

Deficiency zero for random reaction networks under a stochastic block model framework

David F. Anderson*, Tung D. Nguyen†

July 19, 2021

Abstract

Deficiency zero is an important network structure and has been the focus of many celebrated results within reaction network theory. In our previous paper *Prevalence of deficiency zero reaction networks in an Erdős-Rényi framework*, we provided a framework to quantify the prevalence of deficiency zero among randomly generated reaction networks. Specifically, given a randomly generated binary reaction network with n species, with an edge between two arbitrary vertices occurring independently with probability p_n , we established the threshold function $r(n) = \frac{1}{n^3}$ such that the probability of the random network being deficiency zero converges to 1 if $\frac{p_n}{r(n)} \rightarrow 0$ and converges to 0 if $\frac{p_n}{r(n)} \rightarrow \infty$, as $n \rightarrow \infty$.

With the base Erdős-Rényi framework as a starting point, the current paper provides a significantly more flexible framework by weighting the edge probabilities via control parameters $\alpha_{i,j}$, with $i, j \in \{0, 1, 2\}$ enumerating the types of possible vertices (zeroth, first, or second order). The control parameters can be chosen to generate random reaction networks with a specific underlying structure, such as “closed” networks with very few inflow and outflow reactions, or “open” networks with abundant inflow and outflow. Under this new framework, for each choice of control parameters $\{\alpha_{i,j}\}$, we establish a threshold function $r(n, \{\alpha_{i,j}\})$ such that the probability of the random network being deficiency zero converges to 1 if $\frac{p_n}{r(n, \{\alpha_{i,j}\})} \rightarrow 0$ and converges to 0 if $\frac{p_n}{r(n, \{\alpha_{i,j}\})} \rightarrow \infty$.

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 epidemic 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, which imply a state transition for the associated dynamical system, are termed *reactions*. See Figure 1 for an example of a reaction network.

To each such graph a quantity termed the *deficiency* can be computed, and this quantity is central to many classical and celebrated results in the field [1, 2, 3, 4, 5, 8, 10, 12, 13]. To compute the deficiency, we first note that the vertices of a reaction network, which will be denoted by y

*Department of Mathematics, University of Wisconsin, Madison, USA. anderson@math.wisc.edu.

†Department of Mathematics, University of Wisconsin, Madison, USA. nguyen34@math.wisc.edu.

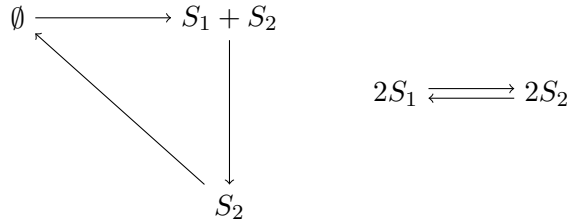


Figure 1: A reaction network with two species: S_1 and S_2 . The vertices are linear combinations of the species over the integers. The directed edges are termed reactions and determine the net change in the counts of the species due to one instance of the reaction. For example, the reaction $S_1 + S_2 \rightarrow S_2$ reduces the count of S_1 by one, but does not affect the count of S_2 .

and/or y' throughout this paper, can be viewed as vectors in $\mathbb{Z}_{\geq 0}^n$. For example, the vertices in Figure 1 can be associated with the vectors $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 2 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 2 \end{bmatrix}$. Moreover, a directed edge between two such vectors, $y \rightarrow y'$, implies a state update of the form $y' - y \in \mathbb{Z}^n$. The set of state update vectors implied by the graph is called the set of “reaction vectors” for the model. Viewing things in this manner the deficiency, δ , for the graph is

$$\delta = \#\text{vertices} - \#\text{connected components} - \dim(\text{span}(\text{reaction vectors})).$$

For example, the deficiency of the reaction network in Figure 1 is $\delta = 5 - 2 - 2 = 1$. Given the significance of deficiency zero, a natural question then arises:

Question: *how prevalent are deficiency zero networks, and, in particular, are they more prevalent in certain natural settings than expected?*

To begin to address this question, our previous paper [6] sought to formulate a framework for deciding the prevalence of deficiency zero among reaction networks with large numbers of species and vertices. In particular, in [6] we considered random reaction networks generated by an Erdős-Rényi random graph framework in the large species limit. We assumed a species set of $\{S_1, \dots, S_n\}$, and, because of their relevance in the biology and chemistry literature, focused on binary reaction networks (whose vertices are of the form \emptyset , S_i , or $S_i + S_j$, see the next section for an explanation of terminology). We then assumed that the probability of an edge, or reaction, between any two vertices, which we denoted by $p_n \in (0, 1)$, was fixed. We then derived a threshold function $r(n) = \frac{1}{n^3}$ such that the probability of the random binary reaction network being deficiency zero converges to 1, as $n \rightarrow \infty$, if $\frac{p_n}{r(n)} \rightarrow 0$ and converges to 0 if $\frac{p_n}{r(n)} \rightarrow \infty$. Here we use the usual notation that for two sequences $\{a_n\}_{n=0}^{\infty}$ and $\{b_n\}_{n=0}^{\infty}$ in \mathbb{R} we write $a_n \ll b_n$ or $b_n \gg a_n$ if $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 0$ and $a_n \sim b_n$ if $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = c$ for some constant c .

While the basic Erdős-Rényi framework 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 $\emptyset \rightleftharpoons S_i$ 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 $S_i \rightleftharpoons S_j$ or $S_i + S_j \rightleftharpoons S_h + S_k$.

To properly generate random reaction networks in those situations, the current paper considers a stochastic block model framework—a generalized Erdős-Rényi framework with weighted edge probabilities [11]. In particular, given that there are n species, we partition the set of all possible reactions into different classes such as $E_n^{0,1} = \{\emptyset \rightleftharpoons S_i\}$, $E_n^{1,1} = \{S_i \rightleftharpoons S_j\}$, $E_n^{0,2} = \{\emptyset \rightleftharpoons S_i + S_j\}$, and so on. In the class with the highest amount of reactions $E_n^{2,2} = \{S_i + S_j \rightleftharpoons S_h + S_k\}$, the edge probability is p_n . For each relevant pair of i and j the edge probability for the reaction class $E_n^{i,j}$ is given by $n^{\alpha_{i,j}} p_n$, where $\{\alpha_{i,j}\}$ are parameters that can be used to control the structure of the random reaction networks. For example, $\alpha_{0,1} = \alpha_{0,2} = 0$ could be used to model closed systems with very few inflow and outflow reactions. Given each choice of $\{\alpha_{i,j}\}$, we will provide a threshold function $r(n, \{\alpha_{i,j}\})$ such that the probability of the random binary reaction network being deficiency zero converges to 1 if $\frac{pn}{r(n, \{\alpha_{i,j}\})} \rightarrow 0$, as $n \rightarrow \infty$, and converges to 0 if $\frac{pn}{r(n, \{\alpha_{i,j}\})} \rightarrow \infty$. For the sake of brevity, we will write $r(n)$ instead of $r(n, \{\alpha_{i,j}\})$ throughout the rest of the work.

The remainder of this paper is organized as follows. In Section 2, we briefly review some key definitions of reaction network theory, including deficiency. In Section 3, we formally set up the stochastic block model framework briefly described above for generating random reaction networks, and provide some concrete examples with illustrations. In Section 4, we provide a set of conditions for deficiency zero in terms of the expected number of reactions in each reaction class. Finally in Section 5, we provide an algorithm to derive the threshold function $r(n)$ for a given choice of control parameters $\{\alpha_{i,j}\}$, which is based on the theoretical results derived in Section 4.

2 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 $y \rightarrow y'$, 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, $y \rightarrow y'$, 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 $\mathbb{Z}_{\geq 0}^n$, 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 $S_1 + S_2 \rightarrow 2S_2$ 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 $y \rightarrow y'$ 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 $\mathcal{S} = \{S_1, \dots, S_n\}$, $\mathcal{C} = \cup_{y \rightarrow y'} \{y, y'\}$, and $\mathcal{R} = \cup_{y \rightarrow y'} \{y \rightarrow y'\}$ 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 1. 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 $y \rightarrow y' \in \mathcal{R}$. Figure 1 is an example of such a graph. We denote by ℓ the number of connected components of the graph. 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. Note that since each connected component must consist of at least two vertices, we have the bound $\ell \leq \frac{|\mathcal{C}|}{2}$.

It is a common practice, which we use here, to specify a reaction network by writing all the reactions, since the sets \mathcal{S} , \mathcal{C} , and \mathcal{R} are contained in this description. For example, the reaction network in Figure 1 has the set of species $\mathcal{S} = \{S_1, S_2\}$, the set of vertices $\mathcal{C} = \{\emptyset, S_2, S_1 + S_2, 2S_1, 2S_2\}$ and the set of reactions $\mathcal{R} = \{\emptyset \rightarrow S_1 + S_2, S_1 + S_2 \rightarrow S_2, S_2 \rightarrow \emptyset, 2S_1 \rightarrow 2S_2, 2S_2 \rightarrow 2S_1\}$.

