reaction vectors are linearly independent. Let I_n be the event that all k_n reaction vectors are linearly independent.

Similar to Lemma 5.5, denote the k_n reaction vectors by $\{v_n^i\}_{i=1}^{k_n} \in \mathbb{Z}^n$ and denote by A_n the event that all reactions have exactly 4 species. We have

$$\begin{aligned}\mathbb{P}(\delta_{R_n} = 0 | R_n \text{ is } k_n\text{-paired}) &= \mathbb{P}(I_n | R_n \text{ is } k_n\text{-paired}) \\ &\geq \mathbb{P}(I_n | A_n, R_n \text{ is } k_n\text{-paired})\mathbb{P}(A_n | R_n \text{ is } k_n\text{-paired}). \end{aligned} \tag{5.12}$$

Utilizing (5.10) in Proposition 5.1, we have

$$\mathbb{P}(I_n^c | A_n, R_n \text{ is } k_n\text{-paired}) \leq c\frac{k_n^4}{n^4} \tag{5.13}$$

for some constant $c > 0$. Thus using (5.5), (5.12) and (5.13), we must have

$$\mathbb{P}(\delta_{R_n} = 0 | R_n \text{ is } k_n\text{-paired}) \geq \left(1 - c\frac{k_n^4}{n^4}\right)\left(1 - \frac{21k_n}{n}\right), \tag{5.14}$$

Since $k_n \ll n$, taking the limit of (5.14) concludes the proof of the lemma. □

Combining Lemmas 5.3, 5.5, and 5.6, we are ready to state the main theorem for this section.

Theorem 5.7. *Suppose $p_n \ll \frac{1}{n^3}$, then*

$$\lim_{n\to\infty} \mathbb{P}(\delta_{R_n} = 0) = 1,$$

Proof. We have

$$\mathbb{P}(\delta_{R_n} = 0) = \mathbb{P}(\delta_{R_n} = 0, R_n \text{ is paired}) + \mathbb{P}(\delta_{R_n} = 0, R_n \text{ is not paired}).$$

Since

$$\mathbb{P}(\delta_{R_n} = 0, R_n \text{ is not paired}) \leq \mathbb{P}(R_n \text{ is not paired}),$$

we must have

$$\lim_{n\to\infty} \mathbb{P}(\delta_{R_n} = 0, R_n \text{ is not paired}) = 0$$

due to Lemma 5.3. Therefore it suffices to show

$$\lim_{n\to\infty} \mathbb{P}(\delta_{R_n} = 0, R_n \text{ is paired}) = 1.$$

Noting that for deficiency zero models, the number of reversible reaction vectors is bounded above by n, we have

$$\begin{aligned}
\mathbb{P}(\delta_{R_n} = 0, R_n \text{ is paired}) &= \sum_{i=1}^{n} \mathbb{P}(\delta_{R_n} = 0, G_n \text{ is } i\text{-paired}) \\
&= \sum_{i=1}^{n} \mathbb{P}(\delta_{R_n} = 0 | R_n \text{ is } i\text{-paired}) \mathbb{P}(R_n \text{ is } i\text{-paired}) \\
&= \sum_{i=1}^{n} \mathbb{P}(\delta_{R_n} = 0 | R_n \text{ is } i\text{-paired}) \frac{N_n!}{i! 2^i (N_n - 2i)!} p_n^i (1-p_n)^{N_n(N_n-1)/2-i} \\
&\geq \sum_{i=1}^{n} \mathbb{P}(\delta_{R_n} = 0 | R_n \text{ is } i\text{-paired}) \frac{(N_n - 2i)^{2i}}{i! 2^i} p_n^i (1-p_n)^{N_n(N_n-1)/2-i}
\end{aligned} \tag{5.15}$$

where the third equality uses that the number of i-paired graphs is $\binom{N_n}{2}\binom{N_n-2}{2}\dots\binom{N_n-2i+2}{2}$, with the repetition of the graphs accounted for by division by $i!$.

Note that because $p_n \ll 1/n^3$ and $N_n \sim n^2$ we have that $N_n^2 p_n \ll n$. Now let k_n satisfy $\lim_{n\to\infty} k_n = \infty$ and $N_n^2 p_n \ll k_n \ll n$. Cutting off the last $n - k_n$ terms from

(5.15), yields

$$\begin{aligned}
\mathbb{P}(\delta_{R_n} = 0, R_n \text{ is paired}) &\geq \sum_{i=1}^{k_n} \mathbb{P}(\delta_{R_n} = 0 | R_n \text{ is } i\text{-paired}) \frac{(N_n - 2i)^{2i}}{i! 2^i} p_n^i (1-p_n)^{N_n(N_n-1)/2-i} \\
&\geq \sum_{i=1}^{k_n} \left(1 - c\frac{i^4}{n^4}\right)\left(1 - \frac{21i}{n}\right) \frac{(N_n - 2i)^{2i}}{i! 2^i} p_n^i (1-p_n)^{N_n(N_n-1)/2-i} \\
&\geq \left(1 - c\frac{k_n^4}{n^4}\right)\left(1 - \frac{21k_n}{n}\right)(1-p_n)^{N_n^2/2} \sum_{i=1}^{k_n} \frac{(N_n - 2i)^{2i}}{i! 2^i} p_n^i \\
&\geq \left(1 - c\frac{k_n^4}{n^4}\right)\left(1 - \frac{21k_n}{n}\right)(1-p_n)^{N_n^2/2} \sum_{i=1}^{k_n} \frac{(N_n - 2k_n)^{2i}}{i! 2^i} p_n^i.
\end{aligned}$$

where the second inequality is obtained from (5.14) in Lemma 5.6.

Let $\lambda_n = \frac{(N_n - 2k_n)^2 p_n}{2}$, and note that $\lambda_n \ll k_n$ since we chose $N_n^2 p_n \ll k_n$. Using Taylor's remainder theorem and Stirling's approximation, we have

$$\sum_{i=1}^{k_n} \frac{\lambda_n^i}{i!} \geq e^{\lambda_n} - \frac{e^{\lambda_n}\lambda_n^{k_n+1}}{(k_n+1)!} \geq e^{\lambda_n}\left(1 - \frac{\lambda_n^{k_n+1}}{\sqrt{2\pi}(k_n+1)^{k_n+1}e^{-k_n+1}}\right) = e^{\lambda_n}\left(1 - \frac{1}{\sqrt{2\pi}}\left(\frac{\lambda_n e}{k_n+1}\right)^{k_n+1}\right).$$

Thus we have

$$\mathbb{P}(\delta_{R_n} = 0, R_n \text{ is paired}) \geq \left(1 - c\frac{k_n^4}{n^4}\right)\left(1 - \frac{21k_n}{n}\right)(1-p_n)^{N_n^2/2} e^{\lambda_n}\left(1 - \frac{1}{\sqrt{2\pi}}\left(\frac{\lambda_n e}{k_n+1}\right)^{k_n+1}\right).$$

Since $\lambda_n \ll k_n \ll n$, the first, second, and last terms converge to one. Hence, it suffices to show

$$\lim_{n\to\infty} (1-p_n)^{N_n^2/2} e^{\lambda_n} = 1,$$

or

$$\lim_{n\to\infty} \frac{N_n^2}{2} \ln(1-p_n) + \lambda_n = 0.$$

Since $p_n \ll 1$, we have $-p_n - p_n^2 \leq \ln(1-p_n) \leq -p_n$. Thus

$$\frac{N_n^2}{2}\ln(1-p_n) + \lambda_n \leq -\frac{N_n^2}{2}p_n + \lambda_n = \frac{p_n}{2}((N_n - 2k_n)^2 - N_n^2) = \frac{p_n}{2}(-4k_n N_n + 4k_n^2).$$

On the other hand, and using the equality above,

$$\frac{N_n^2}{2}\ln(1-p_n)+\lambda_n \geq -\frac{N_n^2}{2}(p_n+p_n^2)+\lambda_n = \frac{p_n}{2}(-4k_nN_n+4k_n^2)-\frac{N_n^2p_n^2}{2}.$$

Since $k_n \ll n$, $N_n \sim n^2$ and $p_n \ll \frac{1}{n^3}$, we have

$$\lim_{n\to\infty}\frac{p_n}{2}(-4k_nN_n+4k_n^2)=0, \quad \text{and,} \quad \lim_{n\to\infty}\frac{N_n^2p_n^2}{2}=0.$$

Thus

$$\lim_{n\to\infty}\frac{N_n^2}{2}\ln(1-p_n)+\lambda_n=0,$$

which concludes the proof of the theorem. □

5.3 A stochastic block model framework for random reaction networks

While the basic Erdős-Rényi framework in Section 5.1 can serve as a good starting point due to its simplicity, in practice one may want to use a more flexible framework that can be easily adapted to different settings where reaction networks may have different underlying structures. For example, one may want to study a closed system where inflow and outflow reactions such as $\emptyset \leftrightharpoons S_i$ are prohibited. On the other hand, one could be interested in an open system where inflow and outflow reactions are abundant. In another setting, perhaps one wants to only allow for reactions that preserve the number of molecules such as $S_i \leftrightharpoons S_j$ or $S_i + S_j \leftrightharpoons S_h + S_k$.

To properly generate random reaction networks in those situations, this section considers a stochastic block model framework–a generalized Erdős-Rényi framework with weighted edge probabilities [25].

Let the set of species be $\mathcal{S} = \{S_1, S_2, \ldots, S_n\}$. We consider binary reaction networks with species in $\mathcal{S}$. The set of all possible vertices is then

$$\mathcal{C}_n^0 = \{\emptyset, S_i, S_i + S_j : \text{for } 1 \leq i \leq n \text{ and } 1 \leq j \leq n.\}$$

Recall from Section 5.1 that we denote $N_n = |\mathcal{C}_n^0|$, the cardinality of $\mathcal{C}_n^0$. We also obtain from Section 5.1 that

$$N_n = \frac{n^2 + 3n + 2}{2},$$

and so

$$n \sim \sqrt{2N_n}.$$

Definition 5.8. *We denote by* $E_n^{0,1}, E_n^{0,2}, E_n^{1,1}, E_n^{1,2}, E_n^{2,2}$ *the sets of edges, or reactions, as follows:*

$$E_n^{0,1} = \{\emptyset \leftrightharpoons S_i : 1 \leq i \leq n\}$$

$$E_n^{0,2} = \{\emptyset \leftrightharpoons S_i + S_j : 1 \leq i, j \leq n\}$$

$$E_n^{1,1} = \{S_i \leftrightharpoons S_j : 1 \leq i, j \leq n; i \neq j\}$$

$$E_n^{1,2} = \{S_i \leftrightharpoons S_j + S_k : 1 \leq i, j, k \leq n\}$$

$$E_n^{2,2} = \{S_i + S_j \leftrightharpoons S_h + S_k : 1 \leq i, j, k, h \leq n; (i,j) \neq (k,h); (i,j) \neq (h,k)\}.$$

Remark 5.9. $E_n^{0,1}, E_n^{0,2}, E_n^{1,1}, E_n^{1,2}, E_n^{2,2}$ *completely partition the set of all possible edges. Note that* $|E_n^{0,1}| \sim n$, $|E_n^{1,1}| \sim |E_n^{0,2}| \sim n^2$, $|E_n^{1,2}| \sim n^3$ *and* $|E_n^{2,2}| \sim n^4$. *In fact, we have* $|E_n^{i,j}| \sim n^{i+j}$. *Finally, note that the terms* edges *and* reactions *can be used interchangeably in the present context.*

We then consider a randomly generated network $G(N_n, p_n)$, which we will simply denote G_n throughout, where the set of vertices is the set of vertices $\mathcal{C}_n^0$, and the probability that there is an edge between two vertices is given as follows

1. an edge in $E_n^{0,1}$ appears in the random graph with probability $p_n^{0,1} = n^{\alpha_{0,1}} p_n$,
2. an edge in $E_n^{0,2}$ appears in the random graph with probability $p_n^{0,2} = n^{\alpha_{0,2}} p_n$,
3. an edge in $E_n^{1,1}$ appears in the random graph with probability $p_n^{1,1} = n^{\alpha_{1,1}} p_n$,
4. an edge in $E_n^{1,2}$ appears in the random graph with probability $p_n^{1,2} = n^{\alpha_{1,2}} p_n$,
5. an edge in $E_n^{2,2}$ appears in the random graph with probability p_n,

where $\alpha_{0,1}, \alpha_{0,2}, \alpha_{1,1}, \alpha_{1,2}$ are parameters that can be used to control the structure of the random graph. Each random graph now corresponds to a reaction network in the following way,

1. each vertex with positive degree in the random graph represents a vertex in the reaction network graph, and
2. each edge in the random graph represents a reaction in the reaction network graph. We can assume all reactions are reversible, i.e., that $y \to y' \in \mathcal{R} \implies y' \to y \in \mathcal{R}$, since deficiency does not depend on the direction of the edges.

Similar to Section 5.1, we will denote the reaction network associated with the graph $G(N_n, p_n)$ by $R(N_n, p_n)$, which we will often simplify to R_n. We will denote the deficiency of R_n by δ_{R_n}.

Remark 5.10. *In the next section it will be more useful to work with the expected and actual number of edges in each set $E_n^{i,j}$ instead of $p_n^{i,j}$. Thus, for convenience we denote by $M_{i,j}(n)$ the number of realized edges from $E_n^{i,j}$ and by $K_{i,j}(n) = \mathbb{E}[M_{i,j}(n)]$ the expected number of realized edges from $E_n^{i,j}$. It is straightforward to see that $M_{i,j}(n)$ has* *a binomial distribution, and from Remark 5.9 that*

$$K_{i,j}(n) \sim n^{i+j} n^{\alpha_{i,j}} p_n$$

for $(i,j) \neq (2,2)$ *and* $K_{2,2}(n) = n^4 p_n$.

With the stochastic block model above, we can model a wide range of reaction networks by tweaking the parameters $\{\alpha_{i,j}\}$. Next, we provide a few examples to illustrate this flexibility.

Example 9 (The case $\alpha_{0,1} = \alpha_{0,2} = \alpha_{1,1} = \alpha_{1,2} = 0$)**.** In this case, we recover the unweighted Erdős-Rényi framework in Section 5.1. From Section 5.2, the threshold function for deficiency zero is $r(n) = \frac{1}{n^3}$. In other words,

$$\lim_{n\to\infty} P(\delta_{R_n} = 0) = \begin{cases} 0 & \text{when} \quad \lim_{n\to\infty} \frac{p_n}{r(n)} = \infty \\ 1 & \text{when} \quad \lim_{n\to\infty} \frac{p_n}{r(n)} = 0 \end{cases}.$$

Lemma 5.3 and Lemma 5.5 tell us that for $\lim_{n\to\infty} \frac{p_n}{r(n)} = 0$, the random reaction networks we observe only contain edges from $E_n^{2,2}$ with high probability. In other words, with the unweighted framework, we only see deficiency zero in “closed systems” (reaction networks with no inflow and outflow) of a very particular type. Reactions such as inflow and outflow, unary-unary, and unary-binary are underrepresented in this case.

