

# Prevalence of deficiency zero reaction networks in an Erdős-Rényi framework

David F. Anderson\*, Tung D. Nguyen†

January 16, 2021

## Abstract

Reaction networks are commonly used within the mathematical biology and mathematical chemistry communities to model the dynamics of interacting species. These models differ from the typical graphs found in random graph theory since their vertices are constructed from elementary building blocks, i.e., the species. In this paper, we consider these networks in an Erdős-Rényi framework and, under suitable assumptions, derive a threshold function for the network to have a deficiency of zero, which is a property of great interest in the reaction network community. Specifically, if the number of species is denoted by  $n$  and if the edge probability is denoted by  $p_n$ , then we prove that the probability of a random binary network being deficiency zero converges to 1 if  $\frac{p_n}{r(n)} \rightarrow 0$ , as  $n \rightarrow \infty$ , and converges to 0 if  $\frac{p_n}{r(n)} \rightarrow \infty$ , as  $n \rightarrow \infty$ , where  $r(n) = \frac{1}{n^3}$ .

## 1 Introduction

Reaction network models are often used to study the dynamics of the abundances of species from various branches of chemistry and biology. Here the word “species” can refer to different (bio)chemical molecules or different animal species, depending on the context. 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 has been central to many classical results pertaining to the associated dynamical systems [4, 5, 12, 15, 17, 6, 10]. To compute the deficiency, we first note that the vertices of a reaction network, which will be denoted by  $y$  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,  $\delta$ , 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  $\delta = 5 - 2 - 2 = 1$ .

---

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

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

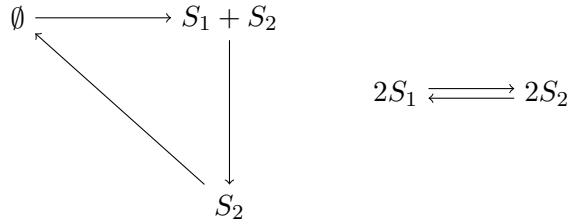


Figure 1: A reaction network with two species:  $S_1$  and  $S_2$ . The vertices are linear combinations of the species over the integers. The directed edges are termed reactions and determine the net change in the counts of the species due to one instance of the reaction. For example, the reaction 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

Given the importance of the deficiency zero property in the literature pertaining to reaction networks, it is natural to ask: how common is this property? There are a number of ways one can tackle this question, including a simple enumeration of all networks of a given size. In fact, the earliest attempt to answer this question can be traced back to work by Horn in 1973 [16]. In that paper, Horn considered all reaction networks with exactly three vertices, each of which satisfies  $\sum_{i=1}^n y_i \leq 2$ , 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  $n \rightarrow \infty$ ) in an Erdős-Rényi random graph framework in which there is an equal probability,  $p_n$ , that there is an edge between any two vertices. 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  $S_1, 2S_1, 3S_1, \dots$ . 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 paper, we restrict ourselves to study so-called “binary” reaction networks, whose vertices satisfy  $\sum_{i=1}^n y_i \leq 2$ . Such models are by far the most common in the literature. Our main finding is that in such a scenario  $r(n) = \frac{1}{n^3}$  is a threshold function in that if  $\frac{p_n}{r(n)} \rightarrow 0$ , as  $n \rightarrow \infty$ , then the probability of deficiency zero converges to 1, whereas if  $\frac{p_n}{r(n)} \rightarrow \infty$ , then the probability of deficiency zero converges to 0. See Theorem 4.1 and Theorem 4.2. Moreover, along the way we prove that in the setting of  $\frac{p_n}{r(n)} \rightarrow 0$ , all the connected components will consist of pairs of vertices, see Lemma 4.2. Intriguingly, paired reaction networks can be found in certain models of autocatalytic cycles related to the study of the origin of life [18].

The remainder of this paper is organized as follows. In Section 2, we briefly review some key terminology of reaction network theory, and provide some preliminary results related to deficiency. In Section 3, we set up the Erdős-Rényi random graph framework for reaction networks. In Section 4, we present our main results, which quantify the prevalence of deficiency zero reaction networks in our chosen framework. Finally, in Section 5, we end with a brief discussion.