Definition 2.2. The linear subspace $S = \text{span}\{y' - y\}$ generated by all reaction vectors is called the *stoichiometric subspace* of the network. Denote $\dim(S) = s$. \triangle

Definition 2.3. A vertex, $y \in \mathbb{Z}_{\geq 0}^n$, 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*. \triangle

Definition 2.4. A reaction network $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ is called *binary* if each vertex is binary, unary, or of zeroth order. \triangle

In later sections, we will focus on *binary* reaction networks due to their relevancy in the chemistry and biology literature.

2.2 Deficiency and related results

Definition 2.5. The *deficiency* of a reaction network $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ is $\delta = |\mathcal{C}| - \ell - s$, where $|\mathcal{C}|$ is the number of vertices, ℓ is the number of connected components of the associated graph, and $s = \dim(\text{span}\{y' - y : y \rightarrow y' \in \mathcal{R}\})$ is the dimension of the stoichiometric subspace of the network.

For each $j \leq \ell$, we let \mathcal{C}_j denote the collection of vertices in the j th connected component, 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. \triangle

Remark 3. From the definition of deficiency, the empty network has deficiency zero.

As we are interested in networks with deficiency zero, it is an important fact that a deficiency zero network cannot have too many vertices. The following lemma from [6] gives an upper bound.

Lemma 2.1 (Lemma 2.1(e) from [6]). *Let $n \in \mathbb{N}$ and let $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ be a reaction network with n species. Assume that the reaction network has deficiency zero, then*

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

Next, we will provide another useful fact about deficiency from [6]: the deficiency of a reaction network cannot decrease if we add a reaction to it. This means deficiency zero is a monotone decreasing property, which guarantees that a threshold function for deficiency zero exists (see [7]).

Lemma 2.2 (Lemma 2.1(g) from [6]). *Let $R = \{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ and $\widehat{R} = \{\widehat{\mathcal{S}}, \widehat{\mathcal{C}}, \widehat{\mathcal{R}}\}$ be two reaction networks with $\widehat{\mathcal{R}} \setminus \mathcal{R} = \{y \rightarrow y'\}$, a single reaction. Then*

$$\delta_{\widehat{R}} \geq \delta_R.$$

Definition 2.6. Let $R = \{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ be a reaction network, and $\widetilde{\mathcal{R}} \subset \mathcal{R}$. Then we denote by $\pi_{\widetilde{\mathcal{R}}}(R)$ the reaction network whose set of reactions is $\widetilde{\mathcal{R}}$, and whose species and vertices are the subsets of \mathcal{S} and \mathcal{C} that are associated with $\widetilde{\mathcal{R}}$, according to Remark 1.

Note that in Definition 2.6, $\pi_{\widetilde{\mathcal{R}}}(R)$ can be thought of as a “sub-network”, or a projection of R onto the subset of species, vertices, and reactions associated with $\widetilde{\mathcal{R}}$. The following corollary is a direct consequence of Lemma 2.2.

Corollary 2.1. *Let $R = \{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ be a reaction network, and $\widetilde{\mathcal{R}} \subset \mathcal{R}$. Then*

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

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

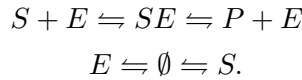
We introduce two more Lemmas related to the deficiency of a network. Their proofs are similar to the proof of Lemma 2.2, and thus they are omitted for the sake of brevity. The first lemma is well-known.

Lemma 2.3. *Let $R = \{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ be a reaction network whose vertices are either unary or of zeroth order, then $\delta_R = 0$.*

Lemma 2.4. *Let $R = \{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ be a reaction network and let $\widetilde{\mathcal{R}}$ be a subset of \mathcal{R} in which precisely one reaction of each reversible pair is removed. If $\widetilde{\mathcal{R}}$ consists of linearly independent reaction vectors, then $\delta_R = 0$.*

We illustrate the concept of deficiency with some reaction networks taken from the biology and chemistry literature.

Example 1 (Enzyme kinetics [4]).

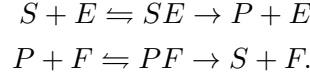


In this example, the reaction network has $|\mathcal{C}| = 6$ vertices, there are $l = 2$ connected components, and the dimension of the stoichiometric subspace is $s = 4$. Thus the deficiency is

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

△

Example 2 (Futile cycle enzyme [14]).



In this example, the reaction network has $|\mathcal{C}| = 6$ vertices, there are $l = 2$ connected components and the dimension of the stoichiometric subspace can be calculated, which yields $s = 3$. Thus the deficiency is

$$\delta = 6 - 2 - 3 = 1.$$

△

3 A stochastic block model framework for binary CRNs

In this section we setup a stochastic block model for generating random reaction networks.

Let the set of species be $\mathcal{S} = \{S_1, S_2, \dots, S_n\}$. 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 yields

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

and so

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

Definition 3.1. 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:

$$\begin{aligned} E_n^{0,1} &= \{\emptyset \rightleftharpoons S_i : 1 \leq i \leq n\} \\ E_n^{0,2} &= \{\emptyset \rightleftharpoons S_i + S_j : 1 \leq i, j \leq n\} \\ E_n^{1,1} &= \{S_i \rightleftharpoons S_j : 1 \leq i, j \leq n; i \neq j\} \\ E_n^{1,2} &= \{S_i \rightleftharpoons S_j + S_k : 1 \leq i, j, k \leq n\} \\ E_n^{2,2} &= \{S_i + S_j \rightleftharpoons S_h + S_k : 1 \leq i, j, k, h \leq n; (i, j) \neq (k, h); (i, j) \neq (h, k)\}. \end{aligned}$$

Remark 4. $E_n^{0,1}, E_n^{0,2}, E_n^{1,1}, E_n^{1,2}, E_n^{2,2}$ 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 $y \rightarrow y' \in \mathcal{R} \implies y' \rightarrow y \in \mathcal{R}$, since deficiency does not depend on the direction of the edges.

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. *In later sections 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 4 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 3 (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 [6]. From [6], and Theorem 5.1 below, the threshold function for deficiency zero is $r(n) = \frac{1}{n^3}$. In other words,

$$\lim_{n \rightarrow \infty} P(\delta_{R_n} = 0) = \begin{cases} 0 & \text{when } \lim_{n \rightarrow \infty} \frac{p_n}{r(n)} = \infty \\ 1 & \text{when } \lim_{n \rightarrow \infty} \frac{p_n}{r(n)} = 0 \end{cases}.$$

Lemma 4.2 and Lemma 4.3 in [6] then tell us that for $\lim_{n \rightarrow \infty} \frac{p_n}{r(n)} = 0$, 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. \triangle

Example 4 (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 $n \rightarrow \infty$. 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 5.1 below, the threshold function for this case is $r(n) = \frac{1}{n^3}$ \triangle

Example 5 (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 2 for a realization of this system with a specific choice of parameters. From Theorem 5.1 below, the threshold function for this case is $r(n) = \frac{1}{n^{10/3}}$. \triangle

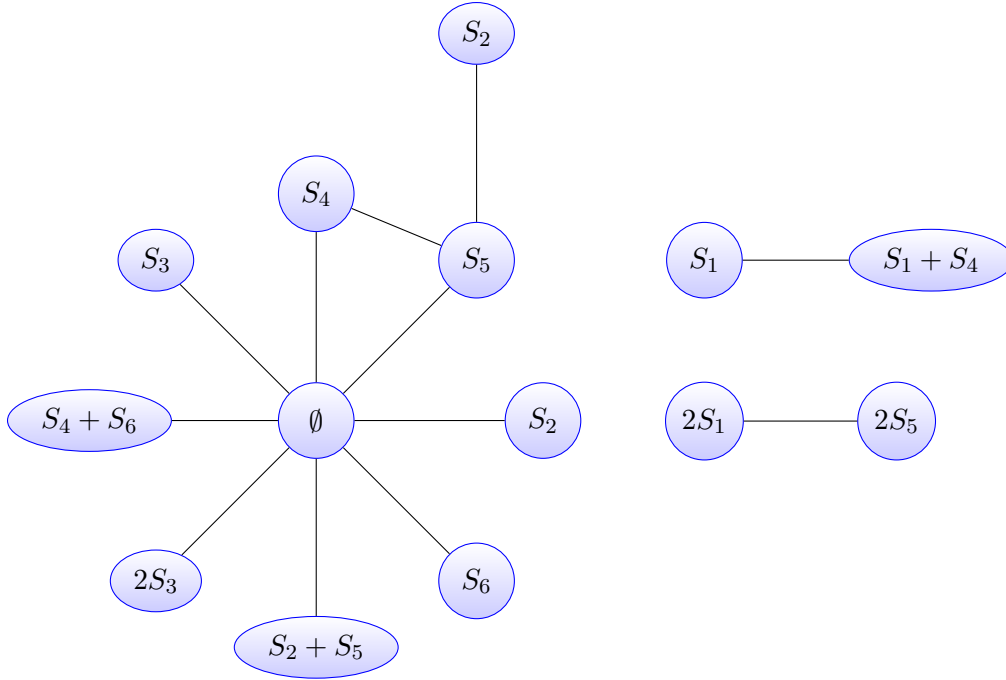


Figure 2: A realization of the open system in Example 5 with $n = 6$ and $p = \frac{0.8}{n^3}$. *Note: The figure only includes non-isolated vertices.*

