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

David F. Anderson∗, Tung D. Nguyen†

October 28, 2019

Abstract

In the mathematical study of reaction networks, many of the classical results pertain to models that have a deficiency of zero. In particular, for deterministic models it is well known that weak reversibility and a deficiency zero of the reaction network imply that the model is complex balanced. In the stochastic setting it is known that weak reversibility and a deficiency of zero imply the existence of a stationary distribution that is a product of Poissons.

Given that deficiency zero models play such a significant role in the mathematical study of reaction networks, a natural question is how prevalent are they? In order to answer this question, we consider reaction networks under an Erdős-Rényi random graph framework. In particular, we start with \( n \) species, and then let our possible vertices be all zeroth, first, and second order complexes that can be produced from the \( n \) species. Edges, or reversible reactions, between two arbitrary complexes then occur independently with probability \( p_n \). We establish a function, \( r(n) \), termed a threshold function, such that the probability of the random network being deficiency zero converges to 1 if \( p_n \ll r(n) \) and converges to 0 if \( p_n \gg r(n) \).

1 Introduction

Reaction network models are often used to study the dynamics of the abundances of species from various branches of chemistry, ecology, and biology. These networks take the form of directed graphs in which the vertices, often termed complexes, are linear combinations of the species and the directed edges are termed reactions. To each such graph a quantity termed the deficiency can be computed, and this quantity has been central to most classical results in the field going all the way back to the seminal works of Horn, Jackson, and Feinberg in 1972 [12, 18, 20]. In particular, the oldest deficiency result is called the Deficiency Zero Theorem, which was proved in [18] and reformulated in [12, 20]. The theorem states that if the deficiency of the graph is zero and if the graph is weakly reversible (i.e., each connected component is strongly connected), then there exists a unique, locally stable equilibrium for an associated ordinary differential equation model of the molecular concentrations. Moreover, the theorem guarantees that the equilibrium is a so-called “complex-balanced” equilibrium and that every equilibrium of the model is also complex-balanced.

There is another classical result about reaction networks whose associated graph has a deficiency of zero that is widely used in the stochastic setting. In [6], it was shown that reaction networks

∗Department of Mathematics, University of Wisconsin, Madison, USA. anderson@math.wisc.edu
†Department of Mathematics, University of Wisconsin, Madison, USA. nguyen34@math.wisc.edu.
with a deficiency of zero and that are weakly reversible admit a stationary distribution that is a product of Poisson distributions. A partial converse to this result was shown in [10].

Further, deficiency zero networks have appeared frequently in recent developments of reaction network theory. In [4], an important assumption was made in which the “fast subnetwork” consisting of low-abundance species has deficiency zero. In [5, 8], the stationary distribution of deficiency zero reaction networks with non-mass action kinetics was examined.

Given the significant role of deficiency zero in reaction network theory, one may ask: are such networks common? The earliest attempt to answer this question can be traced back to some work by Horn in 1973 [19]. In that paper, Horn considered all reaction networks with exactly 3 binary complexes, but no condition on the number of species. Horn found 43 isomorphism classes of such networks, and among these, 41 have deficiency zero.

Our goal in this paper is to quantify the prevalence of deficiency zero networks when the reaction networks have a large number of complexes. Of course, the prevalence will depend upon the particular class of networks considered and deficiency zero reaction networks may be prevalent in one context but uncommon in others. In this paper, we choose to study reaction networks via limit theorems in an Erdős-Rényi random graph framework in which there is an equal probability that there is a reaction between any two complexes. This framework will be discussed in more detail later in the paper. Under this framework, we can answer a number of questions, such as the following.

1. If $p_n$ is the probability that two complexes have a reaction between them, for which values of $p_n$ do we have a high probability of observing a deficiency zero reaction network?

2. If we do observe a network with deficiency zero, what does it look like? Does it contain a few big clusters, or many small components?

The Erdős-Rényi framework we choose here, in which all possible reactions have an equal probability of appearing in the resulting network, is just one possible choice to study random reaction networks. Having equal probabilities puts as few assumptions on our model as possible, thereby making it a reasonable starting point for our analysis. However, it could be that one wants 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 or species are more likely to appear in the network than others (such as when in-flows and out-flows of species are common). Another change to our modeling assumptions could arise in situations where some species are chemostated, which keeps their concentrations constant. In this 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. These contexts will be discussed at the end of the paper. However, it is worth noting that some of the machinery we use in the current work can be adapted to these modeling situations with additional assumptions. Therefore, we view this work as a starting point that we can push to different directions in various practical settings.