## 2 Chemical reaction networks

Here we formally introduce reaction networks and deficiency. Moreover, we collect some preliminary results related to the deficiency of a reaction network.

### 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*.  $\triangle$

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  $\ell$  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 1.** Note that since each connected component must consist of at least two vertices, we have the bound  $\ell \leq \frac{|\mathcal{C}|}{2}$ .

**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  $\dim(S) = s$ .  $\triangle$

Note that  $\dim(S) = s \leq n$ , where  $n$  is the number of species. This fact will be used a number of times in this paper.

**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.  $\triangle$

As discussed in the introduction, we will focus on *binary* reaction networks in this paper. The following type of network will play a key role in the current paper.

**Definition 2.5.** 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.  $\triangle$

## 2.2 Deficiency of a reaction network

**Definition 2.6.** 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,  $\ell$  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.  $\triangle$

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

**Lemma 2.1.** *Let  $n \geq 1$  and let  $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$  be a reaction network with  $n$  species.*

(a)  $\delta$  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  $\zeta_j$  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  $\widehat{\mathcal{R}} \setminus \mathcal{R} = \{y \rightarrow y'\}$ , a single reaction. Let  $\widehat{\delta}$  and  $\delta$  be the deficiencies of the two networks. Then

$$\widehat{\delta} \geq \delta.$$

*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 1), 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

$$0 \geq \frac{|\mathcal{C}|}{2} - n, \quad (1)$$

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  $\zeta_j$  are linearly independent.
- (g) Let  $\ell, s$  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  $y \rightarrow y'$  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

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

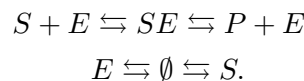
- Case 4:  $y, y' \notin \mathcal{C}$ . In this case, we have  $|\widehat{\mathcal{C}}| = |\mathcal{C}| + 2$ , and  $\hat{\ell} = \ell + 1$ . Similar to the previous cases, we still have  $\hat{s} \leq s + 1$  and thus

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

**Remark 2.** Lemma 2.1(g) implies that if we remove a reaction from a reaction network with deficiency zero, then the resulting network also has deficiency zero. This means deficiency zero is a monotone decreasing property, which guarantees that a threshold function for deficiency zero exists (see [9]).

We will illustrate the concept of deficiency via two examples.

**Example 1** (Enzyme kinetics [5]). Consider a reaction network with species  $\{S, E, SE, P\}$  and associated graph



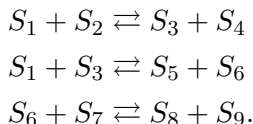
In this example, the reaction network has  $|\mathcal{C}| = 6$  vertices, there are  $\ell = 2$  connected components, and the dimension of the stoichiometric subspace is  $s = 4$ . Thus the deficiency is

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

□

The following example demonstrates that it is sometimes most natural to use Lemma 2.1(f) to verify that a network has a deficiency of zero.

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



This network is *paired* in the sense of Definition 2.5. 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 2.1(f) implies that the deficiency of this network is zero. □

### 3 Erdős-Rényi model for binary reaction networks

As alluded to in the introduction, the vast majority of reaction network models found in the literature are binary. Hence, those are the focus of the current paper.

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 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  $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = c$  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,

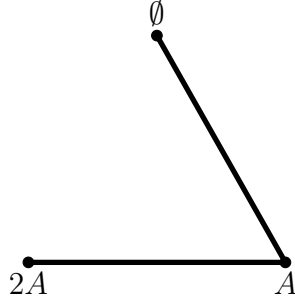


Figure 2: A realization of a random graph when  $n = 1$  and  $p \in (0, 1)$ . The corresponding reaction network is  $\emptyset \rightleftharpoons A \rightleftharpoons 2A$ .

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 we do not need to worry about direction—see Lemma 2.1(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  $\delta_{R_n}$ .

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. These examples can be skipped by those wishing to get directly to the main results.