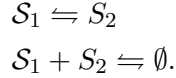
4 Conditions for deficiency zero in terms of $K_{i,j}(n)$

In this section, we will provide a set of conditions on $K_{i,j}(n)$ that guarantee $\lim_{n \rightarrow \infty} \mathbb{P}(\delta_{R_n} = 0) = 0$. We will also show that under the “converse” of these conditions, $\lim_{n \rightarrow \infty} \mathbb{P}(\delta_{R_n} = 0) = 1$. These conditions are used in Section 5 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 zero.

4.1 Conditions on $K_{i,j}(n)$ for $\lim_{n \rightarrow \infty} \mathbb{P}(\delta_{R_n} = 0) = 0$

We start this section by providing some examples which illustrate different ways to break deficiency zero.

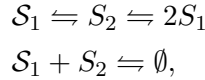
Example 6. Consider a reaction network with only 2 species $\mathcal{S} = \{S_1, S_2\}$



The reaction network has deficiency

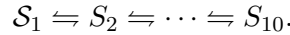
$$\delta = |\mathcal{C}| - \ell - s = 4 - 2 - 2 = 0.$$

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 $2S_1 \rightleftharpoons S_2$, then the new network is



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. \triangle

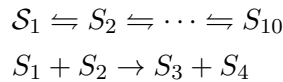
Example 7. Consider a reaction network with 10 species $\mathcal{S} = \{S_1, \dots, S_{10}\}$, which is given below



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 stoichiometric 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 $S_1 + S_2 \rightarrow S_3 + S_4$, then the new network is



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. \triangle

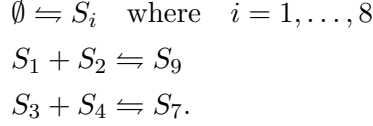
Example 8. Consider a reaction network with 10 species $\mathcal{S} = \{S_1, \dots, S_{10}\}$, and a high number of reactions in $E_n^{0,1}$



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 $\{S_1, \dots, S_8\}$, which breaks deficiency zero. For example, consider the new network



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 $E_n^{2,2}$. \triangle

It turns out that the three examples above are representative of all cases when we have $\lim_{n \rightarrow \infty} \mathbb{P}(\delta_{R_n} = 0) = 0$. We provide rigorous conditions in the following theorem.

Theorem 4.1. *If one of the following conditions holds, then $\lim_{n \rightarrow \infty} \mathbb{P}(\delta_{R_n} = 0) = 0$.*

$$(C1.1) \text{ Either } K_{0,2}(n) \gg n, K_{1,2}(n) \gg n, \text{ or } K_{2,2}(n) \gg n.$$

$$(C1.2) \text{ } K_{1,1}(n) \gg n \text{ and either } K_{0,2}(n) \gg 1, K_{1,2}(n) \gg 1, \text{ or } K_{2,2}(n) \gg 1.$$

$$(C1.3) \text{ Either } K_{0,1}(n)^2 K_{0,2}(n) \gg n^2, K_{0,1}(n)^3 K_{1,2}(n) \gg n^3, \text{ or } K_{0,1}(n)^4 K_{2,2}(n) \gg n^4.$$

Remark 6. *In Theorem 4.1, 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 2.1). 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 2.3). 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 stoichiometric 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 $\lim_{n \rightarrow \infty} \mathbb{P}(\delta_{R_n} = 0) = 0$.

Lemma 4.1. *If either $K_{0,2}(n) \gg n$, $K_{1,2}(n) \gg n$, or $K_{2,2}(n) \gg n$, then we have*

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

Proof. Recall from Lemma 2.1 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 4.1 and Theorem 4.1 in [6].

1. First, we assume that $K_{0,2}(n) \gg n$. From Lemma 2.1, we have that $\delta_{R_n} = 0$ implies $|\mathcal{C}| \leq 2n$, which in turns implies $M_{0,2}(n) \leq 2n - 1$. Thus

$$\begin{aligned} \mathbb{P}(\delta_{R_n} = 0) &\leq \mathbb{P}(M_{0,2}(n) \leq 2n - 1) \\ &= \mathbb{P}(K_{0,2}(n) - M_{0,2}(n) \geq K_{0,2}(n) - (2n - 1)) \\ &\leq \frac{\text{Var}(M_{0,2}(n))}{(K_{0,2}(n) - (2n - 1))^2}. \end{aligned}$$

Since $M_{0,2}(n)$ has a binomial distribution, $\text{Var}(M_{0,2}(n)) \leq \mathbb{E}[M_{0,2}(n)] = K_{0,2}(n)$. Together with the fact that $K_{0,2}(n) \gg n$, we have $\frac{\text{Var}(M_{0,2}(n))}{(K_{0,2}(n) - (2n - 1))^2} \rightarrow 0$, as $n \rightarrow \infty$, and thus $\lim_{n \rightarrow \infty} \mathbb{P}(\delta_{R_n} = 0) = 0$.

2. Next, we assume $K_{1,2}(n) \gg n$. We observe that based on Corollary 2.1, $\delta_{R_n} = 0$ must imply $\delta_{\pi_{E_n^{1,2}}(R_n)} = 0$, where, recalling Definition 2.6, $\pi_{E_n^{1,2}}(R_n)$ 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 2.1. $\delta_{\pi_{E_n^{1,2}}(R_n)} = 0$ 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

$$\mathbb{P}(\delta_{R_n} = 0) \leq \mathbb{P}\left(|I| > \frac{n(n+1)}{2} - 2n\right).$$

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 $\text{Var}(|I|)$ since $|I|$ is binomially distributed. Using $\mathbb{E}|I|$ and $\text{Var}(|I|)$, a rigorous proof for

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(|I| > \frac{n(n+1)}{2} - 2n\right) = 0$$

can be carried out by precisely the same argument as Lemma 4.1 in [6]. We omit it for the sake of brevity.

3. Finally, we assume $K_{2,2}(n) \gg n$. We observe that based on Corollary 2.1, $\delta_{R_n} = 0$ 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

$$\mathbb{P}(\delta_{R_n} = 0) \leq \mathbb{P}(\delta_{\pi_{E_n^{2,2}}(R_n)} = 0).$$

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 4.1 and Theorem 4.1 in [6]. \square

The following proposition will be useful in the proof that Condition C1.2 implies $\lim_{n \rightarrow \infty} \mathbb{P}(\delta_{R_n} = 0) = 0$.

Proposition 4.1. *Let $R = \{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ be a reaction network with $\mathcal{S} = \{S_1, S_2, \dots, S_n\}$. 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 $\widehat{\mathcal{R}} = R \cup \{\emptyset \rightleftharpoons S_i + S_j, \emptyset \rightleftharpoons S_p + S_q\}$ 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 2.3, the deficiency of R is necessarily zero (since it contains only unary vertices). Starting from R , adding the pair of reversible reactions $\emptyset \rightleftharpoons S_i + S_j, \emptyset \rightleftharpoons S_p + S_q$ 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 $\emptyset \rightleftharpoons S_i + S_j$ and $\emptyset \rightleftharpoons S_p + S_q$ 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 4.2. *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*

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

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

$$\lim_{n \rightarrow \infty} \mathbb{P}(M_{0,2}(n) \geq 2) = 1.$$

Now it suffices to show

$$\lim_{n \rightarrow \infty} \mathbb{P}(\delta_{R_n} = 0, M_{0,2}(n) \geq 2) = 0.$$

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 $\emptyset \rightleftharpoons S_i + S_j, \emptyset \rightleftharpoons S_p + S_q$ and note that $\{i, j\} \neq \{p, q\}$. Note that by symmetry the distribution of the pair $(\emptyset \rightleftharpoons S_i + S_j, \emptyset \rightleftharpoons S_p + S_q)$ 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

$$\lim_{n \rightarrow \infty} \mathbb{P}(\delta_{B_n^+} = 0) = 0.$$

Conditioning on the size of B_n , the largest component of $G_n^{1,1}$, yields

$$\mathbb{P}(\delta_{B_n^+} = 0) = \sum_{k=1}^n \mathbb{P}(\delta_{B_n^+} = 0 | |B_n| = k) \mathbb{P}(|B_n| = k). \quad (1)$$