Example 10 (A closed system with $\alpha_{0,1} = \alpha_{0,2} = 0, \alpha_{1,1} = 2, \alpha_{1,2} = 1$)**.** In this case, we have the expected number of edges in $E_n^{0,1}$ is $K_{0,1}(n) \sim np_n$ and the expected number of edges in $E_n^{0,2}$ is $K_{0,2}(n) \sim n^2 p_n$. It is easy to check that the parameters $\alpha_{i,j}$ are selected such that

$$K_{1,1}(n) \sim K_{1,2}(n) \sim K_{2,2}(n) \sim n^4 p_n \quad \text{and} \quad K_{0,1}(n), K_{0,2}(n) \ll n^4 p_n.$$

Thus the random reaction networks we observe will have similar expected amount of reactions in $E_n^{1,1}$, $E_n^{1,2}$, $E_n^{2,2}$. We also have that the expected number of reactions in $E_n^{0,1}$ and $E_n^{0,2}$ is significantly less. In particular, if $p_n \ll \frac{1}{n^2}$, the probability of seeing any reaction in $E_n^{0,1}$ and $E_n^{0,2}$ goes to 0 as $n \to \infty$. Hence, the random networks we observe will not have inflow and outflow reactions with high probability. Thus, this scheme is suitable to model closed systems without underrepresenting unary-unary and unary-binary reactions, unlike the case in Example 3. From Theorem 5.23 below, the threshold function for this case is $r(n) = \frac{1}{n^3}$

Example 11 (An open system with $\alpha_{0,1} = 3$, $\alpha_{1,1} = \alpha_{0,2} = 2$, $\alpha_{1,2} = 1$)**.** In this case, the expected number of realized edges $K_{i,j}(n) \sim n^4 p_n$ for all (i,j). Thus, this scheme is suitable to model an "open system" with inflow and outflow reactions, and with similar amount of reactions from each type. See Figure 7 for a realization of this system with a specific choice of parameters. From Theorem 5.23 below, the threshold function for this case is $r(n) = \frac{1}{n^{10/3}}$.

5.4 Prevalence of deficiency zero reaction networks under a stochastic block model framework

In Section 5.4.1, we will provide a set of conditions on $K_{i,j}(n)$ that guarantee $\lim_{n\to\infty} \mathbb{P}(\delta_{R_n} = 0) = 0$. In Section 5.4.2, we will also show that under the "converse" of these conditions, $\lim_{n\to\infty} \mathbb{P}(\delta_{R_n} = 0) = 1$. Then in Section 5.4.3, we will use these conditions to form an algorithm to find the threshold function for deficiency zero. Specifically, given any choice of $\{\alpha_{i,j}\}$, the algorithm provides a single threshold function $r(n)$ for deficiency

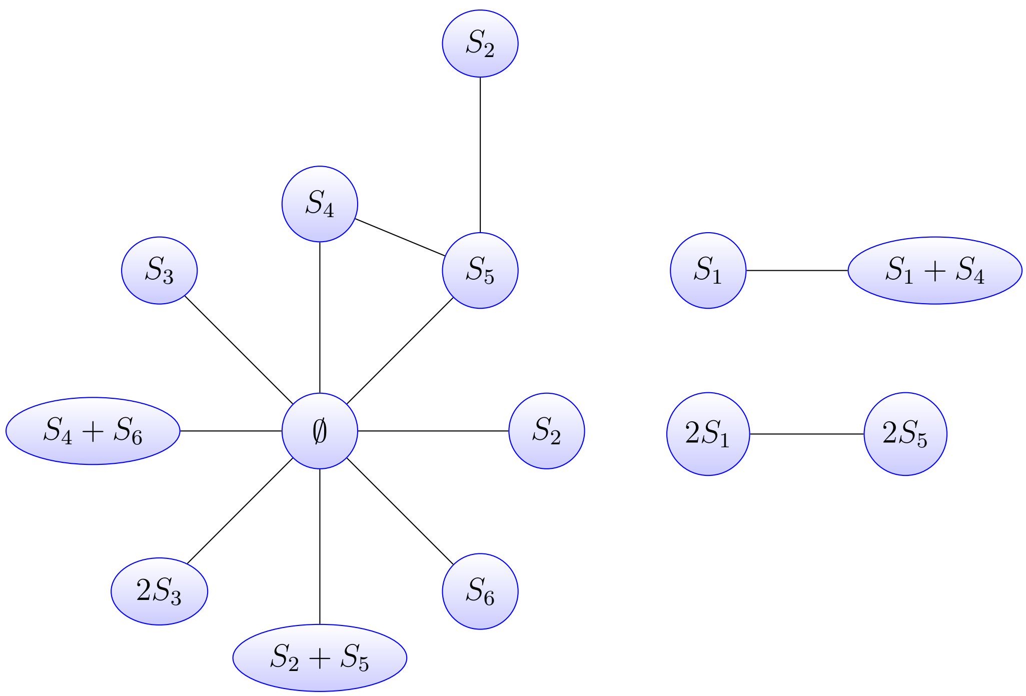

Figure 7: A realization of the open system in Example 5 with $n = 6$ and $p = \frac{0.8}{n^3}$. *Note: The figure only includes non-isolated vertices.*

zero.

5.4.1 Conditions on $K_{i,j}(n)$ for $\lim_{n\to\infty} \mathbb{P}(\delta_{R_n} = 0) = 0$

We start this section by providing some examples which illustrate different ways to break deficiency zero.

Example 12. Consider a reaction network with only 2 species $\mathcal{S} = \{S_1, S_2\}$

$$S_1 \leftrightharpoons S_2$$
$$S_1 + S_2 \leftrightharpoons \emptyset.$$

The reaction network has deficiency

$$\delta = |\mathcal{C}| - \ell - s = 4 - 2 - 2 = 0.$$

However, since there are only 2 species, we must have $s \leq 2$. Thus if we add more vertices and reactions, it is easy to get a positive deficiency from the new reaction network. For example, if we add $2S_1 \leftrightharpoons S_2$, then the new network is

$$S_1 \leftrightharpoons S_2 \leftrightharpoons 2S_1$$
$$S_1 + S_2 \leftrightharpoons \emptyset,$$

and the new deficiency is $\delta' = |\mathcal{C}'| - \ell' - s' = 5 - 2 - 2 = 1$. In this example, we break deficiency zero by having too many vertices with respect to the number of species.

Example 13. Consider a reaction network with 10 species $\mathcal{S} = \{S_1, \ldots, S_{10}\}$, which is given below

$$S_1 \leftrightharpoons S_2 \leftrightharpoons \cdots \leftrightharpoons S_{10}.$$

The reaction network has deficiency

$$\delta = |\mathcal{C}| - \ell - s = 10 - 1 - 9 = 0.$$

Note that all unary vertices are already in the network, and the dimension of the stochiometric subspace, which is 9, is nearly at the maximum possible value of 10. If we add one or two more reactions in $E_n^{1,2}, E_n^{0,2}$, or $E_n^{2,2}$, then it is easy to break deficiency zero since the dimension of the original network is almost at its maximum. For example, if we add $S_1 + S_2 \to S_3 + S_4$, then the new network is

$$S_1 \leftrightharpoons S_2 \leftrightharpoons \cdots \leftrightharpoons S_{10}$$
$$S_1 + S_2 \to S_3 + S_4$$

and the new deficiency is $\delta' = |\mathcal{C}'| - \ell' - s' = 12 - 2 - 9 = 1$. In this example, we break deficiency zero by adding too many more reactions when the dimension of the stoichiometric subspace is already nearly full from the unary reactions.

Example 14. Consider a reaction network with 10 species $\mathcal{S} = \{S_1, \ldots, S_{10}\}$, and a high number of reactions in $E_n^{0,1}$

$$\emptyset \leftrightharpoons S_i \quad \text{where} \quad i = 1, \ldots, 8.$$

The reaction network has deficiency

$$\delta = |\mathcal{C}| - \ell - s = 9 - 1 - 8 = 0.$$

If we add a high enough number of reactions in $E_n^{1,2}, E_n^{0,2}$, or $E_n^{2,2}$, then it is likely that we add a reaction whose species are in $\{S_1, \ldots, S_8\}$, which breaks deficiency zero. For example, consider the new network

$$\emptyset \leftrightharpoons S_i \quad \text{where} \quad i = 1, \ldots, 8$$

$$S_1 + S_2 \leftrightharpoons S_9$$

$$S_3 + S_4 \leftrightharpoons S_7.$$

The new deficiency is $\delta' = |\mathcal{C}'| - \ell' - s' = 12 - 2 - 9 = 1$. In this example, we break deficiency zero by having a high number of reaction in $E_n^{0,1}$ and a high enough number of reaction in $E_n^{1,2}, E_n^{0,2}$, or $E_n^{2,2}$.

It turns out that the three examples above are representative of all cases when we have $\lim_{n\to\infty} \mathbb{P}(\delta_{R_n} = 0) = 0$. We provide rigorous conditions in the following theorem.

Theorem 5.11. *If one of the following conditions holds, then* $\lim_{n\to\infty} \mathbb{P}(\delta_{R_n} = 0) = 0$.

(C1.1) Either $K_{0,2}(n) \gg n$, $K_{1,2}(n) \gg n$, *or* $K_{2,2}(n) \gg n$.

(C1.2) $K_{1,1}(n) \gg n$ *and either* $K_{0,2}(n) \gg 1$, $K_{1,2}(n) \gg 1$, *or* $K_{2,2}(n) \gg 1$.

(C1.3) Either $K_{0,1}(n)^2 K_{0,2}(n) \gg n^2$, $K_{0,1}(n)^3 K_{1,2}(n) \gg n^3$, *or* $K_{0,1}(n)^4 K_{2,2}(n) \gg n^4$.

Remark 5.12. *In Theorem 5.11, the three conditions are not purely technical; there is intuition behind each condition as described in the examples at the beginning of this section, and below.*

1. *Condition C1.1 refers to the case when there are too many vertices in the reaction network, which makes its deficiency strictly positive (see Lemma 2.13(e)). Note that* $K_{0,1}(n) \gg n$ *and* $K_{1,1}(n) \gg n$ *can not break deficiency zero in this regard.*

Obviously, it is impossible to have $K_{0,1}(n) \gg n$ since $|E_n^{0,1}| = n$. The condition $K_{1,1}(n) \gg n$ by itself still results in the network being deficiency zero (see Lemma 2.13(h)). However, the condition $K_{1,1}(n) \gg n$ together with a non-trivial number of reactions from $E_n^{0,2}, E_n^{1,2}, E_n^{2,2}$ can break deficiency zero. This is stated formally in Condition C1.2.

2. *Condition C1.2 refers to the case when the dimension of the stochiometric subspace s is almost fully exhausted from reactions in $E_n^{1,1}$. Recall that $\delta = |\mathcal{C}| - \ell - s$, so in this case as we add more reactions in $E_n^{0,2}, E_n^{1,2}, E_n^{2,2}$, $|\mathcal{C}| - \ell$ increases but s does not, making the deficiency positive.*

3. *Condition C1.3 refers to the case where there is a high probability of some inflow or outflow reaction in $E_n^{0,1}$ and a reaction in another edge set being linearly dependent, which in turn makes the deficiency positive. It will also be apparent later that having a nontrivial number of inflow or outflow reactions in $E_n^{0,1}$ makes it more difficult to have deficiency zero.*

We prove the theorem via a series of lemmas. We begin by showing that if Condition (C1.1) holds, then $\lim_{n\to\infty} \mathbb{P}(\delta_{R_n} = 0) = 0$.

Lemma 5.13. *If either $K_{0,2}(n) \gg n$, $K_{1,2}(n) \gg n$, or $K_{2,2}(n) \gg n$, then we have*

$$\lim_{n\to\infty} \mathbb{P}(\delta_{R_n} = 0) = 0.$$

Proof. Recall from Lemma 2.13(e) that there cannot be too many vertices in a network with deficiency zero. In particular, $\delta_{R_n} = 0$ implies $|\mathcal{C}| \leq 2n$. We will argue that in each of the three cases the number of non-isolated vertices in G_n, which correspond with the vertices of the associated reaction network R_n, is likely to be much higher than the bound $2n$, implying the network has positive deficiency. The first case is straightforward, and the remaining two cases follow the same technique as Lemma 5.1 and Theorem 5.2 in Section 5.2.

1. First, we assume that $K_{0,2}(n) \gg n$. From Lemma 2.13(e), we have that $\delta_{R_n} = 0$ implies $|\mathcal{C}| \leq 2n$, which in turns implies $M_{0,2}(n) \leq 2n-1$. Thus

$$\begin{aligned}\mathbb{P}(\delta_{R_n} = 0) \leq \mathbb{P}(M_{0,2}(n) \leq 2n-1) \\ &= \mathbb{P}(K_{0,2}(n) - M_{0,2}(n) \geq K_{0,2}(n) - (2n-1)) \\ &\leq \frac{\mathrm{Var}(M_{0,2}(n))}{(K_{0,2}(n) - (2n-1))^2}.\end{aligned}$$

   Since $M_{0,2}(n)$ has a binomial distribution, $\mathrm{Var}(M_{0,2}(n)) \leq \mathbb{E}[M_{0,2}(n)] = K_{0,2}(n)$. Together with the fact that $K_{0,2}(n) \gg n$, we have $\frac{\mathrm{Var}(M_{0,2}(n))}{(K_{0,2}(n)-(2n-1))^2} \to 0$, as $n \to \infty$, and thus $\lim_{n\to\infty} \mathbb{P}(\delta_{R_n} = 0) = 0$.