**Example 3** (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 2 shows one possible realization of the random graph when  $p \in (0, 1)$ . The corresponding reaction network for the particular graph shown in Figure 2 is  $\emptyset \rightleftharpoons A \rightleftharpoons 2A$ .

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  $\ell$  and  $s$  be defined as usual. Recall that 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,

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

□

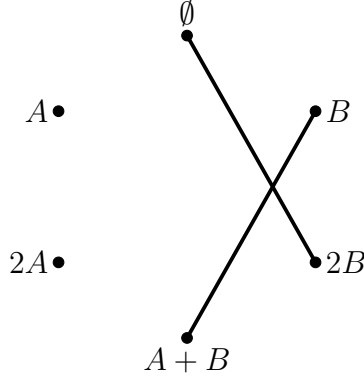
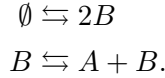


Figure 3: A realization of a random graph when  $n = 2$  and  $p \in (0, 1)$ .

**Example 4** (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 3 illustrates a possible realization of the random graph when  $p \in (0, 1)$ . The corresponding reaction network for the particular graph shown in Figure 2 is



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  $\ell$  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 1 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 1,  $\ell \leq \frac{|\mathcal{C}|}{2}$ . 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:  $\{\emptyset \rightleftharpoons A, A \rightleftharpoons 2A, \emptyset \rightleftharpoons 2A, B \rightleftharpoons A + B\}$ ,  $\{\emptyset \rightleftharpoons B, B \rightleftharpoons 2B, \emptyset \rightleftharpoons 2B, A \rightleftharpoons A + B\}$ , and  $\{A \rightleftharpoons B, 2A \rightleftharpoons 2B, A + B \rightleftharpoons 2A, A + B \rightleftharpoons 2B\}$ . 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  $\{\emptyset \rightleftharpoons A, A \rightleftharpoons 2A, \emptyset \rightleftharpoons 2A\}$ ,  $\{\emptyset \rightleftharpoons B, B \rightleftharpoons 2B, \emptyset \rightleftharpoons 2B\}$ , and  $\{A + B \rightleftharpoons 2A, A + B \rightleftharpoons 2B, 2A \rightleftharpoons 2B\}$ . 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

$$\lim_{n \rightarrow \infty} \mathbb{P}(\delta_{R_n} = 0) = \begin{cases} 0, & \text{if } \lim_{n \rightarrow \infty} \frac{p_n}{r(n)} = \infty \\ 1 & \text{if } \lim_{n \rightarrow \infty} \frac{p_n}{r(n)} = 0. \end{cases} \quad (2)$$

In this paper, we show  $r(n) = \frac{1}{n^3}$ .

## 4 The threshold function for deficiency zero

In Sections 4.1 and 4.2, we will show that the limits in (2) 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  $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 0$ , 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  $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = c$  for some constant  $c \in \mathbb{R}_{>0}$ .

### 4.1 The case $\lim_{n \rightarrow \infty} \frac{p_n}{r(n)} = \infty$

Lemma 2.1(e) provides an upper bound on  $|\mathcal{C}|$  for deficiency zero reaction network. We will utilize this bound to show  $\lim_{n \rightarrow \infty} \mathbb{P}(\delta_{R_n} = 0) = 0$  when  $\lim_{n \rightarrow \infty} \frac{p_n}{r(n)} = \infty$ . 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 4.1.** *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*

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

*Proof.* We require both  $\mathbb{E}(|I|)$  and  $\text{Var}(|I|)$ . 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

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

where the first inequality follows from Bernoulli's inequality.

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

$$\lim_{n \rightarrow \infty} \mathbb{P}(|I| \geq N_n - 2n) = 0. \tag{3}$$

It suffices to prove (3) 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  $\lim_{n \rightarrow \infty} (1 - p_n)^{N_n-1} = \lim_{n \rightarrow \infty} (1 - p_n)^{\frac{1}{p_n}p_n(N_n-1)} = \lim_{n \rightarrow \infty} e^{-p_n(N_n-1)} = 0$ , we have

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

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

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

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

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

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:

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

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

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

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

If  $\alpha_n \gg n$ , we have

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

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

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

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

We are now ready to provide the first main theorem.