From Proposition 4.1, 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

$$\begin{aligned} \mathbb{P}(\delta_{B_n^+} = 0 | |B_n| = k) &\leq \mathbb{P}(\text{not all of } S_i, S_j, S_p, S_q \text{ are contained in } B_n | |B_n| = k) \\ &= 1 - \mathbb{P}(S_i, S_j, S_p, S_q \in B_n | |B_n| = k). \end{aligned} \quad (2)$$

We will compute the probability as follows

$$\mathbb{P}(S_i, S_j, S_p, S_q \in B_n | |B_n| = k) = \mathbb{P}(S_p, S_q \in B_n | S_i, S_j \in B_n, |B_n| = k) \mathbb{P}(S_i, S_j \in B_n | |B_n| = k). \quad (3)$$

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 $\emptyset \rightleftharpoons S_i + S_j$ 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 $\emptyset \rightleftharpoons 2S_i$ with $S_i \in B_n$. Since there are a total of $\binom{n}{2} + n$ elements in $E_n^{0,2}$ we have

$$\mathbb{P}(S_i, S_j \in B_n | |B_n| = k) = \frac{\binom{k}{2} + k}{\binom{n}{2} + n} = \frac{k(k+1)}{n(n+1)} \geq \left(\frac{k}{n}\right)^2, \quad (4)$$

where the inequality holds since $k \leq n$. Similarly, we have

$$\mathbb{P}(S_p, S_q \in B_n | S_i, S_j \in B_n, |B_n| = k) \geq \frac{\binom{k}{2} + k - 1}{\binom{n}{2} + n - 1} \geq \left(\frac{k}{n}\right)^2, \quad (5)$$

where the inequality holds for $k \geq 4$, which can be verified in a straightforward manner. From (2), (3), (4), and (5) we have that for $k \geq 4$

$$\mathbb{P}(\delta_{B_n^+} = 0 | |B_n| = k) \leq 1 - \left(\frac{k}{n}\right)^4. \quad (6)$$

Finally, combining (1) and (6), we have

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(\delta_{B_n^+} = 0) &\leq \sum_{k=4}^n \left(1 - \left(\frac{k}{n}\right)^4\right) \mathbb{P}(|B_n| = k) + \sum_{k=1}^3 \mathbb{P}(|B_n| = k) \\ &= \mathbb{E}\left(1 - \left(\frac{|B_n|}{n}\right)^4\right) + \sum_{k=1}^3 \left(\frac{k}{n}\right)^4 \mathbb{P}(|B_n| = k) \\ &\leq 1 - \left(\mathbb{E}\left(\frac{|B_n|}{n}\right)\right)^4 + \frac{98}{n^4}, \end{aligned} \quad (7)$$

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, [9], $\frac{|B_n|}{n} - f(c_n) \xrightarrow{\mathbb{P}} 0$, 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 $\lim_{n \rightarrow \infty} f(c_n) = 1$. Since both $\frac{|B_n|}{n}$ and $f(c_n)$ are bounded by 1, we have

$$\lim_{n \rightarrow \infty} \mathbb{E} \left(\frac{|B_n|}{n} \right) = \lim_{n \rightarrow \infty} f(c_n) = 1,$$

which completes the proof. \square

Finally, we have the proof related to Condition C1.3.

Lemma 4.3. *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*

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

Proof. It suffices to show $K_{0,1}(n)^2 K_{0,2}(n) \gg n^2$ implies $\lim_{n \rightarrow \infty} \mathbb{P}(\delta_{R_n} = 0) = 0$. 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

$$\mathbb{P}(\delta_{R_n} = 0) = \sum_{\substack{i \leq n \\ j \leq n(n+1)/2}} \mathbb{P}(\delta_{R_n} = 0 | M_{0,1}(n) = i, M_{0,2}(n) = j) \mathbb{P}(M_{0,1}(n) = i, M_{0,2}(n) = j). \quad (8)$$

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 $\emptyset \rightleftharpoons S_p, \emptyset \rightleftharpoons S_q, \emptyset \rightleftharpoons S_p + S_q$ has positive deficiency, so any network containing it also has positive deficiency according to Corollary 2.1. 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 $\emptyset \rightleftharpoons S_p + S_q$ (note that p and q can be the same) where $\emptyset \rightleftharpoons S_p$ and $\emptyset \rightleftharpoons S_q$ are already present is

$$\frac{\binom{i}{2} + i}{\frac{n(n+1)}{2}} = \frac{i(i+1)}{n(n+1)}.$$

We may then use 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) to conclude

$$\mathbb{P}(\delta_{R_n} = 0 | M_{0,1}(n) = i, M_{0,2}(n) = j) \leq \left(1 - \frac{i(i+1)}{n(n+1)} \right)^j. \quad (9)$$

Combining (8) and (9), we have

$$\begin{aligned} \mathbb{P}(\delta_{R_n} = 0) &\leq \sum_{\substack{i \leq n \\ j \leq n(n+1)/2}} \left(1 - \frac{i(i+1)}{n(n+1)} \right)^j \mathbb{P}(M_{0,1}(n) = i, M_{0,2}(n) = j) \\ &= \mathbb{E} \left[\left(1 - \frac{M_{0,1}(n)(M_{0,1}(n)+1)}{n(n+1)} \right)^{M_{0,2}(n)} \right]. \end{aligned}$$

We have $\frac{M_{0,1}(n)(M_{0,1}(n)+1)}{n(n+1)} \geq \frac{M_{0,1}(n)^2}{2n^2}$, thus

$$\mathbb{P}(\delta_{R_n} = 0) \leq \mathbb{E} \left[\left(1 - \frac{M_{0,1}(n)^2}{2n^2} \right)^{M_{0,2}(n)} \right] \leq \mathbb{E} \left[e^{-\frac{M_{0,1}(n)^2 M_{0,2}(n)}{2n^2}} \right], \quad (10)$$

where the second inequality follows the fact that $1 - x \leq e^{-x}$. Notice further that $e^{-x} \leq \frac{1}{x+1}$ for $x \geq 0$, hence we have

$$\mathbb{E} \left[e^{-\frac{M_{0,1}(n)^2 M_{0,2}(n)}{2n^2}} \right] \leq \mathbb{E} \left[\frac{2n^2}{M_{0,1}(n)^2 M_{0,2}(n) + 2n^2} \right] = 2n^2 \mathbb{E} \left[\frac{1}{M_{0,1}(n)^2 M_{0,2}(n) + 2n^2} \right]. \quad (11)$$

Since $M_{0,1}(n) \leq n$ and $M_{0,2}(n) \leq \frac{n(n+1)}{2}$, we have for n large enough

$$\begin{aligned} M_{0,1}(n)^2 M_{0,2}(n) + 2n^2 &\geq (M_{0,1}(n)^2 + 1)(M_{0,2}(n) + 1) \\ &\geq \frac{1}{2}(M_{0,1}(n) + 1)^2(M_{0,2}(n) + 1) \\ &\geq \frac{1}{4}(M_{0,1}(n) + 1)(M_{0,1}(n) + 2)(M_{0,2}(n) + 1), \end{aligned} \quad (12)$$

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 $\frac{1}{2}(a+b)^2 \leq a^2 + b^2$ inequality, and the last inequality comes from $M_{0,1}(n) + 1 \geq \frac{1}{2}(M_{0,1}(n) + 2)$, which is true as long as $M_{0,1}(n) \geq 0$.

Combining (10), (11), (12), and noticing that $M_{0,1}(n)$ and $M_{0,2}(n)$ are independent, we have

$$\mathbb{P}(\delta_{R_n} = 0) \leq 8n^2 \mathbb{E} \left[\frac{1}{(M_{0,1}(n) + 1)(M_{0,1}(n) + 2)} \right] \mathbb{E} \left[\frac{1}{M_{0,2}(n) + 1} \right]. \quad (13)$$

Since $M_{0,1}(n) \sim B(n, K_{0,1}(n)/n)$, from Lemma A.1, we have

$$\mathbb{E} \left[\frac{1}{(M_{0,1}(n) + 1)(M_{0,1}(n) + 2)} \right] \leq \frac{1}{K_{0,1}(n)^2}. \quad (14)$$

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

$$\mathbb{E} \left[\frac{1}{M_{0,2}(n) + 1} \right] \leq \frac{1}{K_{0,2}(n)}. \quad (15)$$

Thus from (13), (14), (15) 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. \square

4.2 Conditions on $K_{i,j}(n)$ for $\lim_{n \rightarrow \infty} \mathbb{P}(\delta_{R_n} = 0) = 1$

Note that the conditions below are essentially the converse of Theorem 4.1.