2. Next, we assume $K_{1,2}(n) \gg n$. We observe that based on Corollary 2.15, $\delta_{R_n} = 0$ must imply $\delta_{\pi_{E_n^{1,2}}(R_n)} = 0$, where, recalling Definition 2.14, $\pi_{E_n^{1,2}}(R_n)$ is the subnetwork of R_n with reactions in $E_n^{1,2}$. Thus we have

$$\mathbb{P}(\delta_{R_n} = 0) \leq \mathbb{P}(\delta_{\pi_{E_n^{1,2}}(R_n)} = 0).$$

   Again, we make use of the upper bound in Lemma 2.13(e). $\delta_{\pi_{E_n^{1,2}}(R_n)} = 0$ must imply the number of non-isolated vertices in $\pi_{E_n^{1,2}}(R_n)$ is bounded by $2n$. Let I be the set of isolated binary vertices in $\pi_{E_n^{1,2}}(R_n)$. Since there are $\frac{n(n+1)}{2}$ binary vertices, we must then have

$$|I| > \frac{n(n+1)}{2} - 2n,$$

and as a result

$$\mathbb{P}(\delta_{R_n} = 0) \le \mathbb{P}\left(|I| > \frac{n(n+1)}{2} - 2n\right).$$

The probability that a binary vertex is isolated in $\pi_{E_n^{1,2}}(R_n)$ is $(1 - p_n^{1,2})^n$, because there are precisely n unary vertices. Thus, summing over the binary vertices yields

$$\mathbb{E}|I| = \frac{n(n+1)}{2}(1 - p_n^{1,2})^n.$$

We can also derive $\mathrm{Var}(|I|)$ since $|I|$ is binomially distributed. Using $\mathbb{E}|I|$ and $\mathrm{Var}(|I|)$, a rigorous proof for

$$\lim_{n\to\infty} \mathbb{P}\left(|I| > \frac{n(n+1)}{2} - 2n\right) = 0$$

can be carried out by precisely the same argument as Lemma 5.1 in Section 5.2. We omit it for the sake of brevity.

3. Finally, we assume $K_{2,2}(n) \gg n$. We observe that based on Corollary 2.15, $\delta_{R_n} = 0$ must imply $\delta_{\pi_{E_n^{2,2}}(R_n)} = 0$, where $\pi_{E_n^{2,2}}(R_n)$ is the subnetwork of R_n with reactions in $E_n^{2,2}$. Thus we have

$$\mathbb{P}(\delta_{R_n} = 0) \le \mathbb{P}(\delta_{\pi_{E_n^{2,2}}(R_n)} = 0).$$

Note that $K_{2,2}(n) \gg n$ implies $n^4 p_n \gg n$, and thus $p_n \gg \frac{1}{n^3}$. The remainder of the proof follows along the same lines as the proof of Lemma 5.1 and Theorem 5.2 in Section 5.2. $\square$

The following proposition will be useful in the proof that Condition C1.2 implies $\lim_{n\to\infty} \mathbb{P}(\delta_{R_n} = 0) = 0$.

Proposition 5.2. *Let $R = \{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ be a reaction network with $\mathcal{S} = \{S_1, S_2, \ldots, S_n\}$. Assume that all vertices in R are unary, and R has only one connected component. Let*

$i,j,p,q \in \{1,\dots,n\}$ be such that $\{i,j\} \neq \{p,q\}$, and let $\widehat{\mathcal{R}} = R \cup \{\emptyset \leftrightharpoons S_i + S_j, \emptyset \leftrightharpoons S_p + S_q\}$ and $\widehat{R}$ be the reaction network associated with $\widehat{\mathcal{R}}$. Then $\delta_{\widehat{R}} = 1$.

Note that in the above proposition we are allowing $i = j$ and/or $i = p$.

Proof. Due to Lemma 2.13(h), the deficiency of R is necessarily zero (since it contains only unary vertices). Starting from R, adding the pair of reversible reactions $\emptyset \leftrightharpoons S_i + S_j, \emptyset \leftrightharpoons S_p + S_q$ to form $\widehat{R}$ increases the number of vertices by three, and increases the number of connected components by 1. It is straightforward to check that since the vertices $\{S_i, S_j, S_p, S_q\}$ are contained within $\mathcal{C}$, the addition of the reaction vectors for $\emptyset \leftrightharpoons S_i + S_j$ and $\emptyset \leftrightharpoons S_p + S_q$ only increases the size of the dimension of the stoichiometric subspace by 1. Hence, we have $\delta_{\widehat{R}} = \delta_R + 3 - 1 - 1 = 1$. $\square$

We now show that Condition (C1.2) yields the desired result.

Lemma 5.14. *If $K_{1,1}(n) \gg n$ and either $K_{0,2}(n) \gg 1$, $K_{1,2}(n) \gg 1$, or $K_{2,2}(n) \gg 1$, then we have*

$$\lim_{n\to\infty} \mathbb{P}(\delta_{R_n} = 0) = 0.$$

Proof. Suppose $K_{0,2}(n) \gg 1$. The other two cases can be handled in a same manner.

$M_{0,2}(n)$ is binomially distributed with mean $K_{0,2}(n) \gg 1$. Thus, standard methods show

$$\lim_{n\to\infty} \mathbb{P}(M_{0,2}(n) \geq 2) = 1.$$

Now it suffices to show

$$\lim_{n\to\infty} \mathbb{P}(\delta_{R_n} = 0, M_{0,2}(n) \geq 2) = 0.$$

Let $G_n^{1,1}$ be the subgraph of G_n consisting of all vertices S_i (even those that are isolated) and all edges in $E_n^{1,1}$ that are realized in G_n. Let B_n be the largest component in $G_n^{1,1}$

and let $|B_n|$ be its size (number of vertices). When $M_{0,2}(n) \geq 2$, we let B_n^+ be the union of B_n with two edges chosen uniformly at random from $E_n^{0,2}$ that are realized in G_n. If $M_{0,2}(n) \leq 1$ we choose the two reactions uniformly at random from $E_n^{0,2}$. We denote the chosen two edges by $\emptyset \leftrightharpoons S_i + S_j, \emptyset \leftrightharpoons S_p + S_q$ and note that $\{i, j\} \neq \{p, q\}$. Note that by symmetry the distribution of the pair $(\emptyset \leftrightharpoons S_i + S_j, \emptyset \leftrightharpoons S_p + S_q)$ is the same as if we simply chose two reactions from $E_n^{0,2}$ uniformly at random. Since $\delta_{B_n^+} \leq \delta_{R_n}$, we must have

$$\mathbb{P}(\delta_{R_n} = 0, M_{0,2}(n) \geq 2) \leq \mathbb{P}(\delta_{B_n^+} = 0, M_{0,2}(n) \geq 2) \leq \mathbb{P}(\delta_{B_n^+} = 0).$$

Thus it suffices to show

$$\lim_{n\to\infty} \mathbb{P}(\delta_{B_n^+} = 0) = 0.$$

Conditioning on the size of B_n, the largest component of $G_n^{1,1}$, yields

$$\mathbb{P}(\delta_{B_n^+} = 0) = \sum_{k=1}^{n} \mathbb{P}(\delta_{B_n^+} = 0 || B_n| = k)\mathbb{P}(|B_n| = k). \tag{5.16}$$

From Proposition 5.2, we know that if B_n^+ has a deficiency of zero, then not all of S_i, S_j, S_p, S_q are contained in B_n. Thus we have

$$\begin{aligned} \mathbb{P}(\delta_{B_n^+} = 0 || B_n| = k) &\leq \mathbb{P}(\text{not all of } S_i, S_j, S_p, S_q \text{ are contained in } B_n || B_n| = k) \\ &= 1 - \mathbb{P}(S_i, S_j, S_p, S_q \in B_n || B_n| = k). \end{aligned} \tag{5.17}$$

We will compute the probability as follows

$$\mathbb{P}(S_i, S_j, S_p, S_q \in B_n || B_n| = k) = \mathbb{P}(S_p, S_q \in B_n | S_i, S_j \in B_n, |B_n| = k)\mathbb{P}(S_i, S_j \in B_n || B_n| = k). \tag{5.18}$$

We first consider the probability $\mathbb{P}(S_i, S_j \in B_n || B_n| = k)$. Since $|B_n| = k$, there are exactly $\binom{k}{2}$ ways of choosing a reaction of the form $\emptyset \leftrightharpoons S_i + S_j$ with $i \neq j$ and $S_i, S_j \in B_n$. Similarly, for the case $i = j$, there are exactly k ways of choosing a reaction of the form $\emptyset \leftrightharpoons 2S_i$ with $S_i \in B_n$. Since there are a total of $\binom{n}{2} + n$ elements in $E_n^{0,2}$ we have

$$\mathbb{P}(S_i, S_j \in B_n \| B_n| = k) = \frac{\binom{k}{2} + k}{\binom{n}{2} + n} = \frac{k(k+1)}{n(n+1)} \geq \left(\frac{k}{n}\right)^2, \tag{5.19}$$

where the inequality holds since $k \leq n$. Similarly, we have

$$\mathbb{P}(S_p, S_q \in B_n | S_i, S_j \in B_n, |B_n| = k) = \frac{\binom{k}{2} + k - 1}{\binom{n}{2} + n - 1} \geq \left(\frac{k}{n}\right)^2, \tag{5.20}$$

where the inequality holds for $k \geq 4$, which can be verified in a straightforward manner. From (5.17), (5.18), (5.19), and (5.20) we have that for $k \geq 4$

$$\mathbb{P}(\delta_{B_n^+} = 0 \| B_n| = k) \leq 1 - \left(\frac{k}{n}\right)^4. \tag{5.21}$$

Finally, combining (5.16) and (5.21), we have

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}(\delta_{B_n^+} = 0) &\leq \sum_{k=4}^{n} \left(1 - \left(\frac{k}{n}\right)^4\right) \mathbb{P}(|B_n| = k) + \sum_{k=1}^{3} \mathbb{P}(|B_n| = k) \\
&= \mathbb{E}\left(1 - \left(\frac{|B_n|}{n}\right)^4\right) + \sum_{k=1}^{3} \left(\frac{k}{n}\right)^4 \mathbb{P}(|B_n| = k) \\
&\leq 1 - \left(\mathbb{E}\left(\frac{|B_n|}{n}\right)\right)^4 + \frac{98}{n^4},
\end{aligned} \tag{5.22}$$

where the last inequality is due to Jensen's inequality. Since $K_{1,1}(n) \gg n$, we have the edge probability for the edges in $E_n^{1,1}$ satisfy

$$p_n^{1,1} \sim \frac{K_{1,1}(n)}{n^2} \gg \frac{n}{n^2} = \frac{1}{n}.$$

From, [20], $\frac{|B_n|}{n} - f(c_n) \xrightarrow{\mathbb{P}} 0$, where $f(c_n) = 1 - \frac{1}{c_n} \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!}(c_n e^{-c_n})^k$ and $c_n = np_n^{1,1}$. Since $np_n^{1,1} \gg 1$, it is straightforward to verify that $\lim_{n\to\infty} f(c_n) = 1$. Since both $\frac{|B_n|}{n}$ and $f(c_n)$ are bounded by 1, we have

$$\lim_{n\to\infty} \mathbb{E}\left(\frac{|B_n|}{n}\right) = \lim_{n\to\infty} f(c_n) = 1,$$

which completes the proof. □

Finally, we have the proof related to Condition C1.3.

Lemma 5.15. *If either $K_{0,1}(n)^2 K_{0,2}(n) \gg n^2$, $K_{0,1}(n)^3 K_{1,2}(n) \gg n^3$ or $K_{0,1}(n)^4 K_{2,2}(n) \gg n^4$, then we have*

$$\lim_{n\to\infty} \mathbb{P}(\delta_{R_n} = 0) = 0.$$

Proof. It suffices to show $K_{0,1}(n)^2 K_{0,2}(n) \gg n^2$ implies $\lim_{n\to\infty} \mathbb{P}(\delta_{R_n} = 0) = 0$. The other two cases follow the same argument. Recall that $M_{0,1}(n)$ has a binomial distribution with $|E_n^{0,1}| = n$ trials and mean $\mathbb{E}M_{0,1}(n) = K_{0,1}(n)$, and $M_{0,2}(n)$ is a binomial distribution with $|E_n^{0,2}| = n(n+1)/2$ trials and mean $\mathbb{E}M_{0,2}(n) = K_{0,2}(n)$. Thus we have

$$\mathbb{P}(\delta_{R_n} = 0) = \sum_{\substack{i\le n \\ j\le n(n+1)/2}} \mathbb{P}(\delta_{R_n} = 0 | M_{0,1}(n) = i, M_{0,2}(n) = j)\mathbb{P}(M_{0,1}(n) = i, M_{0,2}(n) = j). \tag{5.23}$$

Consider the event $\delta_{R_n} = 0$ conditioned on $M_{0,1}(n) = i, M_{0,2}(n) = j$. Note that a reaction network of the form $\emptyset \leftrightharpoons S_p, \quad \emptyset \leftrightharpoons S_q, \quad \emptyset \leftrightharpoons S_p + S_q$ has positive deficiency, so any network containing it also has positive deficiency according to Corollary 2.15. Thus $\delta_{R_n} = 0$ implies there is no such subnetwork in R_n.

There are $\frac{n(n+1)}{2}$ reactions in $E_n^{0,2}$, thus the probability that there is reaction of the form $\emptyset \leftrightharpoons S_p + S_q$ (note that p and q can be the same) where $\emptyset \leftrightharpoons S_p$ and $\emptyset \leftrightharpoons S_q$ are already present is

$$\frac{\binom{i}{2} + i}{\frac{n(n+1)}{2}} = \frac{i(i+1)}{n(n+1)}.$$

We may then us a sequential argument (on the j elements from $E_n^{0,2}$ that have been realized) that is similar to the one used around (5.20) to conclude

$$\mathbb{P}(\delta_{R_n} = 0 | M_{0,1}(n) = i, M_{0,2}(n) = j) \le \left(1 - \frac{i(i+1)}{n(n+1)}\right)^j. \tag{5.24}$$

Combining (5.23) and (5.24), we have

$$\mathbb{P}(\delta_{R_n}=0) \le \sum_{\substack{i\le n\\ j\le n(n+1)/2}} \left(1-\frac{i(i+1)}{n(n+1)}\right)^{j} \mathbb{P}(M_{0,1}(n)=i, M_{0,2}(n)=j)$$
$$= \mathbb{E}\left[\left(1-\frac{M_{0,1}(n)(M_{0,1}(n)+1)}{n(n+1)}\right)^{M_{0,2}(n)}\right].$$

We have $\frac{M_{0,1}(n)(M_{0,1}(n)+1)}{n(n+1)} \ge \frac{M_{0,1}(n)^2}{2n^2}$, thus

$$\mathbb{P}(\delta_{R_n}=0) \le \mathbb{E}\left[\left(1-\frac{M_{0,1}(n)^2}{2n^2}\right)^{M_{0,2}(n)}\right] \le \mathbb{E}\left[e^{-\frac{M_{0,1}(n)^2 M_{0,2}(n)}{2n^2}}\right], \tag{5.25}$$

where the second inequality follows the fact that $1-x \le e^{-x}$. Notice further that $e^{-x} \le \frac{1}{x+1}$ for $x \ge 0$, hence we have

$$\mathbb{E}\left[e^{-\frac{M_{0,1}(n)^2 M_{0,2}(n)}{2n^2}}\right] \le \mathbb{E}\left[\frac{2n^2}{M_{0,1}(n)^2 M_{0,2}(n)+2n^2}\right] = 2n^2\mathbb{E}\left[\frac{1}{M_{0,1}(n)^2 M_{0,2}(n)+2n^2}\right]. \tag{5.26}$$

Since $M_{0,1}(n) \le n$ and $M_{0,2}(n) \le \frac{n(n+1)}{2}$, we have for n large enough

$$\begin{aligned} M_{0,1}(n)^2 M_{0,2}(n)+2n^2 &\ge (M_{0,1}(n)^2+1)(M_{0,2}(n)+1)\\ &\ge \frac{1}{2}(M_{0,1}(n)+1)^2(M_{0,2}(n)+1)\\ &\ge \frac{1}{4}(M_{0,1}(n)+1)(M_{0,1}(n)+2)(M_{0,2}(n)+1), \end{aligned} \tag{5.27}$$

where the first inequality can be verified by expanding the right hand side and utilizing the inequalities on $M_{0,1}(n)$ and $M_{0,2}(n)$, the second inequality follows by the well known $\frac{1}{2}(a+b)^2 \le a^2+b^2$ inequality, and the last inequality comes from $M_{0,1}(n)+1 \ge \frac{1}{2}(M_{0,1}(n)+2)$, which is true as long as $M_{0,1}(n) \ge 0$.