**Theorem 4.1.** *For  $\lim_{n \rightarrow \infty} \frac{p_n}{r(n)} = \infty$ , the following holds*

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

*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 2.1(e) implies that if the network is deficiency zero, we must have

$$|I| \geq N_n - 2n. \quad (4)$$

From (4), we have

$$\mathbb{P}(\delta_{R_n} = 0) \leq \mathbb{P}(|I| \geq N_n - 2n). \quad (5)$$

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 4.1. Hence, using Lemma 4.1 we have

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

## 4.2 The case $\lim_{n \rightarrow \infty} \frac{p_n}{r(n)} = 0$

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

We will show in Lemma 4.2 that as  $n \rightarrow \infty$ , a random reaction network as described in Section 3 with  $\lim_{n \rightarrow \infty} \frac{p_n}{r(n)} = 0$  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 4.2.** *Suppose  $\lim_{n \rightarrow \infty} \frac{p_n}{r(n)} = 0$ . Then*

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

*Proof.* We have

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

It is a well-known fact in random graph theory (for example, see [13]) that for  $p_n \ll \frac{1}{n^3} \ll \frac{1}{N_n}$  we have

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

Thus it suffices to show

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

We follow the notation in [8] 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 [8], 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  $\{\text{not paired, only trees}\} \subset \cup_{k=3}^{N_n} \{T_k(n) > 0\}$ , 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  $\lim_{n \rightarrow \infty} N_n^3 e^3 p_n^2 = 0$  and  $\sum_{k=0}^{N_n-3} (N_n e p_n)^k$  is bounded. Thus we have

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

and the proof is complete  $\square$

**Remark 3.** 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 4.3.** 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

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

*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  $n \rightarrow \infty$ , 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  $j + 1^{\text{st}}$  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

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

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

Therefore, for  $n$  large enough, we have

$$\mathbb{P}(A_n | R_n \text{ is } k_n\text{-paired}) = \prod_{j=0}^{k_n-1} \mathbb{P}(A_n^{j+1} | \cap_{i=1}^j A_n^i, R_n \text{ is } k_n\text{-paired}) \geq \left(1 - \frac{21}{n}\right)^{k_n} \geq 1 - \frac{21k_n}{n} \quad (6)$$

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

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

and the proof is complete.  $\square$

Lemma 4.3 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  $n \rightarrow \infty$ , this set of reaction vectors will be linearly independent.

**Proposition 4.1.** *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,  $\Gamma_n$  will have full column rank with probability converging to one, as  $n \rightarrow \infty$ .*

*Proof.* Let  $I_n$  be the event that all column vectors of  $\Gamma_n$  are linearly independent. It suffices to show

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

We denote the  $k_n$  column vectors of  $\Gamma_n$  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

$$\mathbb{P}(I_n^c) \leq \sum_{\ell=2}^{k_n} \sum_{|T|=\ell} \mathbb{P}(V_n^T \text{ is minimally dependent}) = \sum_{\ell=2}^{k_n} \binom{k_n}{\ell} \mathbb{P}(B_\ell) \quad (7)$$

where  $B_\ell$  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  $M_\ell$  whose columns are the vectors in  $V_n^T$ . Note that the set  $V_n^T$  being minimally dependent implies that  $M_\ell$  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  $M_\ell$  has at least 2 entries. Since each column of  $M_\ell$  has exactly 4 entries,  $M_\ell$  has exactly  $4\ell$  entries. Therefore, the number of non-zero rows in  $M_\ell$  must be at most  $2\ell$  and the number of zero rows in  $M_\ell$  must be at least  $n - 2\ell$ . Combining all of the arguments above, we must have

$$\mathbb{P}(B_\ell) \leq \mathbb{P}(M_\ell \text{ has at least } n - 2\ell \text{ zero rows}). \quad (8)$$