Theorem 4.2. *If all of the following conditions hold, then $\lim_{n \rightarrow \infty} \mathbb{P}(\delta_{R_n} = 0) = 1$.*

(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 $\lim_{n \rightarrow \infty} \mathbb{P}(\delta_{R_n} = 0) = 1$. 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

$$\lim_{n \rightarrow \infty} \mathbb{P}(M_{0,2}(n) = 0) = \lim_{n \rightarrow \infty} \mathbb{P}(M_{1,2}(n) = 0) = \lim_{n \rightarrow \infty} \mathbb{P}(M_{2,2}(n) = 0) = 1.$$

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 2.3. 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 satisfied, 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 $\lim_{n \rightarrow \infty} \mathbb{P}(\delta_{R_n} = 0) = 1$. Combining the above shows that Theorem 4.2 will be proved by showing that $\lim_{n \rightarrow \infty} \mathbb{P}(\delta_{R_n} = 0) = 1$ 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 4.4. *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*

- $\lim_{n \rightarrow \infty} Q_{i,j}(n) > 0$ 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 $\lim_{n \rightarrow \infty} Q_{1,1}(n) > 0$.

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

$$R_{0,1}(n) = \min \left\{ \sqrt{K_{0,1}(n) \frac{n}{\sqrt{K_{0,2}(n)}}}, \sqrt{n K_{0,1}(n)} \right\}.$$

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

$$Q_{0,2}(n) = \max \left\{ 1, \min \left\{ \sqrt{K_{0,2}(n) \frac{n^2}{R_{0,1}(n)^2}}, \sqrt{n K_{0,2}(n)} \right\} \right\}$$

where the minimum is interpreted asymptotically as $n \rightarrow \infty$. 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

- $K_{0,1}(n) \ll S_{0,1}(n) \ll n$, $K_{1,2}(n) \ll Q_{1,2}(n) \ll n$, $\lim_{n \rightarrow \infty} Q_{1,2}(n) > 0$ and $S_{0,1}(n)^3 Q_{1,2}(n) \ll n^3$.
- $K_{0,1}(n) \ll T_{0,1}(n) \ll n$, $K_{2,2}(n) \ll Q_{2,2}(n) \ll n$, $\lim_{n \rightarrow \infty} Q_{2,2}(n) > 0$ 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 $n \rightarrow \infty$. We now have all the $Q_{i,j}(n)$, and all the desired properties are straightforward to confirm. \square

We turn to the main proof of Theorem 4.2. The main proof utilizes some technical results, which will be proven in several lemmas after the main proof.

Proof of Theorem 4.2. Assume that conditions (C2.1*) and (C2.3) hold. We have

$$\mathbb{P}(\delta_{R_n} = 0) = \mathbb{P}(\delta_{R_n} = 0, \cap_{i,j} \{M_{i,j}(n) \leq Q_{i,j}(n)\}) + \mathbb{P}(\delta_{R_n} = 0, \cup_{i,j} \{M_{i,j}(n) > Q_{i,j}(n)\}) \quad (16)$$

We will show that the second term goes to zero. Since each $M_{i,j}(n)$ has a binomial distribution, we have

$$\begin{aligned} \mathbb{P}(M_{i,j}(n) > Q_{i,j}(n)) &= \mathbb{P}(M_{i,j}(n) - K_{i,j}(n) > Q_{i,j}(n) - K_{i,j}(n)) \\ &\leq \frac{\text{Var}(M_{i,j}(n))}{(Q_{i,j}(n) - K_{i,j}(n))^2} \leq \frac{K_{i,j}(n)}{(Q_{i,j}(n) - K_{i,j}(n))^2}. \end{aligned}$$

Since $K_{i,j}(n) \ll Q_{i,j}(n)$ and $\lim_{n \rightarrow \infty} Q_{i,j}(n) > 0$, we have $\lim_{n \rightarrow \infty} \mathbb{P}(M_{i,j}(n) > Q_{i,j}(n)) = 0$ for all (i, j) . Thus

$$\lim_{n \rightarrow \infty} \mathbb{P}(\cup_{i,j} \{M_{i,j}(n) > Q_{i,j}(n)\}) = 0, \quad (17)$$

and consequently,

$$\lim_{n \rightarrow \infty} \mathbb{P}(\delta_{R_n} = 0, \cup_{i,j} \{M_{i,j}(n) > Q_{i,j}(n)\}) = 0. \quad (18)$$

Now we consider the first term in (16). We have

$$\begin{aligned} & \mathbb{P}(\delta_{R_n} = 0, \cap_{i,j} \{M_{i,j}(n) \leq Q_{i,j}(n)\}) \\ &= \sum_{k_{i,j}(n)=0}^{Q_{i,j}(n)} \mathbb{P}(\delta_{R_n} = 0 | \cap_{i,j} \{M_{i,j}(n) = k_{i,j}(n)\}) \mathbb{P}(\cap_{i,j} \{M_{i,j}(n) = k_{i,j}(n)\}) \end{aligned}$$

We will prove in Lemma 4.8 below that

$$\mathbb{P}(\delta_{R_n} = 0 | \cap_{i,j} \{M_{i,j}(n) = k_{i,j}(n)\}) \geq 1 - C_3 \frac{Q(n)}{n}, \quad (19)$$

where $Q(n)$ is a function satisfying $Q(n) \ll n$ and C_3 is independent from n and $k_{i,j}(n)$. Thus

$$\begin{aligned} \mathbb{P}(\delta_{R_n} = 0, \cap_{i,j} \{M_{i,j}(n) \leq Q_{i,j}(n)\}) &\geq \left(1 - C_3 \frac{Q(n)}{n}\right) \sum_{k_{i,j}(n)=0}^{Q_{i,j}(n)} \mathbb{P}(\cap_{i,j} \{M_{i,j}(n) = k_{i,j}(n)\}) \\ &= \left(1 - C_3 \frac{Q(n)}{n}\right) \mathbb{P}(\cap_{i,j} \{M_{i,j}(n) \leq Q_{i,j}(n)\}). \end{aligned} \quad (20)$$

Equation (17) gives us $\lim_{n \rightarrow \infty} \mathbb{P}(\cap_{i,j} \{M_{i,j}(n) \leq Q_{i,j}(n)\}) = 1$, thus from (20) we have

$$\lim_{n \rightarrow \infty} \mathbb{P}(\delta_{R_n} = 0, \cap_{i,j} \{M_{i,j}(n) \leq Q_{i,j}(n)\}) = 1. \quad (21)$$

Combining (16), (18), and (21) we have

$$\lim_{n \rightarrow \infty} \mathbb{P}(\delta_{R_n} = 0) = 1. \quad \square$$

To complete this section, we will provide a series of lemmas, eventually leading to Lemma 4.8, which yields the critical bound (19)

$$\mathbb{P}(\delta_{R_n} = 0 | \cap_{i,j} \{M_{i,j}(n) = k_{i,j}(n)\}) \geq 1 - C_3 \frac{Q(n)}{n}.$$

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 $\emptyset \rightleftharpoons 2S_1$, in which there is only one species, or we could have $\emptyset \rightleftharpoons S_1 + S_2$, 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 $n \rightarrow \infty$, the realized reactions from each set will consist of the maximal number of distinct species.

Lemma 4.5. *Suppose conditions (C2.1*) and (C2.3) hold and that $Q_{i,j}(n)$ are defined as in Lemma 4.4. 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*

$$\lim_{n \rightarrow \infty} \mathbb{P}(A_n | \cap_{i,j} \{M_{i,j}(n) = k_{i,j}(n)\}) = 1.$$

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 $\emptyset \rightleftharpoons S_i + S_j$. 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

$$\begin{aligned} & \mathbb{P}(A_n^{0,2} | M_{0,2}(n) = k_{0,2}(n)) \\ &= \left(1 - \frac{n}{n(n+1)/2}\right) \left(1 - \frac{n}{n(n+1)/2 - 1}\right) \cdots \left(1 - \frac{n}{n(n+1)/2 - k_{0,2}(n) + 1}\right) \\ &= \left(1 - \frac{2}{n+1}\right) \left(1 - \frac{2}{n+1 - \frac{2}{n}}\right) \cdots \left(1 - \frac{2}{n+1 - \frac{2(k_{0,2}(n)-1)}{n}}\right) \\ &\geq \left(1 - \frac{2}{n}\right)^{k_{0,2}(n)} \\ &\geq 1 - \frac{2k_{0,2}(n)}{n}, \end{aligned} \tag{22}$$

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 $S_i \rightleftharpoons S_i + S_j$ (where i and j are not necessarily different) or $S_i \rightarrow 2S_j$ (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

$$\begin{aligned} & \mathbb{P}(A_n^{1,2} | M_{1,2}(n) = k_{1,2}(n)) \\ &= \left(1 - \frac{n(2n-1)}{n^2(n+1)/2}\right) \left(1 - \frac{n(2n-1)}{n^2(n+1)/2 - 1}\right) \cdots \left(1 - \frac{n(2n-1)}{n^2(n+1)/2 - k_{1,2}(n) + 1}\right) \\ &\geq \left(1 - \frac{4}{n}\right)^{k_{1,2}(n)} \\ &\geq 1 - \frac{4k_{1,2}(n)}{n}, \end{aligned} \tag{23}$$

where the first inequality here follows a similar argument to the first inequality in (22).