Combining (5.25), (5.26), (5.27), and noticing that $M_{0,1}(n)$ and $M_{0,2}(n)$ are independent, we have

$$\mathbb{P}(\delta_{R_n}=0) \le 8n^2\mathbb{E}\left[\frac{1}{(M_{0,1}(n)+1)(M_{0,1}(n)+2)}\right]\mathbb{E}\left[\frac{1}{M_{0,2}(n)+1}\right]. \tag{5.28}$$

Since $M_{0,1}(n) \sim B(n, K_{0,1}(n)/n)$, from Lemma A.3, we have

$$\mathbb{E}\left[\frac{1}{(M_{0,1}(n)+1)(M_{0,1}(n)+2)}\right] \leq \frac{1}{K_{0,1}(n)^2}. \tag{5.29}$$

We also have $M_{0,2}(n) \sim B(n(n+1)/2, \frac{K_{0,2}(n)}{n(n+1)/2})$. Repeating the same argument as above, we have

$$\mathbb{E}\left[\frac{1}{M_{0,2}(n)+1}\right] \leq \frac{1}{K_{0,2}(n)}. \tag{5.30}$$

Thus from (5.28), (5.29), (5.30) we have

$$\mathbb{P}(\delta_{R_n} = 0) \leq \frac{8n^2}{K_{0,1}(n)^2 K_{0,2}(n)}.$$

Since $K_{0,1}(n)^2 K_{0,2}(n) \gg n^2$ the proof is complete. □

5.4.2 Conditions on $K_{i,j}(n)$ for $\lim_{n\to\infty} \mathbb{P}(\delta_{R_n} = 0) = 1$

Note that the conditions below are essentially the converse of Theorem 5.11.

Theorem 5.16. *If all of the following conditions hold, then* $\lim_{n\to\infty} \mathbb{P}(\delta_{R_n} = 0) = 1$.

(C2.1) $K_{0,2}(n) \ll n$, $K_{1,2}(n) \ll n$, *and* $K_{2,2}(n) \ll n$.

(C2.2) One of the following conditions holds

(C2.2.1) $K_{1,1}(n) \ll n$

(C2.2.2) $K_{0,2}(n) \ll 1$, $K_{1,2}(n) \ll 1$, *and* $K_{2,2}(n) \ll 1$.

(C2.3) $K_{0,1}(n)^2 K_{0,2}(n) \ll n^2$, $K_{0,1}(n)^3 K_{1,2}(n) \ll n^3$, *and* $K_{0,1}(n)^4 K_{2,2}(n) \ll n^4$.

We will begin by arguing that it is sufficient to prove that a slightly simplified set of conditions implies that $\lim_{n\to\infty} \mathbb{P}(\delta_{R_n} = 0) = 1$. First assume that conditions (C2.1), (C2.2.2), and (C2.3) hold. Condition (C2.2.2), combined with the fact that each $M_{i,j}(n)$ has a binomial distribution with mean $K_{i,j}(n)$, yields

$$\lim_{n\to\infty} \mathbb{P}(M_{0,2}(n) = 0) = \lim_{n\to\infty} \mathbb{P}(M_{1,2}(n) = 0) = \lim_{n\to\infty} \mathbb{P}(M_{2,2}(n) = 0) = 1.$$

Hence, with probability approaching 1, the realized network only has edges in $E_n^{0,1}$ and $E_n^{1,1}$, and has a deficiency of zero by Lemma 2.13(h). Hence, the proof in this situation is done, and we can now simply assume that the conditions (C2.1), (C2.2.1), and (C2.3) are satisfied.

However, another slight simplification can take place. Note that $K_{0,1}(n) \leq n$ (since $|E_n^{0,1}| = n$), and if $K_{0,1}(n) \sim n$, then from condition (C2.3), we would have that condition (C2.2.2) is satisfies, which we already know implies the result. Hence, we only need consider the case $K_{0,1}(n) \ll n$. For the other cases where there exist a subsequence along which $K_{0,1}(n) \sim n$ and another subsequence along which $K_{0,1}(n) \ll n$, we can apply the two corresponding arguments for the two subsequences, both of which when combined will still result in $\lim_{n\to\infty} \mathbb{P}(\delta_{R_n} = 0) = 1$. Combining the above shows that Theorem 5.16 will be proved by showing that $\lim_{n\to\infty} \mathbb{P}(\delta_{R_n} = 0) = 1$ so long as the following conditions are satisfied:

(C2.1*) All $K_{i,j}(n) \ll n$.

(C2.3) $K_{0,1}(n)^2 K_{0,2}(n) \ll n^2$, $K_{0,1}(n)^3 K_{1,2}(n) \ll n^3$, and $K_{0,1}(n)^4 K_{2,2}(n) \ll n^4$.

Showing the above is the goal for the remainder of this section. In the first lemma, we construct some "buffer" functions, $Q_{i,j}(n)$ that are, asymptotically, between $K_{i,j}(n)$ and n, and also satisfy a version of condition (C2.3).

Lemma 5.17. *If conditions (C2.1*) and (C2.3) hold, then there exists* $Q_{0,1}(n)$*,* $Q_{0,2}(n)$*,* $Q_{1,1}(n)$*,* $Q_{1,2}(n)$*,* $Q_{2,2}(n)$ *such that*

- $\lim_{n\to\infty} Q_{i,j}(n) > 0$ *for all* (i,j)*.*
- $K_{i,j}(n) \ll Q_{i,j}(n) \ll n$ *for all* (i,j)*.*
- $Q_{0,1}(n)^2 Q_{0,2}(n) \ll n^2$*,* $Q_{0,1}(n)^3 Q_{1,2}(n) \ll n^3$*, and* $Q_{0,1}(n)^4 Q_{2,2}(n) \ll n^4$*,*

Proof. We begin with $Q_{1,1}(n)$, which will be straightforward. Set

$$Q_{1,1}(n) = \max\{1, \sqrt{nK_{1,1}(n)}\}.$$

From $K_{1,1}(n) \ll n$ in (C2.1*) we have that $K_{1,1}(n) \ll Q_{1,1}(n) \ll n$ and $\lim_{n\to\infty} Q_{1,1}(n) > 0$.

We turn to constructing $Q_{0,2}$. In order to eventually convert the condition $K_{0,1}(n)^2 K_{0,2}(n) \ll n^2$ to the condition $Q_{0,1}(n)^2 Q_{0,2}(n) \ll n^2$, we will first construct a function $R_{0,1}(n)$, which satisfies $R_{0,1}(n)^2 K_{0,2}(n) \ll n^2$. We will then use $R_{0,1}(n)$ to build $Q_{0,2}(n)$ satisfying $R_{0,1}(n)^2 Q_{0,2}(n) \ll n^2$. After producing the pair $(R_{0,1}(n), Q_{0,1}(n))$, we turn to producing similar pairs $(S_{0,1}(n), Q_{1,2}(n))$ and $(T_{0,1}(n), Q_{2,2}(n))$, each satisfying similar inequalities. We will then define $Q_{0,1}(n)$ via the functions $R_{0,1}(n), S_{0,1}(n), T_{0,1}(n)$, and the proof will be complete.

Proceeding, we note that since $K_{0,1}(n)^2 K_{0,2} \ll n^2$, we have $K_{0,1}(n) \ll \frac{n}{\sqrt{K_{0,2}(n)}}$. By (C2.1*), we have $K_{0,1}(n) \ll n$ as well. Let

$$R_{0,1}(n) = \min\left\{\sqrt{K_{0,1}(n)\frac{n}{\sqrt{K_{0,2}(n)}}}, \sqrt{nK_{0,1}(n)}\right\}.$$

The asymptotic inequalities above yield $K_{0,1}(n) \ll R_{0,1}(n) \ll n$ and $R_{0,1}(n)^2 K_{0,2}(n) \ll n^2$. The final inequality implies $K_{0,2}(n) \ll \frac{n^2}{R_{0,1}(n)^2}$. We also have $K_{0,2}(n) \ll n$ from condition (C2.1*). Finally, let

$$Q_{0,2}(n) = \max\left\{1, \min\left\{\sqrt{K_{0,2}(n)\frac{n^2}{R_{0,1}(n)^2}}, \sqrt{nK_{0,2}(n)}\right\}\right\}$$

where the minimum is interpreted asymptotically as $n \to \infty$. Then we have $K_{0,2}(n) \ll Q_{0,2}(n) \ll n$ and $R_{0,1}(n)^2 Q_{0,2}(n) \ll n^2$.

We mimic the above strategy and produce pairs of functions $(S_{0,1}(n), Q_{1,2}(n))$ and $(T_{0,1}(n), Q_{2,2}(n))$ such that

- $K_{0,1}(n) \ll S_{0,1}(n) \ll n$, $K_{1,2}(n) \ll Q_{1,2}(n) \ll n$, $\lim_{n\to\infty} Q_{1,2}(n) > 0$ and $S_{0,1}(n)^3 Q_{1,2}(n) \ll n^3$.
- $K_{0,1}(n) \ll T_{0,1}(n) \ll n$, $K_{2,2}(n) \ll Q_{2,2}(n) \ll n$, $\lim_{n\to\infty} Q_{2,2}(n) > 0$ and $T_{0,1}(n)^4 Q_{2,2}(n) \ll n^4$.

Finally, let

$$Q_{0,1}(n) = \max\{1, \min\{R_{0,1}(n), S_{0,1}(n), T_{0,1}(n)\}\},$$

where the minimum is interpreted asymptotically as $n \to \infty$. We now have all the $Q_{i,j}(n)$, and all the desired properties are straightforward to confirm. □

We turn to the main proof of Theorem 5.16. The main proof utilizes some technical results, which will be proven in several lemmas after the main proof.

Proof of Theorem 5.16. Assume that conditions (C2.1*) and (C2.3) hold. We have

$$\mathbb{P}(\delta_{R_n} = 0) = \mathbb{P}(\delta_{R_n} = 0, \cap_{i,j}\{M_{i,j}(n) \le Q_{i,j}(n)\}) + \mathbb{P}(\delta_{R_n} = 0, \cup_{i,j}\{M_{i,j}(n) > Q_{i,j}(n)\}) \tag{5.31}$$

We will show that the second term goes to zero. Since each $M_{i,j}(n)$ has a binomial distribution, we have

$$\begin{aligned}\mathbb{P}(M_{i,j}(n) > Q_{i,j}(n)) &= \mathbb{P}(M_{i,j}(n) - K_{i,j}(n) > Q_{i,j}(n) - K_{i,j}(n)) \\ &\le \frac{\mathrm{Var}(M_{i,j}(n))}{(Q_{i,j}(n) - K_{i,j}(n))^2} \le \frac{K_{i,j}(n)}{(Q_{i,j}(n) - K_{i,j}(n))^2}.\end{aligned}$$

Since $K_{i,j}(n) \ll Q_{i,j}(n)$ and $\lim_{n\to\infty} Q_{i,j}(n) > 0$, we have $\lim_{n\to\infty} \mathbb{P}(M_{i,j}(n) > Q_{i,j}(n)) = 0$ for all (i,j). Thus

$$\lim_{n\to\infty} \mathbb{P}(\cup_{i,j}\{M_{i,j}(n) > Q_{i,j}(n)\}) = 0, \tag{5.32}$$

and consequently,

$$\lim_{n\to\infty} \mathbb{P}(\delta_{R_n} = 0, \cup_{i,j}\{M_{i,j}(n) > Q_{i,j}(n)\}) = 0. \tag{5.33}$$

Now we consider the first term in (5.31). We have

$$\begin{aligned}&\mathbb{P}(\delta_{R_n} = 0, \cap_{i,j}\{M_{i,j}(n) \le Q_{i,j}(n)\}) \\ &\qquad = \sum_{k_{i,j}(n)=0}^{Q_{i,j}(n)} \mathbb{P}(\delta_{R_n} = 0 | \cap_{i,j} \{M_{i,j}(n) = k_{i,j}(n)\}) \mathbb{P}(\cap_{i,j}\{M_{i,j}(n) = k_{i,j}(n)\})\end{aligned}$$

We will prove in Lemma 5.22 below that

$$\mathbb{P}(\delta_{R_n} = 0 | \cap_{i,j} \{M_{i,j}(n) = k_{i,j}(n)\}) \ge 1 - C_3 \frac{Q(n)}{n}, \tag{5.34}$$

where $Q(n)$ is a function satisfying $Q(n) \ll n$ and C_3 is independent from n and $k_{i,j}(n)$. Thus

$$\begin{aligned}\mathbb{P}(\delta_{R_n} = 0, \cap_{i,j}\{M_{i,j}(n) \le Q_{i,j}(n)\}) &\ge \left(1 - C_3 \frac{Q(n)}{n}\right) \sum_{k_{i,j}(n)=0}^{Q_{i,j}} \mathbb{P}(\cap_{i,j}\{M_{i,j}(n) = k_{i,j}(n)\}) \\ &= \left(1 - C_3 \frac{Q(n)}{n}\right) \mathbb{P}(\cap_{i,j}\{M_{i,j}(n) \le Q_{i,j}(n)\}).\end{aligned} \tag{5.35}$$

Equation (5.32) gives us $\lim_{n\to\infty} \mathbb{P}(\cap_{i,j}\{M_{i,j}(n) \leq Q_{i,j}(n)\}) = 1$, thus from (5.35) we have

$$\lim_{n\to\infty} \mathbb{P}(\delta_{R_n} = 0, \cap_{i,j}\{M_{i,j}(n) \leq Q_{i,j}(n)\}) = 1. \tag{5.36}$$

Combining (5.31), (5.33), and (5.36) we have

$$\lim_{n\to\infty} \mathbb{P}(\delta_{R_n} = 0) = 1. \qquad \square$$

To complete this section, we will provide a series of lemmas, eventually leading to Lemma 5.22, which yields the critical bound (5.34)

$$\mathbb{P}(\delta_{R_n} = 0 | \cap_{i,j} \{M_{i,j}(n) = k_{i,j}(n)\}) \geq 1 - C_3 \frac{Q(n)}{n}.$$

First we make an observation about the most probable number of species in realized reactions from each set $E_n^{i,j}$. Note that a reaction in the set $E_n^{0,2}$ can have either one or two distinct species appearing in it. For example, we could have $\emptyset \leftrightharpoons 2S_1$, in which there is only one species, or we could have $\emptyset \leftrightharpoons S_1 + S_2$, in which there are two species. Similarly, reactions from the set $E_n^{1,2}$ can have one, two, or three distinct species, and reactions from the set $E_n^{2,2}$ can have two, three, or four distinct species. The following lemma states that when the number of realized reactions in each set is not too large, as quantified below, then, with probability approaching one as $n \to \infty$, the realized reactions from each set will consist of the maximal number of distinct species.