We denote the row vectors of  $M_\ell$  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

$$\mathbb{P}(M_\ell \text{ has at least } n - 2\ell \text{ zero rows}) \leq \sum_{|R|=n-2\ell} \mathbb{P}(W_n^R = 0) = \binom{n}{n-2\ell} \mathbb{P}(C_\ell) \quad (9)$$

where  $C_\ell$  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  $C_\ell$  involves picking  $\ell$  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 4.3). 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

$$\mathbb{P}(C_\ell) \leq \prod_{j=0}^{\ell-1} \frac{\binom{2\ell}{2} \binom{2\ell-2}{2}}{\binom{n}{2} - 2j \binom{n-2}{2} - 2j} \leq \prod_{j=0}^{\ell-1} \frac{\binom{2\ell}{2} \binom{2\ell-2}{2}}{\frac{1}{4} \binom{n}{2} \binom{n-2}{2}} \leq 4 \left( \frac{2\ell}{n} \right)^4.$$

Plugging the above into (9), we see

$$\begin{aligned} \mathbb{P}(M_\ell \text{ has at least } n - 2\ell \text{ zero rows}) &\leq \binom{n}{n-2\ell} 4 \left( \frac{2\ell}{n} \right)^{4\ell} \leq \frac{n^{2\ell}}{(2\ell)!} 4 \left( \frac{2\ell}{n} \right)^{4\ell} \\ &\leq \frac{4n^{2\ell}}{\sqrt{2\pi} (2\ell/e)^{2\ell}} \left( \frac{2\ell}{n} \right)^{4\ell} = \frac{4}{\sqrt{2\pi}} \left( \frac{2\ell e}{n} \right)^{2\ell}. \end{aligned} \quad (10)$$

Now combining (7), (8), and (10), we have

$$\begin{aligned}
\mathbb{P}(I_n^c) &\leq \sum_{\ell=2}^{k_n} \binom{k_n}{\ell} \frac{4}{\sqrt{2\pi}} \left(\frac{2\ell e}{n}\right)^{2\ell} \leq \sum_{\ell=2}^{k_n} \frac{k_n^\ell}{\ell!} \frac{4}{\sqrt{2\pi}} \left(\frac{2\ell e}{n}\right)^{2\ell} \\
&\leq \sum_{\ell=2}^{k_n} \frac{k_n^\ell}{\sqrt{2\pi}(\ell/e)^\ell} \frac{4}{\sqrt{2\pi}} \left(\frac{2\ell e}{n}\right)^{2\ell} = \sum_{\ell=2}^{k_n} \frac{2}{\pi} \left(\frac{4\ell e^3 k_n}{n^2}\right)^\ell \\
&\leq \sum_{\ell=2}^{\infty} \frac{2}{\pi} \left(\frac{4e^3 k_n^2}{n^2}\right)^\ell \leq c \frac{k_n^4}{n^4}.
\end{aligned} \tag{11}$$

for some constant  $c > 0$ , since  $k_n \ll n$ . Thus we have

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

which concludes the proof.  $\square$

We return to the setting of reaction networks with our final key lemma.

**Lemma 4.4.** *Suppose that  $k_n \ll n$ . Then we have*

$$\lim_{n \rightarrow \infty} \mathbb{P}(\delta_{R_n} = 0 | R_n \text{ is } k_n\text{-paired}) = 1. \tag{12}$$

*Proof.* Let  $R_n$  be a  $k_n$ -paired reaction network, where  $k_n \ll n$ . From Lemma 2.1,  $R_n$  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 4.3, 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

$$\begin{aligned}
\mathbb{P}(\delta_{R_n} = 0 | R_n \text{ is } k_n\text{-paired}) &= \mathbb{P}(I_n | R_n \text{ is } k_n\text{-paired}) \\
&\geq \mathbb{P}(I_n | A_n, R_n \text{ is } k_n\text{-paired}) \mathbb{P}(A_n | R_n \text{ is } k_n\text{-paired}).
\end{aligned} \tag{13}$$

Utilizing (11) in Proposition 4.1, we have

$$\mathbb{P}(I_n^c | A_n, R_n \text{ is } k_n\text{-paired}) \leq c \frac{k_n^4}{n^4} \tag{14}$$

for some constant  $c > 0$ . Thus using (6), (13) and (14), we must have

$$\mathbb{P}(\delta_{R_n} = 0 | R_n \text{ is } k_n\text{-paired}) \geq \left(1 - c \frac{k_n^4}{n^4}\right) \left(1 - \frac{21k_n}{n}\right), \tag{15}$$

Since  $k_n \ll n$ , taking the limit of (15) concludes the proof of the lemma.  $\square$

Combining Lemmas 4.2, 4.3, and 4.4, we are ready to state the main theorem for this section.