We note that the approach we are taking is in some ways similar to the strategy of “significant feature detection” in the field of network biology [23]. The strategy involves studying the prevalence of certain topological features in networks found empirically, in comparison to networks generated by randomization processes. If a feature is statistically significant in real networks in contrast to randomized networks, then that feature could be the result of some underlying “design” processes such as evolution [23].
The remainder of this paper is organized as follows. In Section 2, we collect some of the notation we use throughout the paper. In Section 3, we briefly review some key definitions of reaction network theory, and introduce some classical and current results related to deficiency zero. In Section 4, we set up the Erdős-Rényi random framework which will be connected to the reaction network setting. In Section 5, we present our main results, which quantify the prevalence of deficiency zero reaction networks in our chosen framework. Finally in Section 6, we discuss directions for future research.

2 Notation

We introduce some common notation that will be used throughout the paper.

1. For \( u, v \in \mathbb{R}_{\geq 0}^n \), we use the conventions
\[
u^v = \prod_{i=1}^{n} u_i^{v_i},
\]
with \( 0^0 \) always taken to be zero, and
\[
u! = \prod_{i=1}^{n} u_i!.
\]

2. Let \( \{a_n\}_{n=0}^{\infty} \) and \( \{b_n\}_{n=0}^{\infty} \in \mathbb{R} \) be two sequences. We write \( a_n \sim b_n \) if
\[
\lim_{n \to \infty} \frac{a_n}{b_n} = c,
\]
for some constant \( c \). We write \( a_n \ll b_n \) or \( b_n \gg a_n \) if
\[
\lim_{n \to \infty} \frac{a_n}{b_n} = 0.
\]

3 Chemical reaction networks

3.1 Reaction networks and key definitions

Let \( \{S_1, \ldots, S_n\} \) be a set of \( n \) species undergoing a finite number of reaction types. We denote a particular reaction by \( y \to y' \), where \( y \) and \( y' \) are linear combinations of species on \( \mathbb{N} \) 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 called complexes of the system. More specifically \( y \) is called the source complex and \( y' \) is called the product complex. A complex can be both a source complex and a product complex. For convenience, we associate each complex with a vector in \( \mathbb{Z}_{\geq 0}^n \), whose coordinates are the number of molecules of the corresponding species in the complex. As is common in the reaction network literature, both ways of representing complexes will be used interchangeably throughout the paper. For example, if the system has 2 species \( \{S_1, S_2\} \), the reaction \( S_1 + S_2 \to 2S_2 \) has \( y = S_1 + S_2 \), which is associated with the vector \( (1, 1) \), and \( y' = 2S_2 \), which is associated with the vector \( (0, 2) \).

**Definition 3.1.** Let \( \mathcal{S} = \{S_1, \ldots, S_n\} \), \( \mathcal{C} = \cup_{y \to y'} \{y, y'\} \), and \( \mathcal{R} = \cup_{y \to y'} \{y \to y'\} \) be the sets of species, complexes and reactions respectively. The triple \( \{\mathcal{S}, \mathcal{C}, \mathcal{R}\} \) is called a reaction network.
To each reaction network \( \{S, C, R\} \), there is a unique directed graph constructed as follows. The nodes of the graph are the complexes. A directed edge is placed from \( y \) to \( y' \) if and only if \( y \to y' \in R \). Each connected component of the graph is called a linkage class. We denote by \( \ell \) the number of linkage class.

**Definition 3.2.** A reaction network \( \{S, C, R\} \) is called weakly reversible if each connected component of the associated directed graph is strongly connected.

**Definition 3.3.** The linear subspace \( S = \text{span}\{y' - y\} \) generated by all reaction vectors is called the stoichiometric subspace of the network. For \( c \in \mathbb{R}^n_\geq \) we say \( c + S = \{x \in \mathbb{R}^n | x = c + s \text{ for some } s \in S\} \) is a stoichiometric compatibility class, \((c + S) \cap \mathbb{R}^m_\geq \) is a non-negative stoichiometric compatibility class, and \((c + S) \cap \mathbb{R}^m_\gt \) is a positive stoichiometric compatibility class. Denote \( \dim(S) = s \).

**Definition 3.4.** A complex is called binary if the sum of its coefficients is 2. A complex is called unary if the sum of its coefficients is 1 (only contains 1 molecule of 1 species). The complex \( \emptyset \) is said to be of zeroth order.