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

$$\begin{aligned}
& \mathbb{P}(A_n^{2,2} | M_{2,2}(n) = k_{2,2}(n)) \\
&= \left(1 - \frac{\frac{n(n-1)(2n-1)}{2}}{\frac{n(n+1)(n-1)(n+2)}{8}}\right) \cdots \left(1 - \frac{\frac{n(n-1)(2n-1)}{2}}{\frac{n(n+1)(n-1)(n+2)}{8} - k_{2,2}(n) + 1}\right) \\
&\geq \left(1 - \frac{8}{n}\right)^{k_{2,2}(n)} \\
&\geq 1 - \frac{8k_{2,2}(n)}{n},
\end{aligned} \tag{24}$$

where the first inequality here follows a similar argument as the first inequality in (22).

From (22), (23), (24), and independence, we have

$$\begin{aligned}
& \mathbb{P}(A_n | \cap_{i,j} \{M_{i,j}(n) = k_{i,j}(n)\}) \\
&\geq \left(1 - \frac{2k_{0,2}(n)}{n}\right) \left(1 - \frac{4k_{1,2}(n)}{n}\right) \left(1 - \frac{8k_{2,2}(n)}{n}\right) \\
&\geq \left(1 - \frac{2Q_{0,2}(n)}{n}\right) \left(1 - \frac{4Q_{1,2}(n)}{n}\right) \left(1 - \frac{8Q_{2,2}(n)}{n}\right),
\end{aligned}$$

and the limit follows. \square

In our next major lemma, Lemma 4.7, we require the notion of a minimally dependent set, which we define below.

Definition 4.1. 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 4.6. *Let M be a matrix whose columns v_1, v_2, \dots, v_m are minimally dependent. Then M has no row with only one non-zero entry.*

Proof. Since v_1, \dots, v_m are dependent, there exist constants $\alpha_1, \dots, \alpha_m$, 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 i th 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 $\emptyset \rightleftharpoons S_1, \emptyset \rightleftharpoons S_2, \emptyset \rightleftharpoons S_3, \emptyset \rightleftharpoons S_1 + S_2$, whose reaction vectors are dependent, but not minimally dependent because the proper subset containing $\emptyset \rightleftharpoons S_1, \emptyset \rightleftharpoons S_2, \emptyset \rightleftharpoons S_1 + S_2$ 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 4.7. *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 4.6 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

$$i_1 \leq z_{i_1+i_2}. \tag{25}$$

To show that the inequality (25) 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, \dots, 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. \square

Finally, we present the main lemma, giving the bound needed for Theorem 4.2. 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 4.7 will play a critical role in this approach.

Lemma 4.8. *Suppose conditions (C2.1*) and (C2.3) hold and that $Q_{i,j}(n)$ are as in Lemma 4.4. Suppose further that $k_{i,j}(n) \leq Q_{i,j}(n)$ for each relevant pair (i, j) . Then*

$$\mathbb{P}(\delta_{R_n} = 0 \mid \cap_{i,j} \{M_{i,j}(n) = k_{i,j}(n)\}) \geq 1 - C_3 \frac{Q(n)}{n},$$

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

$$\begin{aligned} & \mathbb{P}(\delta_{R_n} = 0 \mid \cap_{i,j} \{M_{i,j}(n) = k_{i,j}(n)\}) \\ & \geq \mathbb{P}(\delta_{R_n} = 0 \mid A_n, \cap_{i,j} \{M_{i,j}(n) = k_{i,j}(n)\}) \mathbb{P}(A_n \mid \cap_{i,j} \{M_{i,j}(n) = k_{i,j}(n)\}) \\ & = (1 - \mathbb{P}(\delta_{R_n} > 0 \mid A_n, \cap_{i,j} \{M_{i,j}(n) = k_{i,j}(n)\})) \mathbb{P}(A_n \mid \cap_{i,j} \{M_{i,j}(n) = k_{i,j}(n)\}). \end{aligned} \quad (26)$$

From Lemma 2.3 and Lemma 2.4, 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 $i_1 \leq k_{0,1}(n), \dots, i_5 \leq k_{2,2}(n)$. Then we have

$$\begin{aligned} & \mathbb{P}(\delta_{R_n} > 0 \mid A_n, \cap_{i,j} \{M_{i,j}(n) = k_{i,j}(n)\}) \\ & \leq \sum_{\substack{I \leq K_n \\ i_3+i_4+i_5 > 0 \\ i_1 \leq 2i_3+3i_4+4i_5}} \binom{k_{0,1}(n)}{i_1} \binom{k_{1,1}(n)}{i_2} \binom{k_{0,2}(n)}{i_3} \binom{k_{1,2}(n)}{i_4} \binom{k_{2,2}(n)}{i_5} \mathbb{P}(B_I \mid A_n, \cap_{i,j} \{M_{i,j}(n) = k_{i,j}(n)\}), \end{aligned} \quad (27)$$

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 4.7.

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

$$\begin{aligned}
\mathbb{P}(B_I|A_n, \cap_{i,j}\{M_{i,j}(n) = k_{i,j}(n)\}) &\leq \binom{n}{\ell} \left(\frac{\ell}{n}\right)^{i_1+2i_2+2i_3+3i_4+4i_5} \\
&\leq \frac{n^\ell}{\ell!} \left(\frac{\ell}{n}\right)^{\ell + \frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}} \\
&\leq \frac{n^\ell}{\ell^\ell e^{-\ell}} \left(\frac{\ell}{n}\right)^{\ell + \frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}} \\
&\leq \left(\frac{e\ell}{n}\right)^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}}, \tag{28}
\end{aligned}$$

where the third inequality is due to the inequality $x! \geq x^x e^{-x}$. Combining (27) and (28), we have

$$\begin{aligned}
&\mathbb{P}(\delta_{R_n} > 0|A_n, \cap_{i,j}\{M_{i,j}(n) = k_{i,j}(n)\}) \\
&\leq \sum_{\substack{I \leq K_n \\ i_3+i_4+i_5 > 0 \\ i_1 \leq 2i_3+3i_4+4i_5}} \binom{k_{0,1}(n)}{i_1} \binom{k_{1,1}(n)}{i_2} \dots \binom{k_{2,2}(n)}{i_5} \left(\frac{e\ell}{n}\right)^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}} \\
&\leq \sum_{\substack{I \leq K_n \\ i_3+i_4+i_5 > 0 \\ i_1 \leq 2i_3+3i_4+4i_5}} \left(\frac{ek_{0,1}(n)}{i_1}\right)^{i_1} \dots \left(\frac{ek_{2,2}(n)}{i_5}\right)^{i_5} \left(\frac{e\ell}{n}\right)^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}} \\
&\leq \sum_{\substack{I \leq K_n \\ i_3+i_4+i_5 > 0 \\ i_1 \leq 2i_3+3i_4+4i_5}} \left(\frac{(5ek_{0,1}(n))^{i_1} \dots (5ek_{2,2}(n))^{i_5}}{(i_1+i_2+i_3+i_4+i_5)^{i_1+i_2+i_3+i_4+i_5}}\right) \left(\frac{e\ell}{n}\right)^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}} \\
&\leq \sum_{\substack{I \leq K_n \\ i_3+i_4+i_5 > 0 \\ i_1 \leq 2i_3+3i_4+4i_5}} \left(\frac{(5eQ_{0,1}(n))^{i_1} \dots (5eQ_{2,2}(n))^{i_5}}{(i_1+i_2+i_3+i_4+i_5)^{i_1+i_2+i_3+i_4+i_5}}\right) \left(\frac{e\ell}{n}\right)^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}},
\end{aligned}$$

where the second inequality is again due to $x! \geq x^x e^{-x}$ and the third inequality is due to Corollary A.1. 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

$$\begin{aligned}
&\mathbb{P}(\delta_{R_n} > 0|A_n, \cap_{i,j}\{M_{i,j}(n) = k_{i,j}(n)\}) \\
&\leq \sum_{\substack{I \leq K_n \\ i_3+i_4+i_5 > 0 \\ i_1 \leq 2i_3+3i_4+4i_5}} \frac{(5eQ_{0,1}(n))^{i_1} \dots (5eQ_{2,2}(n))^{i_5} (i_1+i_2+i_3+i_4+i_5)^{\ell-(i_1+i_2+i_3+i_4+i_5)}}{((2e)^{-1}n)^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}}} \\
&\leq \sum_{\substack{I \leq K_n \\ i_3+i_4+i_5 > 0 \\ i_1 \leq 2i_3+3i_4+4i_5}} \frac{(5eQ_{0,1}(n))^{i_1} \dots (5eQ_{2,2}(n))^{i_5} (i_1+i_2+i_3+i_4+i_5)^{-\frac{i_1}{2}+\frac{i_4}{2}+i_5}}{((2e)^{-1}n)^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}}} \\
&= S_n + T_n, \tag{29}
\end{aligned}$$