**Theorem 4.2.** *Suppose  $p_n \ll \frac{1}{n^3}$ , then*

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



*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

$$\lim_{n \rightarrow \infty} \mathbb{P}(\delta_{R_n} = 0, R_n \text{ is not paired}) = 0$$

due to Lemma 4.2. Therefore it suffices to show

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

Noting that for deficiency zero models, the number of reversible reaction vectors is bounded above by  $n$ , we have

$$\begin{aligned} \mathbb{P}(\delta_{R_n} = 0, R_n \text{ is paired}) &= \sum_{i=1}^n \mathbb{P}(\delta_{R_n} = 0, G_n \text{ is } i\text{-paired}) \\ &= \sum_{i=1}^n \mathbb{P}(\delta_{R_n} = 0 | R_n \text{ is } i\text{-paired}) \mathbb{P}(R_n \text{ is } i\text{-paired}) \\ &= \sum_{i=1}^n \mathbb{P}(\delta_{R_n} = 0 | R_n \text{ is } i\text{-paired}) \frac{N_n!}{i!2^i(N_n - 2i)!} p_n^i (1 - p_n)^{N_n(N_n-1)/2-i} \\ &\geq \sum_{i=1}^n \mathbb{P}(\delta_{R_n} = 0 | R_n \text{ is } i\text{-paired}) \frac{(N_n - 2i)^{2i}}{i!2^i} p_n^i (1 - p_n)^{N_n(N_n-1)/2-i} \quad (16) \end{aligned}$$

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  $\lim_{n \rightarrow \infty} k_n = \infty$  and  $N_n^2 p_n \ll k_n \ll n$ . Cutting off the last  $n - k_n$  terms from (16), 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 (15) in Lemma 4.4.

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

$$\lim_{n \rightarrow \infty} (1 - p_n)^{N_n^2/2} e^{\lambda_n} = 1,$$

or

$$\lim_{n \rightarrow \infty} \frac{N_n^2}{2} \ln(1 - p_n) + \lambda_n = 0.$$

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,

$$\frac{N_n^2}{2} \ln(1 - p_n) + \lambda_n \geq -\frac{N_n^2}{2} (p_n + p_n^2) + \lambda_n = \frac{p_n}{2} (-4k_n N_n + 4k_n^2) - \frac{N_n^2 p_n^2}{2}.$$

Since  $k_n \ll n$ ,  $N_n \sim n^2$  and  $p_n \ll \frac{1}{n^3}$ , we have

$$\lim_{n \rightarrow \infty} \frac{p_n}{2} (-4k_n N_n + 4k_n^2) = 0, \quad \text{and,} \quad \lim_{n \rightarrow \infty} \frac{N_n^2 p_n^2}{2} = 0.$$

Thus

$$\lim_{n \rightarrow \infty} \frac{N_n^2}{2} \ln(1 - p_n) + \lambda_n = 0,$$

which concludes the proof of the theorem.  $\square$

## 5 Discussion

This work stemmed from a natural question pertaining to reaction networks: given the importance of deficiency zero in the reaction network literature, can we quantify how prevalent the condition is? In the Erdős-Rényi framework we have chosen here, we have provided a threshold function,  $r(n) = \frac{1}{n^3}$ , for the property in that if  $p_n/r(n) \rightarrow 0$ , then the probability of deficiency zero converges to 1, and if  $p_n/r(n) \rightarrow \infty$ , then the probability of deficiency zero converges to 0.

We do not make the claim that the framework selected here is the only, or even the most, biologically relevant. Instead, having equal probabilities for each edge puts as few assumptions on our model as possible, thereby making it a reasonable starting point for analysis. In fact, there are multiple avenues for future research, and we list just a few here.