**Definition 3.5.** A reaction network \( \{S, C, R\} \) is called binary if each complex is binary, unary, or of zeroth order.

In later sections, we will focus on binary reaction networks, which was also discussed by Horn in [19].

### 3.2 Dynamical models of reaction networks

#### 3.2.1 Deterministic model

In the deterministic case, the evolution of the species concentration \( x(t) \in \mathbb{R}^n_\geq \) is modeled as the solution to the ODE

\[
\dot{x} = \sum_{y \to y' \in R} (y' - y)\lambda_{y \to y'}(x)
\]

for some functions \( \lambda_{y \to y'} : \mathbb{R}^n_\geq \to \mathbb{R}_\geq \) and an initial condition \( x(0) \). Such functions \( \lambda_{y \to y'}(x) \) are called intensity functions. The most common choice for intensity functions is deterministic mass action kinetics:

\[
\lambda_{y \to y'}(x) = \kappa_{y \to y'}x^y,
\]

where the constants \( \kappa_{y \to y'} \in \mathbb{R}_\gt \) are called rate constants.

Note that under the assumption of mass action kinetics, the solution to (1) exists and is unique for any initial condition, since the rates \( \lambda_{y \to y'} \) are polynomials and therefore locally Lipschitz. In contrast, global existence is not guaranteed, and in case of a blow-up at a finite time \( t^\star \) we consider the solution to (1) only in the interval \([0, t^\star)\).

#### 3.2.2 Stochastic model

In the stochastic case, the evolution of the species count \( X(t) \in \mathbb{Z}^n_\geq \) is modeled as a continuous time Markov chain with state space in \( \mathbb{Z}^n_\geq \). The Kolmogorov’s forward equation for the model is given by

\[
\frac{d}{dt} \mathbb{P}_\mu(x,t) = \sum_{y \to y' \in R} \lambda_{y \to y'}(x - y' + y)\mathbb{P}_\mu(x - y' + y, t) - \sum_{y \to y' \in R} \lambda_{y \to y'}(x)\mathbb{P}_\mu(x, t)
\]
where $P_\mu(x,t)$ represents the probability that $X(t) = x \in \mathbb{Z}_{\geq 0}$ given an initial distribution of $\mu$. The functions $\lambda_{y\to y'}$ are called stochastic intensity functions. The most common choice for stochastic intensity functions are stochastic mass action kinetics:

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

which approximates the deterministic mass action kinetics when $x_i$ are large.

The generator for the Markov chain is the operator $A$, defined by

$$Af(x) = \sum_{y\to y'\in R} \lambda_{y\to y'}(x)(f(x+y' - y) - f(x)),$$

where $f$ is any function defined on the state space. A more detailed construction of the stochastic model can be found in [7]. Typically, in many practical contexts, the stochastic models are simulated via Gillespie’s algorithm [15, 16] or the next reaction method [3, 14]. In case of an explosion occurring at a finite time $T_\infty$, we only consider the process up to $T_\infty$.

### 3.3 Deficiency and related results

**Definition 3.6.** The deficiency of a chemical reaction network $\{S, C, R\}$ is $\delta = |C| - \ell - s$, where $|C|$ is the number of complexes, $\ell$ is the number of linkage classes, and $s$ is the dimension of the stoichiometric subspace of the network.

**Remark 1.** Note that since each linkage class must consist of at least two complexes, we have the bound $\ell \leq \frac{|C|}{2}$.

**Remark 2.** Since we will be studying randomly generated networks in this paper, there is a positive probability that a generated network has no reactions, and hence no complexes. We term such a network the empty network, and note that its deficiency is zero.

The assumption that a network has a deficiency of zero has been central to the most classical results in reaction network theory, both in deterministic and stochastic settings.

**Theorem 3.1 (The Deficiency Zero theorem [12, 18, 20]).** Consider a chemical reaction network $\{S, C, R\}$ which is deficiency zero and weakly reversible. Assume that the network admits deterministic mass action kinetics. Then, for any choice of rate constants $\{\kappa_{y\to y'}\}$, the system has exactly one equilibrium concentration in each positive stoichiometric compatibility class and that equilibrium concentration is locally asymptotically stable.

Furthermore, every equilibrium $c$ of the model is complex-balanced. That is, for each complex $z \in C$

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

where the sum on the left, respectively right, is over those reactions with $z$ as the source, respectively product, complex.