Lemma 5.18. *Suppose conditions (C2.1*) and (C2.3) hold and that $Q_{i,j}(n)$ are defined as in Lemma 5.17. Suppose further that $k_{i,j}(n) \leq Q_{i,j}(n)$. Let $A_n^{0,2}$, $A_n^{1,2}$, and $A_n^{2,2}$ be the events that the realized reactions in $E_n^{0,2}, E_n^{1,2}, E_n^{2,2}$ all have precisely 2,3, and 4 distinct species respectively. Let $A_n = A_n^{0,2} \cap A_n^{1,2} \cap A_n^{2,2}$ Then*

$$\lim_{n\to\infty} \mathbb{P}(A_n | \cap_{i,j} \{M_{i,j}(n) = k_{i,j}(n)\}) = 1.$$

Moreover, we have the explicit bound

$$\mathbb{P}(A_n | \cap_{i,j} \{M_{i,j}(n) = k_{i,j}(n)\}) \geq \left(1 - \frac{2Q_{0,2}(n)}{n}\right)\left(1 - \frac{4Q_{1,2}(n)}{n}\right)\left(1 - \frac{8Q_{2,2}(n)}{n}\right).$$

Proof. First, consider the reactions in $E_n^{0,2}$, which have the form $\emptyset \leftrightharpoons S_i + S_j$. These reactions have 2 species if and only if $i \neq j$. Recall that $|E_n^{0,2}| = n(n+1)/2$, and there are n reactions of the form $2S_i$. Thus we have

$$\begin{aligned}
&\mathbb{P}(A_n^{0,2} | M_{0,2}(n) = k_{0,2}(n)) \\
&\qquad = \left(1 - \frac{n}{n(n+1)/2}\right)\left(1 - \frac{n}{n(n+1)/2 - 1}\right) \cdots \left(1 - \frac{n}{n(n+1)/2 - k_{0,2}(n) + 1}\right) \\
&\qquad = \left(1 - \frac{2}{n+1}\right)\left(1 - \frac{2}{n+1-\frac{2}{n}}\right) \cdots \left(1 - \frac{2}{n+1-\frac{2(k_{0,2}(n)-1)}{n}}\right) \\
&\qquad \geq \left(1 - \frac{2}{n}\right)^{k_{0,2}(n)} \\
&\qquad \geq 1 - \frac{2k_{0,2}(n)}{n},
\end{aligned} \tag{5.37}$$

where the last inequality is due to Bernoulli's inequality.

Next, consider the reactions in $E_n^{1,2}$. These reactions have less than 3 species if it is either $S_i \leftrightharpoons S_i + S_j$ (where i and j are not necessarily different) or $S_i \to 2S_j$ (where $i \neq j$). It is straightforward to check that there are n^2 reactions of the former type, and there are $n(n-1)$ reactions of the latter type, both of which add up to $n(2n-1)$ reactions in $E_n^{1,2}$ with less than 3 species. Since $|E_n^{1,2}| = \frac{n^2(n+1)}{2}$ we have

$$\begin{aligned}
&\mathbb{P}(A_n^{1,2} | M_{1,2}(n) = k_{1,2}(n)) \\
&\qquad = \left(1 - \frac{n(2n-1)}{n^2(n+1)/2}\right)\left(1 - \frac{n(2n-1)}{n^2(n+1)/2 - 1}\right) \cdots \left(1 - \frac{n(2n-1)}{n^2(n+1)/2 - k_{1,2}(n) + 1}\right) \\
&\qquad \geq \left(1 - \frac{4}{n}\right)^{k_{1,2}(n)} \\
&\qquad \geq 1 - \frac{4k_{1,2}(n)}{n},
\end{aligned} \tag{5.38}$$

where the first inequality here follows a similar argument to the first inequality in (5.37).

Finally, consider the reactions in $E_n^{2,2}$. These reactions have less than 4 species if they have the form $2S_i \rightleftharpoons 2S_j$, $2S_i \rightleftharpoons S_j + S_k$ (where $j \neq k$), or $S_i + S_j \rightleftharpoons S_i + S_k$ (where i, j, k are pairwise different). It is straightforward to check that there are $\frac{n(n-1)}{2}$ reactions of the first type, $n(\frac{n(n+1)}{2} - n)$ reactions of the second type, and $(\frac{n(n+1)}{2} - n)(n-2)$ reactions of the third type. In total, there are $\frac{n(n-1)(2n-1)}{2}$ reactions in $E_n^{2,2}$ with less than 4 species. Since $|E_n^{2,2}| = \binom{\frac{n(n+1)}{2}}{2} = \frac{n(n+1)(n-1)(n+2)}{8}$, we have

$$\begin{aligned}
&\mathbb{P}(A_n^{2,2} | M_{2,2}(n) = k_{2,2}(n)) \\
&\qquad = \left(1 - \frac{\frac{n(n-1)(2n-1)}{2}}{\frac{n(n+1)(n-1)(n+2)}{8}}\right) \cdots \left(1 - \frac{\frac{n(n-1)(2n-1)}{2}}{\frac{n(n+1)(n-1)(n+2)}{8} - k_{2,2}(n) + 1}\right) \\
&\qquad \geq \left(1 - \frac{8}{n}\right)^{k_{2,2}(n)} \\
&\qquad \geq 1 - \frac{8k_{2,2}(n)}{n}, \qquad (5.39)
\end{aligned}$$

where the first inequality here follows a similar argument as the first inequality in (5.37).

From (5.37),(5.38),(5.39), and independence, we have

$$\begin{aligned}
&\mathbb{P}(A_n | \cap_{i,j} \{M_{i,j}(n) = k_{i,j}(n)\})] \\
&\qquad \geq \left(1 - \frac{2k_{0,2}(n)}{n}\right)\left(1 - \frac{4k_{1,2}(n)}{n}\right)\left(1 - \frac{8k_{2,2}(n)}{n}\right) \\
&\qquad \geq \left(1 - \frac{2Q_{0,2}(n)}{n}\right)\left(1 - \frac{4Q_{1,2}(n)}{n}\right)\left(1 - \frac{8Q_{2,2}(n)}{n}\right),
\end{aligned}$$

and the limit follows. $\square$

In our next major lemma, Lemma 5.21, we require the notion of a minimally dependent set, which we define below.

Definition 5.19. *We say a set of vectors is minimally dependent if it is linearly dependent and any of its proper subsets are linearly independent.*

We make a quick observation on minimally dependent set.

Lemma 5.20. *Let M be a matrix whose columns $v_1, v_2, \ldots, v_m$ are minimally dependent. Then M has no row with only one non-zero entry.*

Proof. Since $v_1, \ldots, v_m$ are dependent, there exist constants $\alpha_1, \ldots, \alpha_m$, not all of which are zero, such that

$$\alpha_1 v_1 + \cdots + \alpha_m v_m = 0.$$

Suppose by contradiction that M has a row with only one non-zero entry, and suppose that entry belongs to the ith column. Then this must imply $\alpha_i = 0$. However, this implies that

$$\sum_{j \neq i} \alpha_j v_j = 0,$$

with not all α_j equaling zero. This contradicts the set $\{v_i\}_{i=1}^m$ being minimally dependent.

$\square$

An example related to minimal dependence in the context of reaction network is the network $\emptyset \leftrightharpoons S_1, \emptyset \leftrightharpoons S_2, \emptyset \leftrightharpoons S_3, \emptyset \leftrightharpoons S_1 + S_2$, whose reaction vectors are dependent, but not minimally dependent because the proper subset containing $\emptyset \leftrightharpoons S_1, \emptyset \leftrightharpoons S_2, \emptyset \leftrightharpoons S_1 + S_2$ is dependent. In the next lemma, we will show that for a set of reaction vectors to be minimally dependent, there cannot be too many reactions from $E_n^{0,1}$, relative to the numbers from $E_n^{0,2}, E_n^{1,2}, E_n^{2,2}$.

Lemma 5.21. *Suppose a set V with i_1, i_2, i_3, i_4, i_5 reaction vectors in $E_n^{0,1}, E_n^{1,1}, E_n^{0,2}, E_n^{1,2}, E_n^{2,2}$, respectively, is minimally dependent. Assume further that each of the i_3, i_4, and i_5 reactions from $E_n^{0,2}, E_n^{1,2}, E_n^{2,2}$ have precisely 2,3, and 4 species, respectively, and that*

$i_3 + i_4 + i_5 > 0$. *Then we must have*

$$i_1 \leq 2i_3 + 3i_4 + 4i_5.$$

Proof. Consider a matrix M whose first i_1 columns are the reaction vectors from $V \cap E_n^{0,1}$, the next i_2 columns are the reaction vectors from $V \cap E_n^{1,1}$, the next i_3 columns are the reaction vectors from $E_n^{0,2}$, etc. Let P be the sub-matrix consisting of the first $i_1 + i_2$ columns of M (so it is constructed by the reaction vectors from $V \cap E_n^{0,1}$ followed by the reaction vectors from $V \cap E_n^{1,1}$).

Since V is minimally dependent, Lemma 5.20 tells us that M has no row with only one non-zero entry. Let $z_{i_1+i_2}$ be the number of rows of P with exactly one entry. By construction, the final $i_3 + i_4 + i_5$ columns of M have at most

$$2i_3 + 3i_4 + 4i_5$$

non-zero elements. Therefore, we must have

$$z_{i_1+i_2} \leq 2i_3 + 3i_4 + 4i_5,$$

for otherwise there are not enough non-zero terms in the final $i_3 + i_4 + i_5$ columns to cover the rows of P with a single element. The remainder of the proof just consists of showing that

$$i_1 \leq z_{i_1+i_2}. \tag{5.40}$$

To show that the inequality (5.40) holds, we consider adding the column vectors sequentially, and make the following observations.

1. The first i_1 columns of M can, without loss of generality, be taken to be the canonical vectors $e_1, \ldots, e_{i_1}$. Note, therefore, that the sub-matrix consisting of the first i_1 columns of M has exactly i_1 rows that have a single non-zero entry.

2. The rank of the sub-matrix of P consisting of the first $i_1 + k$ columns must be $i_1 + k$ for any $0 \leq k \leq i_2$, for otherwise there is a dependence and V would not be minimally dependent (here we are explicitly using that $i_3 + i_4 + i_5 > 0$).

3. Consider the action of going from a sub-matrix of P consisting of the first $i_1 + k$ columns to one consisting of the first $i_1 + k + 1$ columns, for $k \leq i_2 - 1$. Since each such sub-matrix is full rank (by the point made above), the addition of the next column vector in the construction must have at least one element in a row that was previously all zeros.

4. Since each column vector being added has at most two elements, the number of rows with a single entry can never decrease.

Hence, we have that the number of rows with precisely one non-zero entry at the end of the construction, $z_{i_1+i_2}$ must be at least as large as the number at the beginning of the construction, i_1, and we are done. □

Finally, we present the main lemma, giving the bound needed for Theorem 5.16. Note that a positive deficiency must imply the existence of a minimally independent set. Thus the main approach of the proof revolves around summing over the probabilities of each certain set of reaction vectors being minimally dependent. The constraint in Lemma 5.21 will play a critical role in this approach.

Lemma 5.22. *Suppose conditions (C2.1*) and (C2.3) hold and that $Q_{i,j}(n)$ are as in Lemma 5.17. Suppose further that $k_{i,j}(n) \leq Q_{i,j}(n)$ for each relevant pair (i, j). Then*

$$\mathbb{P}(\delta_{R_n} = 0 | \cap_{i,j} \{M_{i,j}(n) = k_{i,j}(n)\}) \geq 1 - C_3 \frac{Q(n)}{n},$$

where $Q(n)$ is a function satisfying $Q(n) \ll n$ and C_3 is independent from n and $k_{i,j}(n)$.

Proof. We have

$$
\begin{aligned}
&\mathbb{P}(\delta_{R_n} = 0 | \cap_{i,j} \{M_{i,j}(n) = k_{i,j}(n)\}) \\
&\quad \geq \mathbb{P}(\delta_{R_n} = 0 | A_n, \cap_{i,j}\{M_{i,j}(n) = k_{i,j}(n)\})\mathbb{P}(A_n | \cap_{i,j} \{M_{i,j}(n) = k_{i,j}(n)\}) \\
&\quad = (1 - \mathbb{P}(\delta_{R_n} > 0 | A_n, \cap_{i,j}\{M_{i,j}(n) = k_{i,j}(n)\}))\mathbb{P}(A_n | \cap_{i,j} \{M_{i,j}(n) = k_{i,j}(n)\}).
\end{aligned}
\tag{5.41}
$$

From Lemma 2.13(h) and Lemma 2.13(i), the event $\delta_{R_n} > 0$ must imply there exists a minimally dependent set which consists of at least one reaction from $E_n^{0,2}$, $E_n^{1,2}$, or $E_n^{2,2}$. Let $I = (i_1, i_2, i_3, i_4, i_5)$ be a multi-index. Let $K_n = (k_{0,1}(n), k_{1,1}(n), k_{0,2}(n), k_{1,2}(n), k_{2,2}(n))$. For convenience, we write $I \leq K_n$ to represent $i_1 \leq k_{0,1}(n), \ldots, i_5 \leq k_{2,2}(n)$. Then we have

$$
\begin{aligned}
&\mathbb{P}(\delta_{R_n} > 0 | A_n, \cap_{i,j}\{M_{i,j}(n) = k_{i,j}(n)\}) \\
&\leq \sum_{\substack{I \leq K_n \\ i_3+i_4+i_5>0 \\ i_1 \leq 2i_3+3i_4+4i_5}} \binom{k_{0,1}(n)}{i_1}\binom{k_{1,1}(n)}{i_2}\binom{k_{0,2}(n)}{i_3}\binom{k_{1,2}(n)}{i_4}\binom{k_{2,2}(n)}{i_5}\mathbb{P}(B_I | A_n, \cap_{i,j}\{M_{i,j}(n) = k_{i,j}(n)\}),
\end{aligned}
\tag{5.42}
$$

where B_I is the event that a set with i_1, i_2, i_3, i_4, i_5 realized reactions from $E_n^{0,1}$, $E_n^{1,1}$, $E_n^{0,2}$, $E_n^{1,2}$, $E_n^{2,2}$, which also satisfy A_n, is minimally dependent. Note that the constraint $i_1 \leq 2i_3 + 3i_4 + 4i_5$ comes from Lemma 5.21.