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

$$T_n = \sum_{\substack{I \leq K_n \\ i_3+i_4+i_5 > 0 \\ i_1 \leq 2i_3+3i_4+4i_5 \\ i_1 > i_4+2i_5}} \frac{(5eQ_{0,1}(n))^{i_1} \cdots (5eQ_{2,2}(n))^{i_5} (i_1 + i_2 + i_3 + i_4 + i_5)^{-\frac{i_1}{2} + \frac{i_4}{2} + i_5}}{((2e)^{-1}n)^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}}}$$

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

$$\begin{aligned} T_n &\leq \sum_{\substack{I \leq K_n \\ i_3+i_4+i_5 > 0 \\ i_1 \leq 2i_3+3i_4+4i_5 \\ i_1 > i_4+2i_5}} \frac{(5eQ_{0,1}(n))^{i_1} \cdots (5eQ_{2,2}(n))^{i_5}}{((2e)^{-1}n)^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}}} \\ &= \sum_{\substack{I \leq K_n \\ i_3+i_4+i_5 > 0 \\ i_1 \leq 2i_3+3i_4+4i_5 \\ i_1 > i_4+2i_5}} \frac{Q_{0,1}(n)^{i_1} \cdots Q_{2,2}(n)^{i_5} (5e)^{i_1+i_2+i_3+i_4+i_5}}{((2e)^{-1}n)^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}}} \\ &\leq \sum_{\substack{I \leq K_n \\ i_3+i_4+i_5 > 0 \\ i_1 \leq 2i_3+3i_4+4i_5 \\ i_1 > i_4+2i_5}} \frac{Q_{0,1}(n)^{i_1} \cdots Q_{2,2}(n)^{i_5}}{((50e^3)^{-1}n)^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}}}, \end{aligned} \quad (30)$$

where the last inequality is due to the fact that $i_1 + i_2 + i_3 + i_4 + i_5 \leq 2^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}}$. Let

$$Q(n) = \max\{Q_{i,j}(n), Q_{0,1}(n)Q_{0,2}(n)^{1/2}, Q_{0,1}(n)Q_{1,2}(n)^{1/3}, Q_{0,1}(n)Q_{2,2}(n)^{1/4}\}, \quad (31)$$

where the maximum is interpreted asymptotically as $n \rightarrow \infty$. From the way we construct $Q_{i,j}(n)$ in Lemma 4.4 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

$$Q_{0,1}(n)^{i_1 \frac{2i_3}{2i_3+3i_4+4i_5}} Q_{0,2}(n)^{\frac{i_1}{2} \frac{2i_3}{2i_3+3i_4+4i_5}} \leq Q(n)^{i_1 \frac{2i_3}{2i_3+3i_4+4i_5}},$$

$$Q_{0,1}(n)^{i_1 \frac{3i_4}{2i_3+3i_4+4i_5}} Q_{1,2}(n)^{\frac{i_1}{3} \frac{3i_4}{2i_3+3i_4+4i_5}} \leq Q(n)^{i_1 \frac{3i_4}{2i_3+3i_4+4i_5}},$$

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 (30), $Q_{0,2}(n)$ 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

$$\begin{aligned} Q_{0,1}(n)^{i_1} \cdots Q_{2,2}(n)^{i_5} &\leq Q(n)^{i_1+i_2} Q(n)^{i_3+i_4+i_5 - i_1 \frac{i_3+i_4+i_5}{2i_3+3i_4+4i_5}} \\ &= Q(n)^{i_1+i_2+i_3+i_4+i_5 - i_1 \frac{i_3+i_4+i_5}{2i_3+3i_4+4i_5}}. \end{aligned} \quad (32)$$

Note that $i_1 \leq 2i_3 + 3i_4 + 4i_5$, thus

$$i_1 + i_2 + i_3 + i_4 + i_5 - i_1 \frac{i_3 + i_4 + i_5}{2i_3 + 3i_4 + 4i_5} \leq \frac{i_1 + 2i_2 + 2i_3 + 3i_4 + 4i_5}{2}, \quad (33)$$

where the inequality above can be verified in a straightforward manner. Combining (30), (32), and (33), and noting that $i_3 + i_4 + i_5 > 0$, we have

$$\begin{aligned} T_n &\leq \sum_{\substack{I \leq K_n \\ i_3+i_4+i_5 > 0 \\ i_1 \leq 2i_3+3i_4+4i_5 \\ i_1 > i_4+2i_5}} \left(\frac{Q(n)}{(50e^3)^{-1}n} \right)^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}} \\ &\leq \frac{Q(n)}{(50e^3)^{-1}n} \sum_{i_1=0}^{\infty} \left(\frac{Q(n)}{(50e^3)^{-1}n} \right)^{i_1/2} \cdots \sum_{i_5=0}^{\infty} \left(\frac{Q(n)}{(50e^3)^{-1}n} \right)^{2i_5} \\ &\leq C_1 \frac{Q(n)}{n}, \end{aligned} \quad (34)$$

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 $i_1 \leq k_{0,1}(n) \leq Q_{0,1}(n), \dots, i_5 \leq k_{2,2}(n) \leq Q_{2,2}(n)$, implying $i_1, \dots, i_5 \leq Q(n)$. Therefore we have

$$\begin{aligned} S_n &\leq \sum_{\substack{I \leq K_n \\ i_3+i_4+i_5 > 0 \\ i_1 \leq 2i_3+3i_4+4i_5 \\ i_1 \leq i_4+2i_5}} \frac{(5eQ(n))^{i_1+i_2+i_3+i_4+i_5 - \frac{i_1}{2} + \frac{i_4}{2} + i_5}}{((2e)^{-1}n)^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}}} \\ &= \sum_{\substack{I \leq K_n \\ i_3+i_4+i_5 > 0 \\ i_1 \leq 2i_3+3i_4+4i_5 \\ i_1 \leq i_4+2i_5}} \frac{(5eQ(n))^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}}}{((2e)^{-1}n)^{\frac{i_1+2i_2+2i_3+3i_4+4i_5}{2}}} \\ &\leq \frac{5eQ(n)}{(2e)^{-1}n} \sum_{i_1=0}^{\infty} \left(\frac{5eQ(n)}{(2e)^{-1}n} \right)^{i_1/2} \cdots \sum_{i_5=0}^{\infty} \left(\frac{5eQ(n)}{(2e)^{-1}n} \right)^{2i_5} \\ &\leq C_2 \frac{Q(n)}{n}, \end{aligned} \quad (35)$$

where C_2 is independent from n and $k_{i,j}(n)$. From (29), (34), (35), we have

$$\mathbb{P}(\delta_{R_n} > 0 | A_n, \cap_{i,j} \{M_{i,j}(n) = k_{i,j}(n)\}) \leq C_1 \frac{Q(n)}{n} + C_2 \frac{Q(n)}{n}. \quad (36)$$

From Lemma 4.5 and the fact that $Q_{i,j}(n) \leq Q(n)$, we have

$$\mathbb{P}(A_n | \cap_{i,j} \{M_{i,j}(n) = k_{i,j}(n)\}) \geq \left(1 - \frac{2Q(n)}{n}\right) \left(1 - \frac{4Q(n)}{n}\right) \left(1 - \frac{8Q(n)}{n}\right). \quad (37)$$

Plugging (36) and (37) into (26) yields

$$\begin{aligned} & \mathbb{P}(\delta_{R_n} = 0 | \cap_{i,j} \{M_{i,j}(n) = k_{i,j}(n)\}) \\ & \geq \left(1 - C_1 \frac{Q(n)}{n} - C_2 \frac{Q(n)}{n}\right) \left(1 - \frac{2Q(n)}{n}\right) \left(1 - \frac{4Q(n)}{n}\right) \left(1 - \frac{8Q(n)}{n}\right) \\ & \geq 1 - C_3 \frac{Q(n)}{n}, \end{aligned} \quad (38)$$

where the last inequality is obtained from repeatedly applying $(1-a)(1-b) \geq 1-a-b$ (where $a, b \geq 0$). Clearly we must have C_3 independent from n and $k_{i,j}(n)$. \square

5 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. $\lim_{n \rightarrow \infty} \mathbb{P}(\delta_{R_n} = 0) = 0$ for $\lim_{n \rightarrow \infty} \frac{p_n}{r(n)} = \infty$, and
2. $\lim_{n \rightarrow \infty} \mathbb{P}(\delta_{R_n} = 0) = 1$ for $\lim_{n \rightarrow \infty} \frac{p_n}{r(n)} = 0$.