A value $c$ satisfying (2) is called a complex-balanced equilibrium. At such equilibria the flux flowing into a complex is equal to the flux flowing out of that complex.
Theorem 3.2 (Product form stationary distribution [6]). Consider a chemical reaction network \( \{S, C, R\} \) which is deficiency zero and weakly reversible. Assume further that the network has stochastic mass action kinetics. Then for any choice of rate constants \( \{\kappa_{y \to y'}\} \), the model admits a stationary distribution consisting of the product of Poisson distributions,

\[
\pi(x) = \frac{c^x}{x!} e^{-\|c\|_1}, \quad x \in \mathbb{Z}^n_{\geq 0}
\]

where \( c \) is a complex-balanced equilibrium for the deterministic system.

Note that the complex-balanced equilibrium \( c \) in the statement of Theorem 3.2 is guaranteed to exist by Theorem 3.1.

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

Example 1 (Enzyme kinetics [6]).

\[
S + E \rightleftharpoons SE \rightleftharpoons P + E \\
E \rightleftharpoons \emptyset \rightleftharpoons S.
\]

In this example, the reaction network has \( |C| = 6 \) complexes, there are \( l = 2 \) linkage classes and the dimension of the stochiometric subspace is \( s = 4 \). Thus the deficiency is

\[
\delta = 6 - 2 - 4 = 0.
\]

Both theorems above can be applied. The network, when modeled deterministically, has a unique, locally stable equilibrium concentration in each compatibility class. When modeled stochastically, the network admits a stationary distribution that is a product of Poissons.

It is often the case that the reaction \( P + E \to SE \) is not part of the model. In that case, the deficiency is still zero but the above theorems no longer hold since the model is no longer weakly reversible. \[\square\]

Example 2 (Futile cycle enzyme [22]).

\[
S + E \rightleftharpoons SE \to P + E \\
P + F \rightleftharpoons PF \to S + F.
\]

In this example, the reaction network has \( |C| = 6 \) complexes, there are \( l = 2 \) linkage classes and the dimension of the stochiometric subspace can be calculated, which yields \( s = 3 \). Thus the deficiency is

\[
\delta = 6 - 2 - 3 = 1,
\]

and the reaction network is not of deficiency zero. Moreover, the reaction network is not weakly reversible. Thus, both theorems above cannot be applied to this reaction network. In fact, when modeled deterministically, the network may have up to 3 stable steady states [22]. \[\square\]

4 Erdős-Rényi model for binary CRNs

In this section we set up our Erdős-Rényi model. There are several approaches we could take, including the following:
1. we could consider binary reaction networks, and let the number of species go to infinity,
2. we could have a finite number of species, and let the maximum coefficients of all complexes
go to infinity,
3. we could scale both the number of species and the coefficients of the complexes.

The second approach is not suitable for studying large networks since there is an upper bound to
the number of complexes a deficiency zero network can have when the number of species is fixed (see
equation (3) below). The third approach would require two scaling factors, and thus complicate
our set up without a clear motivation. In fact, it is generally assumed in the chemistry and biology
literature that most reaction networks are binary as it is quite unlikely for three molecules to
simultaneously interact. Thus, we choose to work with the first approach and focus on binary
reaction networks only.

Let the set of species be $\mathcal{S} = \{S_1, S_2, \ldots, S_n\}$. We consider binary reaction networks with
species in $\mathcal{S}$. The set of all possible complexes is then
$$\mathcal{C}_0 = \{\emptyset, S_i, S_i + S_j\} \quad \text{for} \quad 1 \leq i, j \leq n.$$ 

For a given $n$, we denote $N_n = |\mathcal{C}_0|$, the cardinality of $\mathcal{C}_0$. Thus, $N_n$ is the total number of
possible unary, binary, and zeroth order complexes 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
$$n \sim \sqrt{2N_n}.$$

We consider an Erdős-Rényi random graph $G_n(n, p_n)$, which we will simply denote $G_n$ through-
out, where the set of vertices is the set of all possible binary complexes $\mathcal{C}_0$, and the probability that
there is an edge between any 2 particular vertices is $p_n$, independently of all other edges. Each
random graph now corresponds to a reaction network in the following way,

1. each vertex with positive degree represents a complex in the reaction network, and
2. each edge represents a reaction (we can assume all reactions are reversible so we do not need
to worry about direction).

Next, we provide two simple examples when the number of species is small, and thus we are
able to explicitly compute the probability of the random network being deficiency zero.