Now we fix an index $I = (i_1, i_2, i_3, i_4, i_5)$ and we fix a particular minimally dependent reaction set V_I with i_1, i_2, i_3, i_4, i_5 reactions in $E_n^{0,1}, E_n^{1,1}, E_n^{0,2}, E_n^{1,2}, E_n^{2,2}$. Let M_I be the matrix whose columns are reaction vectors in V_I. Next, we notice that the total number of non-zero entries in M_I is $i_1 + 2i_2 + 2i_3 + 3i_4 + 4i_5$. Since each non-zero row in M_I must have at least two non-zero entries, the number of non-zero rows is at most $\ell := \lfloor \frac{i_1+2i_2+2i_3+3i_4+4i_5}{2} \rfloor$.

There are $\binom{n}{\ell}$ ways to choose ℓ non-zero rows from n rows. Fix a set of ℓ rows to be non-zero rows. We have the probability that all i_1 reaction vectors in $E_n^{0,1}$ have non-zero entry among these ℓ rows is

$$\frac{\ell}{n}\frac{\ell-1}{n-1}\cdots\frac{\ell-i_1+1}{n-i_1+1} \leq \left(\frac{\ell}{n}\right)^{i_1}.$$

The probability that all i_2 reactions vectors in $E_n^{1,1}$ have non-zero entry among these ℓ rows is

$$\frac{\binom{\ell}{2}}{\binom{n}{2}}\frac{\binom{\ell}{2}-1}{\binom{n}{2}-1}\cdots\frac{\binom{\ell}{2}-i_2+1}{\binom{n}{2}-i_2+1} \leq \left(\frac{\ell}{n}\right)^{2i_2}.$$

Using similar arguments, we have

$$\begin{aligned}
\mathbb{P}(B_I|A_n, \cap_{i,j}\{M_{i,j}(n) = k_{i,j}(n)\}) &\leq \binom{n}{\ell}\left(\frac{\ell}{n}\right)^{i_1+2i_2+2i_3+3i_4+4i_5} \\
&\leq \frac{n^\ell}{\ell!}\left(\frac{\ell}{n}\right)^{\ell+\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}} \\
&\leq \frac{n^\ell}{\ell^\ell e^{-\ell}}\left(\frac{\ell}{n}\right)^{\ell+\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}} \\
&\leq \left(\frac{e\ell}{n}\right)^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}},
\end{aligned} \tag{5.43}$$

where the third inequality is due to the inequality $x! \geq x^x e^{-x}$. Combining (5.42) and

(5.43), we have

$$
\begin{aligned}
&\mathbb{P}(\delta_{R_n} > 0 | A_n, \cap_{i,j}\{M_{i,j}(n) = k_{i,j}(n)\}) \\
\leq & \sum_{\substack{I \leq K_n \\ i_3+i_4+i_5>0 \\ i_1 \leq 2i_3+3i_4+4i_5}} \binom{k_{0,1}(n)}{i_1}\binom{k_{1,1}(n)}{i_2}\cdots\binom{k_{2,2}(n)}{i_5}\left(\frac{e\ell}{n}\right)^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}} \\
\leq & \sum_{\substack{I \leq K_n \\ i_3+i_4+i_5>0 \\ i_1 \leq 2i_3+3i_4+4i_5}} \left(\frac{ek_{0,1}(n)}{i_1}\right)^{i_1}\cdots\left(\frac{ek_{2,2}(n)}{i_5}\right)^{i_5}\left(\frac{e\ell}{n}\right)^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}} \\
\leq & \sum_{\substack{I \leq K_n \\ i_3+i_4+i_5>0 \\ i_1 \leq 2i_3+3i_4+4i_5}} \left(\frac{(5ek_{0,1}(n))^{i_1}\cdots(5ek_{2,2}(n))^{i_5}}{(i_1+i_2+i_3+i_4+i_5)^{i_1+i_2+i_3+i_4+i_5}}\right)\left(\frac{e\ell}{n}\right)^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}} \\
\leq & \sum_{\substack{I \leq K_n \\ i_3+i_4+i_5>0 \\ i_1 \leq 2i_3+3i_4+4i_5}} \left(\frac{(5eQ_{0,1}(n))^{i_1}\cdots(5eQ_{2,2}(n))^{i_5}}{(i_1+i_2+i_3+i_4+i_5)^{i_1+i_2+i_3+i_4+i_5}}\right)\left(\frac{e\ell}{n}\right)^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}},
\end{aligned}
$$

where the second inequality is again due to $x! \geq x^x e^{-x}$ and the third inequality is due to Corollary A.5. Since $\ell = \lfloor \frac{i_1+2i_2+2i_3+3i_4+4i_5}{2} \rfloor \leq 2(i_1+i_2+i_3+i_4+i_5)$, we have

$$
\begin{aligned}
&\mathbb{P}(\delta_{R_n} > 0 | A_n, \cap_{i,j}\{M_{i,j}(n) = k_{i,j}(n)\}) \\
\leq & \sum_{\substack{I \leq K_n \\ i_3+i_4+i_5>0 \\ i_1 \leq 2i_3+3i_4+4i_5}} \frac{(5eQ_{0,1}(n))^{i_1}\cdots(5eQ_{2,2}(n))^{i_5}(i_1+i_2+i_3+i_4+i_5)^{\ell-(i_1+i_2+i_3+i_4+i_5)}}{((2e)^{-1}n)^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}}} \\
\leq & \sum_{\substack{I \leq K_n \\ i_3+i_4+i_5>0 \\ i_1 \leq 2i_3+3i_4+4i_5}} \frac{(5eQ_{0,1}(n))^{i_1}\cdots(5eQ_{2,2}(n))^{i_5}(i_1+i_2+i_3+i_4+i_5)^{-\frac{i_1}{2}+\frac{i_4}{2}+i_5}}{((2e)^{-1}n)^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}}} \\
= & S_n + T_n,
\end{aligned}
\tag{5.44}
$$

where

$$
S_n = \sum_{\substack{I \leq K_n \\ i_3+i_4+i_5>0 \\ i_1 \leq 2i_3+3i_4+4i_5 \\ i_1 \leq i_4+2i_5}} \frac{(5eQ_{0,1}(n))^{i_1}\cdots(5eQ_{2,2}(n))^{i_5}(i_1+i_2+i_3+i_4+i_5)^{-\frac{i_1}{2}+\frac{i_4}{2}+i_5}}{((2e)^{-1}n)^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}}},
$$

consists of the terms with positive exponent for $i_1 + i_2 + i_3 + i_4 + i_5$ and

$$T_n = \sum_{\substack{I \le K_n \\ i_3+i_4+i_5>0 \\ i_1 \le 2i_3+3i_4+4i_5 \\ i_1 > i_4+2i_5}} \frac{(5eQ_{0,1}(n))^{i_1} \cdots (5eQ_{2,2}(n))^{i_5} (i_1+i_2+i_3+i_4+i_5)^{-\frac{i_1}{2}+\frac{i_4}{2}+i_5}}{((2e)^{-1}n)^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}}}$$

consists of the terms with negative exponent for $i_1 + i_2 + i_3 + i_4 + i_5$.

We first deal with T_n, which is the more difficult term to bound. Notice that the exponent $-\frac{i_1}{2} + \frac{i_4}{2} + i_5 < 0$. Therefore we have

$$\begin{aligned} T_n &\le \sum_{\substack{I \le K_n \\ i_3+i_4+i_5>0 \\ i_1 \le 2i_3+3i_4+4i_5 \\ i_1 > i_4+2i_5}} \frac{(5eQ_{0,1}(n))^{i_1} \cdots (5eQ_{2,2}(n))^{i_5}}{((2e)^{-1}n)^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}}} \\ &= \sum_{\substack{I \le K_n \\ i_3+i_4+i_5>0 \\ i_1 \le 2i_3+3i_4+4i_5 \\ i_1 > i_4+2i_5}} \frac{Q_{0,1}(n)^{i_1} \cdots Q_{2,2}(n)^{i_5} (5e)^{i_1+i_2+i_3+i_4+i_5}}{((2e)^{-1}n)^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}}} \\ &\le \sum_{\substack{I \le K_n \\ i_3+i_4+i_5>0 \\ i_1 \le 2i_3+3i_4+4i_5 \\ i_1 > i_4+2i_5}} \frac{Q_{0,1}(n)^{i_1} \cdots Q_{2,2}(n)^{i_5}}{((50e^3)^{-1}n)^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}}}, \end{aligned} \tag{5.45}$$

where the last inequality is due to the fact that $i_1 + i_2 + i_3 + i_4 + i_5 \le 2^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}}$. Let

$$Q(n) = \max\{Q_{i,j}(n), Q_{0,1}(n)Q_{0,2}(n)^{1/2}, Q_{0,1}(n)Q_{1,2}(n)^{1/3}, Q_{0,1}(n)Q_{2,2}(n)^{1/4}\}, \tag{5.46}$$

where the maximum is interpreted asymptotically as $n \to \infty$. From the way we construct $Q_{i,j}(n)$ in Lemma 5.17 we have $Q(n) \ll n$. Next, we split $Q_{0,1}(n)^{i_1}$ into the product of three terms and distribute them into $Q_{0,2}(n), Q_{1,2}(n)$, and $Q_{2,2}(n)$. We have

$$Q_{0,1}(n)^{i_1 \frac{2i_3}{2i_3+3i_4+4i_5}} Q_{0,2}(n)^{\frac{i_1}{2} \frac{2i_3}{2i_3+3i_4+4i_5}} \le Q(n)^{i_1 \frac{2i_3}{2i_3+3i_4+4i_5}},$$

$$Q_{0,1}(n)^{i_1 \frac{3i_4}{2i_3+3i_4+4i_5}} Q_{1,2}(n)^{\frac{i_1}{3} \frac{3i_4}{2i_3+3i_4+4i_5}} \le Q(n)^{i_1 \frac{3i_4}{2i_3+3i_4+4i_5}},$$

and

$$Q_{0,1}(n)^{i_1 \frac{4i_5}{2i_3+3i_4+4i_5}} Q_{2,2}(n)^{\frac{i_1}{4} \frac{4i_5}{2i_3+3i_4+4i_5}} \leq Q(n)^{i_1 \frac{4i_5}{2i_3+3i_4+4i_5}}.$$

Multiplying these inequalities together, we have

$$Q_{0,1}(n)^{i_1} Q_{0,2}(n)^{i_1 \frac{i_3}{2i_3+3i_4+4i_5}} Q_{1,2}(n)^{i_1 \frac{i_4}{2i_3+3i_4+4i_5}} Q_{2,2}(n)^{i_1 \frac{i_5}{2i_3+3i_4+4i_5}} \leq Q(n)^{i_1}.$$

Note that in (5.45), $Q_{0,2}(n)$ has an exponent of i_3. Notice further that $i_1 \frac{i_3}{2i_3+3i_4+4i_5} \leq i_3$, since $i_1 \leq 2i_3 + 3i_4 + 4i_5$. Thus we have

$$Q_{0,2}(n)^{i_3 - i_1 \frac{i_3}{2i_3+3i_4+4i_5}} \leq Q(n)^{i_3 - i_1 \frac{i_3}{2i_3+3i_4+4i_5}}.$$

Similarly, we have

$$Q_{1,2}(n)^{i_4 - i_1 \frac{i_4}{2i_3+3i_4+4i_5}} \leq Q(n)^{i_4 - i_1 \frac{i_4}{2i_3+3i_4+4i_5}}$$

and

$$Q_{2,2}(n)^{i_5 - i_1 \frac{i_5}{2i_3+3i_4+4i_5}} \leq Q(n)^{i_5 - i_1 \frac{i_5}{2i_3+3i_4+4i_5}}.$$

Therefore we have

$$\begin{aligned} Q_{0,1}(n)^{i_1} \cdots Q_{2,2}(n)^{i_5} &\leq Q(n)^{i_1+i_2} Q(n)^{i_3+i_4+i_5 - i_1 \frac{i_3+i_4+i_5}{2i_3+3i_4+4i_5}} \\ &= Q(n)^{i_1+i_2+i_3+i_4+i_5 - i_1 \frac{i_3+i_4+i_5}{2i_3+3i_4+4i_5}}. \end{aligned} \tag{5.47}$$

Note that $i_1 \leq 2i_3 + 3i_4 + 4i_5$, thus

$$i_1 + i_2 + i_3 + i_4 + i_5 - i_1 \frac{i_3 + i_4 + i_5}{2i_3 + 3i_4 + 4i_5} \leq \frac{i_1 + 2i_2 + 2i_3 + 3i_4 + 4i_5}{2}, \tag{5.48}$$

where the inequality above can be verified in a straightforward manner. Combining

(5.45),(5.47), and (5.48), and noting that $i_3+i_4+i_5>0$, we have

$$
\begin{aligned}
T_n &\le \sum_{\substack{I\le K_n\\ i_3+i_4+i_5>0\\ i_1\le 2i_3+3i_4+4i_5\\ i_1>i_4+2i_5}} \left(\frac{Q(n)}{(50e^3)^{-1}n}\right)^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}}\\
&\le \frac{Q(n)}{(50e^3)^{-1}n}\sum_{i_1=0}^{\infty}\left(\frac{Q(n)}{(50e^3)^{-1}n}\right)^{i_1/2}\cdots\sum_{i_5=0}^{\infty}\left(\frac{Q(n)}{(50e^3)^{-1}n}\right)^{2i_5}\\
&\le C_1\frac{Q(n)}{n},
\end{aligned}
\tag{5.49}
$$

where the second inequality is due to the fact that $i_3+i_4+i_5>0$. Since each sum on the right hand side is bounded by 2 for n large enough, the constant C_1 is independent from n and $k_{i,j}(n)$.