From Remark 5, 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 4, 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 5.1. *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 $n \rightarrow \infty$.

Proof. If $\lim_{n \rightarrow \infty} \frac{pn}{r(n)} = \infty$, then it is easy to show that at least one condition in Theorem 4.1 is satisfied. Similarly, if $\lim_{n \rightarrow \infty} \frac{pn}{r(n)} = 0$, then all conditions in Theorem 4.2 are satisfied. \square

Example 9 (A closed system with $\alpha_{0,1} = \alpha_{0,2} = 0, \alpha_{1,1} = 2, \alpha_{1,2} = 1$). In this case, we have $K_{0,1}(n) \sim np_n, K_{0,2}(n) \sim n^2p_n, K_{1,1}(n) \sim K_{1,2}(n) \sim K_{2,2}(n) \sim n^4p_n$. Using Theorem 5.1 yields

$$r(n) = \frac{1}{n^3},$$

which is the same threshold as in the base case in [6]. \triangle

Example 10 (An open system with $\alpha_{0,1} = 3, \alpha_{1,1} = \alpha_{0,2} = 2, \alpha_{1,2} = 1$). . In this case, we have $K_{i,j}(n) \sim n^4p_n$ for all (i, j) . Using Theorem 5.1 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. \triangle

Acknowledgements

We gratefully acknowledge support via the Army Research Office through grant W911NF-18-1-0324, and via the William F. Vilas Trust Estate.

A Appendix

The following lemmas have been used in the manuscript. Their proofs are added for completeness.

Lemma A.1. *Let $X \sim B(n, p)$. Then we have*

$$\mathbb{E}\left[\frac{1}{X+1}\right] \leq \frac{1}{np}, \quad \text{and} \quad \mathbb{E}\left[\frac{1}{(X+1)(X+2)}\right] \leq \frac{1}{(np)^2}.$$

Proof. We have

$$\begin{aligned} \mathbb{E}\left[\frac{1}{X+1}\right] &= \sum_{i=0}^n \frac{1}{i+1} \binom{n}{i} p^i (1-p)^{n-i} = \sum_{i=0}^n \frac{n!}{(i+1)!(n-i)!} p^i (1-p)^{n-i} \\ &= \frac{1}{n+1} \frac{1}{p} \sum_{i=0}^n \binom{n+1}{i+1} p^{i+1} (1-p)^{n-i} \leq \frac{1}{np} (p+1-p)^{n+1} \leq \frac{1}{np}. \end{aligned}$$

Similarly, we have

$$\begin{aligned} \mathbb{E}\left[\frac{1}{(X+1)(X+2)}\right] &= \sum_{i=0}^n \frac{1}{(i+1)(i+2)} \binom{n}{i} p^i (1-p)^{n-i} = \sum_{i=0}^n \frac{n!}{(i+2)!(n-i)!} p^i (1-p)^{n-i} \\ &= \frac{1}{(n+1)(n+2)} \frac{1}{p^2} \sum_{i=0}^n \binom{n+2}{i+2} p^{i+2} (1-p)^{n-i} \\ &\leq \frac{1}{(np)^2} (p+1-p)^{n+2} \leq \frac{1}{(np)^2}. \end{aligned}$$

□

Lemma A.2. *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.1. *Let $x_1, x_2, \dots, x_n \in \mathbb{R}_{\geq 0}$, then we have*

$$\prod_{i=1}^n (nx_i)^{x_i} \geq \left(\sum_{i=1}^n x_i\right)^{\sum_{i=1}^n x_i}. \quad (39)$$

Proof. We will prove the corollary by induction. Clearly (39) holds for $n = 1$. Lemma A.2 shows that (39) holds for $n = 2$. Suppose (39) holds for $n = k$. It suffices to show that (39) holds for $n = 2k$ and $n = k - 1$.

First, we will show that (39) holds for $n = 2k$. Applying the inductive hypothesis for the $n = k$ terms x_1, \dots, x_k and the $n = k$ terms x_{k+1}, \dots, x_{2k} , and then applying Lemma A.2 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 (39) holds for $n = k - 1$. Applying the induction hypothesis for the $n = k$ terms $x_1, \dots, x_{k-1}, \frac{1}{k-1} \sum_{i=1}^{k-1} x_i$, we have

$$\begin{aligned} & \prod_{i=1}^{k-1} (kx_i)^{x_i} \left(\frac{k}{k-1} \sum_{i=1}^{k-1} x_i \right)^{\frac{1}{k-1} \sum_{i=1}^{k-1} x_i} \geq \left(\sum_{i=1}^{k-1} x_i + \frac{1}{k-1} \sum_{i=1}^{k-1} x_i \right)^{\sum_{i=1}^{k-1} x_i + \frac{1}{k-1} \sum_{i=1}^{k-1} x_i} \\ \Rightarrow & \prod_{i=1}^{k-1} (kx_i)^{x_i} \left(\frac{k}{k-1} \sum_{i=1}^{k-1} x_i \right)^{\frac{1}{k-1} \sum_{i=1}^{k-1} x_i} \geq \left(\frac{k}{k-1} \sum_{i=1}^{k-1} x_i \right)^{\frac{k}{k-1} \sum_{i=1}^{k-1} x_i} \\ \Rightarrow & \prod_{i=1}^{k-1} (kx_i)^{x_i} \geq \left(\frac{k}{k-1} \sum_{i=1}^{k-1} x_i \right)^{\sum_{i=1}^{k-1} x_i} \Rightarrow \prod_{i=1}^{k-1} ((k-1)x_i)^{x_i} \geq \left(\sum_{i=1}^{k-1} x_i \right)^{\sum_{i=1}^{k-1} x_i}. \end{aligned}$$

Thus we have show that (39) holds for $n = k - 1$, which concludes the proof of the Corollary. \square

References

- [1] David F. Anderson, Daniele Cappelletti, Masanori Koyama, and Thomas G. Kurtz. Non-explosivity of stochastically modeled reaction networks that are complex balanced. *Bulletin of Mathematical Biology*, 80(10):2561–2579, 2018.
- [2] David F. Anderson, Daniele Cappelletti, and Thomas G. Kurtz. Finite time behavior of stochastically modeled chemical systems with absolute concentration robustness. *SIAM Journal on Applied Dynamical Systems*, 16(3), 2017.
- [3] David. F. Anderson and Simon. L. Cotter. Product-form stationary distributions for deficiency zero networks with non-mass action kinetics. *Bulletin of Mathematical Biology*, 78(12), 2016.
- [4] David F. Anderson, Gheorghe Craciun, and Thomas G. Kurtz. Product-form stationary distributions for deficiency zero chemical reaction networks. *Bulletin of Mathematical Biology*, 72(8), 2010.
- [5] David F. Anderson and Tung D. Nguyen. Results on stochastic reaction networks with non-mass action kinetics. *Mathematical Biosciences and Engineering*, 16(4):2118–2140, 2019.
- [6] David F. Anderson and Tung D. Nguyen. Prevalence of deficiency zero reaction networks in an Erdos-Renyi framework. *Accepted to Journal of Applied Probability*, 2021.
- [7] Béla Bollobás and Andrew Thomason. Threshold functions. *Combinatorica*, 7:35–38, 1987.
- [8] Daniele Cappelletti and Carsten Wiuf. Product-form poisson-like distributions and complex balanced reaction systems. *SIAM Journal on Applied Mathematics*, 76(1), 2014.
- [9] Paul Erdős and Alfréd Rényi. On the evolution of random graphs. In *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, pages 17–61, 1960.
- [10] Martin Feinberg. Complex balancing in general kinetic systems. *Archive for Rational Mechanics and Analysis*, 49:187–194, 1972.
- [11] Paul Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983.

- [12] Fritz Horn. Necessary and sufficient conditions for complex balancing in chemical kinetics. *Archive for Rational Mechanics and Analysis*, 49:172–186, 1972.
- [13] Fritz Horn and Roy Jackson. General mass action kinetics. *Archive for Rational Mechanics and Analysis*, 47:187–194, 1972.
- [14] Liming Wang and Eduardo Sontag. On the number of steady states in a multiple futile cycle. *Journal of Mathematical Biology*, 57(1):29–52, 2008.