**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 $\mathcal{C}_0 = \{\emptyset, A, 2A\}$. The figure below shows one
realization of the random graph with $p = \frac{1}{2}$
The corresponding reaction network is
$$\emptyset \rightleftharpoons A \rightleftharpoons 2A.$$ 

Since the dimension of the stochiometric subspace $s$ is bounded above by the number of species
$n = 1$, there are only two possibilities: $s = 0$ or $s = 1$. When $s = 0$, the network is empty, hence it
has deficiency zero by our convention. Thus we have
$$\mathbb{P}(s = 0) = (1 - p)^3.$$
We now consider the event that \( \{ \delta = 0, s = 1 \} \), which takes place in the setting of this example if and only if precisely two complexes appear in the resulting network. Since having exactly two complexes corresponds to a graph with only one edge, we have

\[
P(\delta = 0, s = 1) = 3p(1 - p)^2.
\]

Combining the 2 cases, the probability of a random binary reaction network with one species being deficiency zero is

\[
P(\delta = 0) = (1 - p)^3 + 3p(1 - p)^2.
\]

\( \square \)

**Example 4** (The case with \( n = 2 \) species). Denote the set of species by \( S = \{ A, B \} \). The set of vertices is \( \mathcal{C}_0 = \{ \emptyset, A, B, 2A, 2B, A + B \} \). The figure below illustrates a realization of the random graph with \( p = \frac{1}{6} \).

The corresponding reaction network is

\[
\emptyset \rightleftharpoons 2B \\
B \rightleftharpoons A + B.
\]

We wish to compute the probability that the deficiency of a resulting network is zero and there are now 3 possibilities: \( s = 0, s = 1, \) or \( s = 2 \). The case \( s = 0 \) is similar to the previous example,

\[
P(\delta = 0, s = 0) = P(s = 0) = (1 - p)^{15},
\]

since we have a total of \( \binom{6}{2} = 15 \) possible edges.
We turn to the event \( \{ \delta = 0, s = 1 \} \). Since we must have \( |C| = \ell + 1 \), and by Remark 1 we have \( \ell \leq \frac{|C|}{2} \), we may conclude that \( |C| \leq 2 \). As the network cannot be empty with \( s = 1 \), we have \( |C| = 2 \). As in the previous example, this corresponds to a graph with only one edge. Thus

\[
\mathbb{P}(\delta = 0, s = 1) = 15p(1 - p)^{14}.
\]

We turn to the event \( \{ \delta = 0, s = 2 \} \). In this case we have \( |C| = \ell + 2 \) and, again by Remark 1, \( \ell \leq \frac{|C|}{2} \). Combining these two facts yields \( |C| \leq 4 \). In addition, the fact that \( s = 2 \) ensures \( |C| \geq 3 \). If \( |C| = 3 \), then \( \ell = 1 \) and the corresponding graph must have either 2 or 3 edges. If \( |C| = 4 \), then \( \ell = 2 \) and the corresponding graph must have 2 edges. Thus

\[
\mathbb{P}(\delta = 0, s = 2) = \mathbb{P}(\delta = 0, s = 2, 2 \text{ edges}) + \mathbb{P}(\delta = 0, s = 2, 3 \text{ edges}).
\]

If 2 edges are present in the graph, by excluding the configurations with positive deficiency, we have

\[
\mathbb{P}(\delta = 0, s = 2, 2 \text{ edges}) = p^2(1 - p)^{13}\left( \frac{15}{2} - 3\binom{4}{2} \right).
\]

If 3 edges are present in the graph, they must be in the same connected component as argued above. Excluding the configurations with positive deficiency, we have

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

Thus the probability of a random binary reaction network with two species being deficiency zero is

\[
\mathbb{P}(\delta = 0) = (1 - p)^{15} + 15p(1 - p)^{14} + p^2(1 - p)^{13}\left( \frac{15}{2} - 3\binom{4}{2} \right) + p^3(1 - p)^{12}\left( \binom{6}{3} - 3 \right).
\]

\[\square\]

As implied by the two previous examples, the computation of \( P(\delta = 0) \) 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 threshold functions \( r_1(n) \) and \( r_2(n) \) such that

\[
\lim_{n \to \infty} \mathbb{P}(\delta_{G_n} = 0) = \begin{cases} 
0 & \text{if } p_n \gg \frac{1}{n^3}, \\
1 & \text{if } p_n \ll \frac{1}{n^3}.
\end{cases}
\]

In this paper, we are able to obtain a single threshold function \( r(n) = r_1(n) = r_2(n) \).