Next we consider S_n. Recall that $i_1\le k_{0,1}(n)\le Q_{0,1}(n),\ldots,i_5\le k_{2,2}(n)\le Q_{2,2}(n)$, implying $i_1,\ldots,i_5\le Q(n)$. Therefore we have

$$
\begin{aligned}
S_n &\le \sum_{\substack{I\le K_n\\ i_3+i_4+i_5>0\\ i_1\le 2i_3+3i_4+4i_5\\ i_1\le i_4+2i_5}} \frac{(5eQ(n))^{i_1+i_2+i_3+i_4+i_5-\frac{i_1}{2}+\frac{i_4}{2}+i_5}}{((2e)^{-1}n)^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}}}\\
&= \sum_{\substack{I\le K_n\\ i_3+i_4+i_5>0\\ i_1\le 2i_3+3i_4+4i_5\\ i_1\le i_4+2i_5}} \frac{(5eQ(n))^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}}}{((2e)^{-1}n)^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}}}\\
&\le \frac{5eQ(n)}{(2e)^{-1}n}\sum_{i_1=0}^{\infty}\left(\frac{5eQ(n)}{(2e)^{-1}n}\right)^{i_1/2}\cdots\sum_{i_5=0}^{\infty}\left(\frac{5eQ(n)}{(2e)^{-1}n}\right)^{2i_5}\\
&\le C_2\frac{Q(n)}{n},
\end{aligned}
\tag{5.50}
$$

where C_2 is independent from n and $k_{i,j}(n)$. From (5.44), (5.49), (5.50), we have

$$
\mathbb{P}(\delta_{R_n}>0|A_n,\cap_{i,j}\{M_{i,j}(n)=k_{i,j}(n)\})\le C_1\frac{Q(n)}{n}+C_2\frac{Q(n)}{n}. \tag{5.51}
$$

From Lemma 5.18 and the fact that $Q_{i,j}(n) \le Q(n)$, we have

$$\mathbb{P}(A_n | \cap_{i,j} \{M_{i,j}(n) = k_{i,j}(n)\}) \ge \left(1 - \frac{2Q(n)}{n}\right)\left(1 - \frac{4Q(n)}{n}\right)\left(1 - \frac{8Q(n)}{n}\right). \quad (5.52)$$

Plugging (5.51) and (5.52) into (5.41) yields

$$\begin{aligned}
&\mathbb{P}(\delta_{R_n} = 0 | \cap_{i,j} \{M_{i,j}(n) = k_{i,j}(n)\}) \\
&\ge \left(1 - C_1\frac{Q(n)}{n} - C_2\frac{Q(n)}{n}\right)\left(1 - \frac{2Q(n)}{n}\right)\left(1 - \frac{4Q(n)}{n}\right)\left(1 - \frac{8Q(n)}{n}\right) \\
&\ge 1 - C_3\frac{Q(n)}{n}, \qquad (5.53)
\end{aligned}$$

where the last inequality is obtained from repeatedly applying $(1-a)(1-b) \ge 1-a-b$ (where $a, b \ge 0$). Clearly we must have C_3 independent from n and $k_{i,j}(n)$. □

5.4.3 The threshold function for deficiency zero

In this section, we provide an algorithm to find the threshold function $r(n)$ for deficiency zero for a given set of $\{\alpha_{i,j}\}$. Specifically, $r(n)$ will satisfy

1. $\lim_{n\to\infty} \mathbb{P}(\delta_{R_n} = 0) = 0$ for $\lim_{n\to\infty} \frac{p_n}{r(n)} = \infty$, and
2. $\lim_{n\to\infty} \mathbb{P}(\delta_{R_n} = 0) = 1$ for $\lim_{n\to\infty} \frac{p_n}{r(n)} = 0$.

From Remark 5.10, we have $K_{i,j}(n) \sim n^{i+j}n^{\alpha_{i,j}}p_n = n^{i+j+\alpha_{i,j}}p_n$. Moreover, from Section 5.4.1 and 5.4.2, we have sets of conditions on the $K_{i,j}(n)$ that determine when a network does or does not have a deficiency of zero. Combining these yields the following theorem. In the theorem below, note that the equations 1-3 correspond to condition (C1.1) (and (C2.1)), the equations 4-7 correspond to condition (C1.2) (and (C2.2)), and the equations 8-10 correspond to condition (C1.3) (and (C2.3)).

Theorem 5.23. *Given a set of parameters* $\{\alpha_{i,j}\}$*, consider the following systems where we solve for* $\{r_i(n)\}$

1. $n^{2+\alpha_{0,2}} r_1(n) = n.$

2. $n^{3+\alpha_{1,2}} r_2(n) = n.$

3. $n^4 r_3(n) = n.$

4. $n^{2+\alpha_{1,1}} r_4(n) = n.$

5. $n^{2+\alpha_{0,2}} r_5(n) = 1.$

6. $n^{3+\alpha_{1,2}} r_6(n) = 1.$

7. $n^4 r_7(n) = 1.$

8. $n^{4+2\alpha_{0,1}+\alpha_{0,2}} r_8(n)^3 = n^2.$

9. $n^{6+3\alpha_{0,1}+\alpha_{1,2}} r_9(n)^4 = n^3.$

10. $n^{8+4\alpha_{0,1}} r_{10}(n)^5 = n^4.$

Then the threshold function is

$$r(n) = \min\{r_1(n), r_2(n), r_3(n), \max\{r_4(n), \min\{r_5(n), r_6(n), r_7(n)\}\}, r_8(n), r_9(n), r_{10}(n)\},$$

where the maximum and minimum are interpreted asymptotically as $n \to \infty$.

Proof. If $\lim_{n\to\infty} \frac{p_n}{r(n)} = \infty$, then it is easy to show that at least one condition in Theorem 5.11 is satisfied. Similarly, if $\lim_{n\to\infty} \frac{p_n}{r(n)} = 0$, then all conditions in Theorem 5.16 are satisfied. $\square$

Example 15 (A closed system with $\alpha_{0,1} = \alpha_{0,2} = 0, \alpha_{1,1} = 2, \alpha_{1,2} = 1$)**.** In this case, we have $K_{0,1}(n) \sim np_n$, $K_{0,2}(n) \sim n^2 p_n$, $K_{1,1}(n) \sim K_{1,2}(n) \sim K_{2,2}(n) \sim n^4 p_n$. Using Theorem 5.23 yields

$$r(n) = \frac{1}{n^3},$$

which is the same threshold as in the base case in Section 5.2.

Example 16 (An open system with $\alpha_{0,1} = 3$, $\alpha_{1,1} = \alpha_{0,2} = 2$, $\alpha_{1,2} = 1$)**.** . In this case, we have $K_{i,j}(n) \sim n^4 p_n$ for all (i, j). Using Theorem 5.23 yields

$$r(n) = \frac{1}{n^{10/3}},$$

which is a lower threshold than the previous case with a closed system. Intuitively, the inflow and outflow reactions make it easier to break deficiency zero of a reaction network.

Appendix A

Appendix

The following lemmas have been used in the manuscript. Their proofs are added for completeness.

Lemma A.1. *Here we need to provide an asymptotic estimate as $C \to \infty$ of the form*

$$\ln \sum_{x \in \mathbb{Z}^n_{\geq 0}} \frac{C^x}{(x!)^d} \sim \sum_n (d_i C_i^{1/d_i} + a \ln C_i + b)$$

where a, b are constants that do not depend on C.

Proof. When $n = 1$, by Stirling estimation (and ignoring the factor of $\sqrt{2\pi}$), we have

$$(x!)^d \sim \left(\sqrt{x}\frac{x^x}{e^x}\right)^d = \sqrt{xd}\frac{(xd)^{xd}}{e^{xd}d^{xd}}x^{(d-1)/2} \sim \Gamma(xd+1)\frac{x^{(d-1)/2}}{d^{xd}} \sim \frac{\Gamma(xd+1+(d-1)/2)}{d^{xd}},$$

where the last estimation is due to the fact that $\lim_{n\to\infty} \frac{\Gamma(n+\alpha)}{n^\alpha \Gamma(n)} = 1$.

Thus

$$\sum_{x \in \mathbb{Z}} \frac{C^x}{(x!)^d} \sim \sum_{x \in \mathbb{Z}} \frac{(Cd^d)^x}{\Gamma(xd + (d+1)/2)}. \tag{A.1}$$

The asymptotic behavior of the right hand side in (A.1) can be found in Example 2.3.1 of [33]. In particular, its asymptotic character is exponential since we are considering C having real values only

$$\sum_{x \in \mathbb{Z}} \frac{(Cd^d)^x}{\Gamma(xd + (d+1)/2)} \sim \frac{1}{d}(Cd^d)^{(1-d)/2d} e^{(Cd^d)^{1/d}} = cC^{(1-d)/2d} e^{dC^{1/d}}$$

where c is some constant depending on d. Thus, taking log we have

$$\ln\left(\sum_{x\in\mathbb{Z}}\frac{C^x}{(x!)^d}\right)\sim dC^{1/d}+a\ln C+b$$

where a, b are some constants depending on d.

When $n>1$, we have

$$\ln\left(\sum_{x\in\mathbb{Z}^n}\frac{C^x}{(x!)^d}\right)=\ln\left(\prod_{n}\sum_{x\in\mathbb{Z}}\frac{C_i^{x_i}}{(x_i!)^{d_i}}\right)\sim\sum_{n}(d_iC_i^{1/d_i}+a\ln C_i+b),$$

where we have applied the $n=1$ case in the final step. □

In the following lemma, for sequences a_V and b_V we write $a_V\sim b_V$, as $V\to\infty$, for

$$\lim_{V\to\infty}(a_V-b_V)=0.$$

Lemma A.2. *Let θ_i satisfy Assumption 1. Then for a fixed $c\in\mathbb{Z}^n_{>0}$,*

$$\frac{1}{V}\ln\left(\sum_{x\in\mathbb{Z}^n}\frac{(V^dc)^x}{\prod_{i=1}^n\theta_i(1)\cdots\theta_i(x_i)}\right)\sim\frac{1}{V}\ln\left(\sum_{x\in\mathbb{Z}^n}\frac{(V^dc)^x}{\prod_{i=1}^n A_i^{x_i}(x_i!)^{d_i}}\right)=\frac{1}{V}\ln\left(\sum_{x\in\mathbb{Z}^n}\frac{(V^dcA^{-1})^x}{\prod_{i=1}^n(x_i!)^{d_i}}\right),$$

as $V\to\infty$.

Proof. Let first consider the case when $n=1$.

Let $\alpha>0$. From Assumption 1 we have $\lim_{x\to\infty}\frac{\theta(x)}{x^d}=A_i$. Therefore, there exists an $N>0$ such that if $x>N$, then

$$A-\alpha A<\frac{\theta(x)}{x^d}<A+\alpha A$$

which is equivalent to

$$1-\alpha<\frac{\theta(x)}{Ax^d}<1+\alpha. \tag{A.2}$$

Consider

$$\begin{aligned}\frac{1}{V}\ln\left(\sum_{x\in\mathbb{Z}}\frac{(V^dc)^x}{\theta(1)\cdots\theta(x)}\right)-\frac{1}{V}\ln\left(\sum_{x\in\mathbb{Z}}\frac{(V^dc)^x}{A^x(x!)^d}\right)\\ =&\frac{1}{V}\ln\left(\frac{\sum_{x\leq N}\frac{(V^dc)^x}{\theta(1)\cdots\theta(x)}+\sum_{x>N}\frac{(V^dc)^x}{\theta(1)\cdots\theta(x)}}{\sum_{x\in\mathbb{Z}}\frac{(V^dc)^x}{A^x(x!)^d}}\right)\\ =&\frac{1}{V}\ln\left(\frac{R+S}{T}\right)\end{aligned}$$

where

$$R=\sum_{x\leq N}\frac{(V^dc)^x}{\theta(1)\cdots\theta(x)}=O(V^{dN})$$
$$T=\sum_{x\in\mathbb{Z}}\frac{(V^dc)^x}{A^x(x!)^d}=O(e^{d(V^dc/A)^{1/d}})=O(e^{dV(c/A)^{1/d}}),$$

where the sum in R is over all $x\in\mathbb{Z}_{\geq 0}$ and the estimation of T is again based on Lemma A.1.

We have

$$\begin{aligned}S&=\sum_{x>N}\frac{(V^dc)^x}{\theta(1)\cdots\theta(x)}\\ &=\sum_{x>N}\frac{(V^dc)^x}{A^x(x!)^d}\frac{A^x(x!)^d}{\theta(1)\cdots\theta(x)}\\ &=\frac{A^N(N!)^d}{\theta(1)\cdots\theta(N)}\sum_{x>N}\frac{(V^dc)^x}{A^x(x!)^d}\frac{A(N+1)^{d_i}\cdots A(x)^d}{\theta(N+1)\cdots\theta(x)}\\ &=c_N\sum_{x>N}\frac{(V^dc)^x}{A^x(x!)^d}\frac{A(N+1)^d\cdots A(x)^d}{\theta(N+1)\cdots\theta(x)}.\end{aligned}$$

Using (A.2), we have

$$c_N\sum_{x>N}\frac{(V^dc)^x}{A^x(x!)^d}\frac{1}{(1+\alpha)^{x-N}}<S<c_N\sum_{x>N}\frac{(V^dc)^x}{A^x(x!)^d}\frac{1}{(1-\alpha)^{x-N}}.$$

Thus

$$c_N \sum_{x>N} \frac{(V^d c(1+\alpha)^{-1})^x}{A^x (x!)^d} < S < c_N \sum_{x>N} \frac{(V^d c(1-\alpha)^{-1})^x}{A^x (x!)^d}$$

where $LHS = O(e^{dV(c/A(1+\alpha))^{1/d}})$ and $RHS = O(e^{dV(c/A(1-\alpha))^{1/d}})$ similar to how we estimated T. This, together with the fact $R \ll S$ for V large enough, gives us

$$\frac{1}{V} \ln \left(a_N \frac{e^{dV(c/A)^{1/d}}}{e^{dV(c/A(1+\alpha))^{1/d}}} \right) < \frac{1}{V} \ln \left(\frac{R+S}{T} \right) < \frac{1}{V} \ln \left(b_N \frac{e^{dV(c/A)^{1/d}}}{e^{dV(c/A(1-\alpha))^{1/d}}} \right)$$

where a_N, b_N are some constants depending on N. Thus

$$\frac{\ln a_N}{V} + d((c/A)^{1/d} - (c/A(1+\alpha))^{1/d}) < \frac{1}{V} \ln \left(\frac{R+S}{T} \right) < \frac{\ln b_N}{V} + d((c/A)^{1/d} - (c/A(1-\alpha))^{1/d}).$$

Now for an $\epsilon > 0$, pick α such that

$$\max\{|d((c/A)^{1/d} - (c/A(1+\alpha))^{1/d})|, |d((c/A)^{1/d} - (c/A(1-\alpha))^{1/d})|\} < \frac{\epsilon}{2}.$$

Then we pick V large enough so that

$$\max\{|a_N/V\big|, |b_N/V|\} < \frac{\epsilon}{2},$$

then

$$\left| \frac{1}{V} \ln \left(\frac{R+S}{T} \right) \right| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Thus $\frac{1}{V} \ln \left(\frac{R+S}{T} \right) \to 0$ as $V \to \infty$ and we have finished the case $n = 1$.