- One may want to study models in which some added structure is known. For example, our assumption of equal probabilities would need to be relaxed in those contexts where different reaction types are more likely to appear in the network than others (such as when in-flows and out-flows of species are common). This would necessitate the use of a stochastic block model framework. We have carried out such an analysis in [7].

- In the setting of molecular biology some proteins may be more active and interact with many other proteins while some proteins may be relatively inactive and have fewer interactions. In such cases, we can study random reaction networks under a more general random graph framework such as the Chung-Lu model, where vertices can be assigned different weights [11].
- Situations can arise in which some species are chemostated, which keeps their concentrations constant. In such a case we may want to focus on the asymptotic behavior of “sub-networks”, which consist of the species not being chemostated, instead of the whole network. The study of sub-networks may also be useful in the multi-scale settings, where we want to focus on a subset of “discrete” species which are in low abundances and behave differently than those in high abundance. [3].
- There are other meaningful topological features beside deficiency zero that we could study with our approach. Some features of interest are deficiency one (together with additional graphical features) as in [19], endotactic, strongly endotactic, and asyphonic as in [1, 2, 14].

The analysis and methods developed here will, to varying degrees, be applicable to each of the situations listed above.

## Acknowledgements

We thank Robin Pemantle for influencing the formulation of the question posed in this paper. We also thank an anonymous reviewer whose comments greatly influenced the final version of this paper. We gratefully acknowledge grant support from the Army Research Office via grant W911NF-18-1-0324.

## References

- [1] Andrea Agazzi, Amir Dembo, and Jean-Pierre Eckmann. Large deviations theory for markov jump models of chemical reaction networks. *Annals of Applied Probability*, 28(3):1821–1855, 2018.
- [2] Andrea Agazzi, Amir Dembo, and Jean-Pierre Eckmann. On the geometry of chemical reaction networks: Lyapunov function and large deviations. *Journal of Statistical Physics*, 172(2):321–352, 2018.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] 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.

- [7] David F Anderson and Tung D Nguyen. Deficiency zero for random reaction networks under a stochastic block model framework. *arXiv preprint arXiv:2010.07201*, 2020.
- [8] Béla Bollobás. *Introduction to Random Graphs*. Cambridge University Press, 2001.
- [9] Béla Bollobás and Andrew Thomason. Threshold functions. *Combinatorica*, 7:35–38, 1987.
- [10] Daniele Cappelletti and Carsten Wiuf. Product-form poisson-like distributions and complex balanced reaction systems. *SIAM Journal on Applied Mathematics*, 76(1), 2014.
- [11] Fan Chung and Linyuan Lu. The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences*, 99(25):15879–15882, 2002.
- [12] Martin Feinberg. Complex balancing in general kinetic systems. *Archive for Rational Mechanics and Analysis*, 49:187–194, 1972.
- [13] Alan Frieze and Michał Karoński. *Random Graphs*. Cambridge University Press, 2016.
- [14] Manoj Gopalkrishnan, Ezra Miller, and Anne Shiu. A geometric approach to the global attractor conjecture. *SIAM Journal on Applied Dynamical Systems*, 13(2):758–797, 2014.
- [15] Fritz Horn. Necessary and sufficient conditions for complex balancing in chemical kinetics. *Archive for Rational Mechanics and Analysis*, 49:172–186, 1972.
- [16] Fritz Horn. Stability and complex balancing in mass-action systems with three short complexes. *Proceedings of the Royal Society of London*, 334:331–342, 1973.
- [17] Fritz Horn and Roy Jackson. General mass action kinetics. *Archive for Rational Mechanics and Analysis*, 47:187–194, 1972.
- [18] Zhen Peng, Alex Plum, Praful Gagrani, and David A. Baum. An ecological framework for the analysis of prebiotic chemical reaction networks and their dynamical behavior. Submitted, arXiv: <https://arxiv.org/abs/2001.02533>, 2020.
- [19] Guy Shinar and Martin Feinberg. Structural sources of robustness in biochemical reaction networks. *Science*, 327(5971):1389–1391, 2010.