## 5 The threshold function for deficiency zero

In this section, we will show that the threshold function for deficiency zero is \( r(n) = \frac{1}{n^3} \). In Sections 5.1 and 5.2, we will respectively prove

1. \( \lim_{n \to \infty} \mathbb{P}(\delta_{G_n} = 0) = 0 \) for \( p_n \gg \frac{1}{n^3} \), and
2. \( \lim_{n \to \infty} \mathbb{P}(\delta_{G_n} = 0) = 1 \) for \( p_n \ll \frac{1}{n^3} \).
5.1 The case $p_n \gg \frac{1}{n^7}$

The strategy of this section is similar to the two examples provided in the previous section: we will use the upper bound on the dimension of the stochiometric subspace $s$.

**Lemma 5.1.** Let $n \in \mathbb{N}$ and let $\{S, C, R\}$ be a reaction network with $n$ species. Assume that the reaction network has deficiency zero, then we must have

$$|C| \leq 2n.$$

**Proof.** From the definition of deficiency $\delta = |C| - l - s$, the fact that $s \leq n$, and $l \leq \frac{|C|}{2}$ (from Remark 1), we have

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

Since the reaction network has deficiency zero, we therefore have

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

which implies $|C| \leq 2n$. \qed

We will use this upper bound to show $\lim_{n \to \infty} P(\delta_{G_n} = 0) = 0$ when $p_n \gg \frac{1}{n^7}$. In fact, we will prove a slightly stronger inequality in the theorem below.

**Theorem 5.1.** Suppose $p_n = \frac{2n + \alpha_n}{N_n(N_n - 1)}$ with $\alpha_n \gg n^{1/2}$, then

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

**Proof.** Let $I$ be the set of isolated vertices in $G_n$, that is $I = \{v \in C_0 : \deg(v) = 0\}$. Note that in the reaction network corresponding to $G_n$, the complexes correspond to vertices in $G_n$ with positive degree. Thus, Lemma 5.1 implies that if the network is deficiency zero, we must have

$$|I| = |C_0| - |C| \geq N_n - 2n. \quad (4)$$

From (4), we have

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

To control the right hand side of (5), we use Chebyshev’s inequality.

We consider the expected number of isolated vertices

$$E[|I|] = E \left\{ \sum_{v \in C_0} 1_{\{\deg(v) = 0\}} \right\}$$

$$= N_n P(\deg(v) = 0)$$

$$= N_n (1 - p_n)^{N_n - 1}.$$

Note that

$$|I|^2 = \sum_{v, w \in C_0} 1_{\{\deg(v) = \deg(w) = 0\}} = \sum_{v \in C_0} 1_{\{\deg(v) = 0\}} + \sum_{v, w \in C_0 : v \neq w} 1_{\{\deg(v) = \deg(w) = 0\}}.$$
Therefore, we have
\[
\Var(|I|) = \mathbb{E} \left[ |I|^2 \right] - (\mathbb{E} |I|)^2
\]
\[
= \mathbb{E} \left[ \sum_{v \in \mathcal{G}_0} 1_{\{\deg(v) = 0\}} + \sum_{v, w \in \mathcal{L}_0, v \neq w} 1_{\{\deg(v) = \deg(w) = 0\}} \right] - N_n^2 (1 - p_n)^{2N_n - 2}
\]
\[
= N_n (1 - p_n)^N_n - 1 + N_n (N_n - 1)(1 - p_n)^{2N_n - 3} - N_n^2 (1 - p_n)^{2N_n - 2}
\]
\[
= N_n (1 - p_n)^N_n - 1 (1 - (1 - p_n)^N_n - 2) + N_n^2 (1 - p_n)^{2N_n - 3}p_n
\]
\[
\leq N_n (1 - p_n)^N_n - 1 (N_n - 2)p_n + N_n^2 (1 - p_n)^{2N_n - 3}p_n
\]
\[
\leq N_n (N_n - 2)p_n + N_n^2 p_n \leq 2N_n^2 p_n,
\]
where the first inequality follows from Bernoulli’s inequality.

We will utilize \(\mathbb{E} |I|\) and \(\Var |I|\) to show that
\[
\lim_{n \to \infty} \mathbb{P}(|I| \geq N_n - 2n) = 0.
\]

By taking subsequences if necessary, it suffices to prove (6) in the three cases below.