For $n > 1$, we have

$$\begin{aligned}
\frac{1}{V}\ln\left(\sum_{x\in\mathbb{Z}^n}\frac{(V^d c)^x}{\prod_{i=1}^n \theta_i(1)\cdots\theta_i(x_i)}\right) &= \frac{1}{V}\ln\left(\prod_{i=1}^n\sum_{x_i\in\mathbb{Z}}\frac{(V_i^{d_i c_i})^{x_i}}{\theta_i(1)\cdots\theta_i(x_i)}\right)\\
&= \frac{1}{V}\sum_{i=1}^n\ln\left(\sum_{x_i\in\mathbb{Z}}\frac{(V_i^{d_i c_i})^{x_i}}{\theta_i(1)\cdots\theta_i(x_i)}\right)\\
&\sim \frac{1}{V}\sum_{i=1}^n\ln\left(\sum_{x_i\in\mathbb{Z}}\frac{(V_i^{d_i c_i})^{x}}{A_i^{x_i}(x_i!)^{d_i}}\right)\\
&= \frac{1}{V}\ln\left(\prod_{i=1}^n\sum_{x_i\in\mathbb{Z}}\frac{(V_i^{d_i c_i})^{x}}{A_i^{x_i}(x_i!)^{d_i}}\right)\\
&= \frac{1}{V}\ln\left(\sum_{x\in\mathbb{Z}^n}\frac{(V^d c)^x}{\prod_{i=1}^n A_i^{x_i}(x_i!)^{d_i}}\right).
\end{aligned}$$

Note that here the asymptotic analysis still holds after finite addition, since the asymptotic relation we prove for the case $n = 1$ is slightly stronger than the usual definition of asymptotic. □

Lemma A.3. *Let $X \sim B(n,p)$. Then we have*

$$\mathbb{E}\left[\frac{1}{X+1}\right] \le \frac{1}{np}, \quad \text{and} \quad \mathbb{E}\left[\frac{1}{(X+1)(X+2)}\right] \le \frac{1}{(np)^2}.$$

Proof. We have

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{X+1}\right] &= \sum_{i=0}^n \frac{1}{i+1}\binom{n}{i}p^i(1-p)^{n-i} = \sum_{i=0}^n \frac{n!}{(i+1)!(n-i)!}p^i(1-p)^{n-i}\\
&= \frac{1}{n+1}\frac{1}{p}\sum_{i=0}^n\binom{n+1}{i+1}p^{i+1}(1-p)^{n-i} \le \frac{1}{np}(p+1-p)^{n+1} \le \frac{1}{np}.
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{(X+1)(X+2)}\right] &= \sum_{i=0}^n \frac{1}{(i+1)(i+2)}\binom{n}{i}p^i(1-p)^{n-i} = \sum_{i=0}^n \frac{n!}{(i+2)!(n-i)!}p^i(1-p)^{n-i}\\
&= \frac{1}{(n+1)(n+2)}\frac{1}{p^2}\sum_{i=0}^n\binom{n+2}{i+2}p^{i+2}(1-p)^{n-i}\\
&\le \frac{1}{(np)^2}(p+1-p)^{n+2} \le \frac{1}{(np)^2}.
\end{aligned}$$

∎

Lemma A.4. *Let $x, y \in \mathbb{R}_{\geq 0}$. Then we have*

$$(2x)^x(2y)^y \geq (x+y)^{x+y}$$

Proof. Clearly the inequality holds when either $x = 0$ or $y = 0$ or both. Suppose $x > 0$ and $y > 0$. We have

$$(2x)^x(2y)^y \geq (x+y)^{x+y} \iff 2^{x+y}\left(\frac{x}{y}\right)^x \geq \left(1+\frac{x}{y}\right)^{x+y} \iff 2^{1+\frac{x}{y}}\left(\frac{x}{y}\right)^{x/y} \geq \left(1+\frac{x}{y}\right)^{1+x/y}.$$

Thus the inequality holds if we have $2^{1+t}t^t \geq (1+t)^{1+t}$, or $(1+t)\ln(2) + t\ln(t) \geq (1+t)\ln(1+t)$ for $t > 0$. Let

$$f(t) = (1+t)\ln(2) + t\ln(t) - (1+t)\ln(1+t).$$

A quick calculation shows $f'(t) = \ln(2t) - \ln(1+t)$, and $f(t)$ has a global minimum at $t = 1$. Thus $f(t) \geq f(1) = 0$, which concludes the proof of the Lemma. ∎

Corollary A.5. *Let $x_1, x_2, \ldots, x_n \in \mathbb{R}_{\geq 0}$, then we have*

$$\prod_{i=1}^{n}(nx_i)^{x_i} \geq \left(\sum_{i=1}^{n} x_i\right)^{\sum_{i=1}^{n} x_i}. \tag{A.3}$$

Proof. We will prove the corollary by induction. Clearly (A.3) holds for $n = 1$. Lemma A.4 shows that (A.3) holds for $n = 2$. Suppose (A.3) holds for $n = k$. It suffices to show that (A.3) holds for $n = 2k$ and $n = k - 1$.

First, we will show that (A.3) holds for $n = 2k$. Applying the inductive hypothesis for the $n = k$ terms $x_1, \ldots, x_k$ and the $n = k$ terms $x_{k+1}, \ldots, x_{2k}$, and then applying Lemma A.4 yields

$$\prod_{i=1}^{2k}(2kx_i)^{2x_i} \geq \left(\sum_{i=1}^{k} 2x_i\right)^{\sum_{i=1}^{k} 2x_i}\left(\sum_{i=k+1}^{2k} 2x_i\right)^{\sum_{i=k+1}^{2k} 2x_i} \geq \left(\sum_{i=1}^{2k} x_i\right)^{2\sum_{i=1}^{2k} x_i}.$$

Taking square root of the inequality above gives us the case $n = 2k$.

Next, we will show that (A.3) holds for $n = k-1$. Applying the induction hypothesis for the $n = k$ terms $x_1, \ldots, x_{k-1}, \frac{1}{k-1}\sum_{i=1}^{k-1} x_i$, we have

$$\prod_{i=1}^{k-1}(kx_i)^{x_i}\left(\frac{k}{k-1}\sum_{i=1}^{k-1}x_i\right)^{\frac{1}{k-1}\sum_{i=1}^{k-1}x_i} \geq \left(\sum_{i=1}^{k-1}x_i + \frac{1}{k-1}\sum_{i=1}^{k-1}x_i\right)^{\sum_{i=1}^{k-1}x_i+\frac{1}{k-1}\sum_{i=1}^{k-1}x_i}$$
$$\Rightarrow \prod_{i=1}^{k-1}(kx_i)^{x_i}\left(\frac{k}{k-1}\sum_{i=1}^{k-1}x_i\right)^{\frac{1}{k-1}\sum_{i=1}^{k-1}x_i} \geq \left(\frac{k}{k-1}\sum_{i=1}^{k-1}x_i\right)^{\frac{k}{k-1}\sum_{i=1}^{k-1}x_i}$$
$$\Rightarrow \prod_{i=1}^{k-1}(kx_i)^{x_i} \geq \left(\frac{k}{k-1}\sum_{i=1}^{k-1}x_i\right)^{\sum_{i=1}^{k-1}x_i} \Rightarrow \prod_{i=1}^{k-1}((k-1)x_i)^{x_i} \geq \left(\sum_{i=1}^{k-1}x_i\right)^{\sum_{i=1}^{k-1}x_i}.$$

Thus we have show that (A.3) holds for $n = k-1$, which concludes the proof of the Corollary. $\square$

Bibliography

[1] A. Agazzi, A. Dembo, and J.-P. Eckmann, *Large deviations theory for markov jump models of chemical reaction networks*, Annals of Applied Probability, 28 (2018), pp. 1821–1855.

[2] A. Agazzi, A. Dembo, and J.-P. Eckmann, *On the geometry of chemical reaction networks: Lyapunov function and large deviations*, Journal of Statistical Physics, 172 (2018), pp. 321–352.

[3] D. F. Anderson, *Boundedness of trajectories for weakly reversible, single linkage class reaction systems*, Journal of Mathematical Chemistry, 49 (2011), p. 2275.

[4] D. F. Anderson, *A proof of the global attractor conjecture in the single linkage class case*, SIAM Journal on Applied Mathematics, 71 (2011), pp. 1487–1508.

[5] D. F. Anderson, D. Cappelletti, J. Kim, and T. D. Nguyen, *Tier structure of strongly endotactic reaction networks*, Stochastic Processes and their Applications, 130 (2020), pp. 7218–7259.

[6] D. F. Anderson, D. Cappelletti, M. Koyama, and T. G. Kurtz, *Non-explosivity of stochastically modeled reaction networks that are complex balanced*, Bulletin of Mathematical Biology, 80 (2018), pp. 2561–2579.

[7] D. F. Anderson, G. Cracian, M. Gopalkrishnan, and C. Wiuf, *Lyapunov functions, stationary distributions, and non-equilibrium potential for reaction networks*, Bulletin of Mathematical Biology, 77 (2015), pp. 1744 – 1767.

[8] D. F. Anderson, G. Craciun, and T. G. Kurtz, *Product-form stationary distributions for deficiency zero chemical reaction networks*, Bulletin of Mathematical Biology, 72 (2010), pp. 1947–1970.

[9] D. F. Anderson and J. Kim, *Some network conditions for positive recurrence of stochastically modeled reaction networks*, SIAM Journal of Applied Mathematics, 78 (2018), pp. 2692–2713.

[10] D. F. Anderson and T. G. Kurtz, *Continuous time Markov chain models for chemical reaction networks*, Springer, 2011, ch. 1 in Design and Analysis of Biomolecular Circuits: Engineering Approaches to Systems and Synthetic Biology.

[11] D. F. Anderson and T. G. Kurtz, *Stochastic analysis of biochemical systems*, Springer, 2015.

[12] D. F. Anderson and T. D. Nguyen, *Prevalence of deficiency zero reaction networks in an erdos-renyi framework*, Submitted, arXiv: https://arxiv.org/abs/1910.12723, (2019).

[13] D. F. Anderson and T. D. Nguyen, *Results on stochastic reaction networks with non-mass action kinetics*, Mathematical Biosciences and Engineering, 16 (2019), pp. 2118–2140.

[14] D. F. Anderson and T. D. Nguyen, *Deficiency zero for random reaction networks under a stochastic block model framework*, Submitted, arXiv: https://arxiv.org/abs/2010.07201, (2020).

[15] B. Bollobás, *Introduction to Random Graphs*, Cambride University Press, 2001.

[16] D. Cappelletti and C. Wiuf, *Product-form poisson-like distributions and complex balanced reaction systems*, SIAM Journal on Applied Mathematics, 76 (2014), p. 411–432.

[17] C. Chan, X. Liu, L. Wang, L. Bardwell, Q. Nie, and G. Enciso, *Protein scaffolds can enhance the bistability of multisite phosphorylation systems*, PLoS Comput Biol, 8 (2012), pp. 1–9.

[18] G. Cracian, F. Nazarov, and C. Pantea, *Persistence and permanence of mass-action and power-law dynamical systems*, SIAM Journal on Applied Mathematics, 73 (2013), pp. 305–329.

[19] G. Craciun, *Toric differential inclusions and a proof of the global attractor conjecture*. arXiv preprint: https://arxiv.org/pdf/1501.02860.pdf, 2016.

[20] P. Erdős and A. Rényi, *On the evolution of random graphs*, in Publication of the Mathematical Institute of the Hungarian Academy of Sciences, 1960, pp. 17–61.

[21] M. Feinberg, *Complex balancing in general kinetic systems*, Archive for Rational Mechanics and Analysis, 49 (1972), pp. 187–194.

[22] A. Frieze and M. Karoński, *Random Graphs*, Cambride University Press, 2016.

[23] G. Gnacadja, *Univalent positive polynomial maps and the equilibrium state of chemical networks of reversible binding reactions*, Advances in Applied Mathematics, 43 (2009), pp. 394–414.

[24] M. Gopalkrishnan, E. Miller, and A. Shiu, *A geometric approach to the global attractor conjecture*, SIAM Journal on Applied Dynamical Systems, 13 (2014), pp. 758–797.

[25] P. Holland, K. B. Laskey, and S. Leinhardt, *Stochastic blockmodels: First steps*, Social Networks, 5 (1983), pp. 109–137.

[26] F. Horn, *Necessary and sufficient conditions for complex balancing in chemical kinetics*, Archive for Rational Mechanics and Analysis, 49 (1972), pp. 172–186.

[27] F. Horn, *Stability and complex balancing in mass-action systems with three short complexes*, Proceedings of the Royal Society of London, 334 (1973), pp. 331–342.

[28] F. Horn and R. Jackson, *General mass action kinetics*, Archive for Rational Mechanics and Analysis, 47 (1972), pp. 187–194.

[29] H.-W. Kang, L. Zheng, , and H. G. Othmer, *A new method for choosing the computational cell in stochastic reaction–diffusion systems*, Journal of mathematical biology, 65 (2012), pp. 1017–1099.

[30] F. P. Kelly, *Reversibility and stochastic networks*, J. Wiley, 1979.

[31] T. G. Kurtz, *Strong approximation theorems for density dependent Markov chains*, Stoch. Proc. Appl., 6 (1977/78), pp. 223–240.

[32] T. G. Kurtz, *Representations of markov processes as multparameter time changes*, Ann. Prob., 8 (1980), pp. 682–715.

[33] R. B. Paris and A. D. Wood, *Asymptotics of high order differential equations*, Pitman Research Notes in Mathematics Series, 1986.

[34] E. D. Sontag, *Structure and stability of certain chemical networks and applications to the kinetic proofreading of t-cell receptor signal transduction*, IEEE Trans. Auto. Cont., 46 (2001), pp. 1028–1047.

[35] P. Whittle, *Systems in stochastic equilibrium*, J. Wiley, 1986.