1. When \(\alpha_n \gg N_n\), we have \(p_n \gg \frac{1}{N_n}\). Applying Markov’s inequality, we have
\[
\mathbb{P}(|I| > N_n - 2n) \leq \frac{\mathbb{E} |I|}{N_n - 2n} = \frac{N_n}{N_n - 2n} (1 - p_n)^{N_n - 1}.
\]

Since \(\lim_{n \to \infty} (1 - p_n)^{N_n - 1} = \lim_{n \to \infty} (1 - p_n)^{\frac{1}{p_n}p_n(N_n - 1)} = \lim_{n \to \infty} e^{-p_n(N_n - 1)} = 0\), we have
\[
\lim_{n \to \infty} \mathbb{P}(|I| > N_n - 2n) = 0.
\]

2. When \(\alpha_n \sim N_n\), we have \(p_n \sim \frac{1}{N_n}\), and thus \(p_n > \frac{c}{N_n}\) for some constant \(c > 0\) and \(n\) large enough. Therefore
\[
\mathbb{E} |I| = N_n (1 - p_n)^{N_n - 1} \leq N_n (1 - \frac{c}{N_n})^{N_n - 1} \leq N_n e^{-c}.
\]

Applying Chebyshev’s inequality yields
\[
\mathbb{P}(|I| > N_n - 2n) \leq \frac{\Var |I|}{(N_n - 2n - \mathbb{E} |I|)^2} \leq \frac{2N_n^2 p_n}{(N_n - 2n - N_n e^{-c})^2} = \frac{2p_n}{(1 - 2n/N_n - e^{-c})^2}.
\]

Since \(p_n \sim \frac{1}{N_n}\) and \(N_n \sim n^2\), we have
\[
\lim_{n \to \infty} \mathbb{P}(|I| > N_n - 2n) = 0.
\]

3. The last case is when \(\alpha_n \ll N_n\), or \(p_n \ll \frac{1}{N_n}\). Using Taylor’s expansion, we have
\[
\mathbb{E} |I| = N_n (1 - p_n)^{N_n - 1} \leq N_n \left(1 - p_n(N_n - 1) + p_n^2 \frac{(N_n - 1)(N_n - 2)}{2}\right).
\]

11
Again, we apply Chebyshev’s inequality:

\[
\mathbb{P}(|I| \geq N_n - 2n) \leq \frac{\text{Var}[|I|]}{(N_n - 2n - E[|I|])^2} \leq \frac{2N_n^2 p_n}{(N_n - 2n - N_n(N_n - 1)p_n - \frac{N_n(N_n - 1)(N_n - 2)}{2} p_n^2)^2} = \frac{2N_n^2 p_n}{(-2n + N_n(N_n - 1)p_n - \frac{N_n(N_n - 1)(N_n - 2)}{2} p_n^2)^2}.
\]

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

\[
\mathbb{P}(|I| \geq N_n - 2n) \leq \frac{\frac{2N_n}{N_n - 1}(2n + \alpha_n)}{\left(-2n + 2n + \alpha_n - \frac{N_n - 2}{2N_n(N_n - 1)}(2n + \alpha_n)^2\right)^2} = \frac{2N_n}{N_n - 1} \left(\frac{2n + \alpha_n}{\alpha_n - \frac{N_n - 2}{2N_n(N_n - 1)}(2n + \alpha_n)^2}\right)^2.
\]

If \( \alpha_n \ll n \) or \( \alpha_n \sim n \), we have

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

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

If \( \alpha_n \gg n \), we have

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

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

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

In all cases above, we have \( \lim_{n \to \infty} \mathbb{P}(|I| \geq N_n - 2n) = 0 \), and from (5),

\[
\lim_{n \to \infty} \mathbb{P}(\delta_{G_n} = 0) = 0.
\]

Note that \( \frac{n}{N_n} \sim \frac{1}{n^2} \). Thus a direct corollary of Theorem 5.1 is the following.

**Corollary 1.** For \( p_n \gg \frac{1}{n^2} \), the following holds

\[
\lim_{n \to \infty} \mathbb{P}(\delta_{G_n} = 0) = 0.
\]
5.2 The case $p_n \ll \frac{1}{n^3}$

The previous section considered when $p_n \gg \frac{1}{n^3}$. Here we focus on the latter case, where $p_n \ll \frac{1}{n^3}$. We will show in Lemma 5.2 that as $n \to \infty$, a random reaction network as described in Section 4 with $p_n \ll \frac{1}{n^3}$ almost surely contains only connected components that consist of 2 vertices. Thus in the corresponding reaction network each linkage class has exactly 2 complexes. For convenience, we introduce a definition to capture this type of reaction network.

**Definition 5.1.** A reaction network is called **paired** if each of its linkage class only contains 2 complexes. A reaction network is called **i-paired** if it is paired and contains $i$ linkage classes.

**Lemma 5.2.** Suppose $p_n \ll \frac{1}{n^3}$. Then

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

**Proof.** We have

$$\mathbb{P}(G_n \text{ is not paired}) = \mathbb{P}(G_n \text{ is not paired}, G_n \text{ contains only trees})$$

$$+ \mathbb{P}(G_n \text{ is not paired}, G_n \text{ contains a cycle}).$$

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

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

Thus it suffices to show

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

We follow the notation in [9] and for $k \geq 2$ let $T_k(n)$ be the number of trees in $G_n$ with $k$ vertices. Using estimates similar to the ones in [9], we have

$$\mathbb{P}(G_n \text{ is not paired, } G_n \text{ contains only trees}) \leq \sum_{k=3}^{N_n} \mathbb{P}(T_k(n) > 0)$$

$$\leq \sum_{k=3}^{N_n} \binom{N_n}{k} k^{k-2} p_n^{k-1}$$

$$\leq \sum_{k=3}^{N_n} \frac{N_n e^k}{\sqrt{2\pi k^3}} \cdot \binom{N_n}{k} k^{k-2} p_n^{k-1}$$

$$= \frac{1}{\sqrt{2\pi}} N_n^3 e^2 p_n^2 \sum_{k=0}^{N_n-3} (N_n e p_n)^k,$$

where the first inequality follows since \{not paired, only trees\} $\subset \bigcup_{k=3}^{N_n} \{T_k(n) > 0\}$, the second follows by choosing the $k$ vertices from the $N_n$ choices and noting there are $k^{k-2}$ possible trees from these vertices (each with $k - 1$ edges), and the third follows from Stirling. Since $p_n \ll \frac{1}{n^3} \sim N_n^{-3/2}$, we have $\lim_{n \to \infty} N_n^3 e^2 p_n^2 = 0$ and $\sum_{k=0}^{N_n-3} (N_n e p_n)^k$ is bounded. Thus we have

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

and the proof is complete.
Remark 3. Note that for \( p_n \ll \frac{1}{n^2} \), the expected number of edges is

\[
p_n \left( \frac{N_n}{2} \right) = p_n \frac{N_n(N_n - 1)}{2} \ll n.
\]

Thus for \( p_n \ll \frac{1}{n^2} \), as \( n \to \infty \), \( G_n \) 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 the source complex and 2 different species in the product complex). 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 \( G_n \) being \( k_n \)-paired and can therefore generate \( G_n \) uniformly from the set of all \( k_n \)-paired graphs.

Lemma 5.3. Suppose that \( k_n \ll n \). Let \( A_n \) be the event that all reactions in \( G_n \) have exactly 4 species. Then we have

\[
\lim_{n \to \infty} \mathbb{P}(A_n|G_n \text{ is } k_n\text{-paired}) = 1.
\]

Proof. Let \( G_n \) be a \( k_n \)-paired reaction network, where \( k_n \ll n \). Denote the \( k_n \) reaction vectors by \( \{v^i_n\}_{k_n=1}^{k_n} \subseteq \mathbb{Z}^n \). We denote by \( A^i_n \) the event that the vector \( v^i_n \) has 4 non-zero elements, thus \( A_n = \bigcap_{i=1}^{k_n} A^i_n \). The proof will proceed by using that

\[
\mathbb{P}(A_n|G_n \text{ is } k_n\text{-paired}) = \prod_{j=0}^{k_n-1} \mathbb{P}(A^i_{n+j+1} \mid \bigcap_{i=1}^j A^i_n, G_n \text{ is } k_n\text{-paired}),
\]

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

First, note that the total number of complexes of the form \( S_k + S_m \) where \( k \neq m \) is \( \left( \begin{array}{c} n \\end{array} \right) \). Suppose we have already picked \( j \) pairs of reversible reactions where each pair has 4 species. Then the number of unpicked complexes of the form \( S_k + S_m \) where \( k \neq m \) is \( \left( \begin{array}{c} n \\end{array} \right) - 2j \). After picking one such \( S_k + S_m \) for the \( j + 1 \)st pair, we need to pick another complex. The number of available complexes of the form \( S_p + S_q \), where \( p, q, m, \) and \( k \) are all different is at least \( \left( \begin{array}{c} n \\end{array} \right) - 2j \), 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 complexes we have already chosen.

Thus for \( n \) large enough, we have

\[
\mathbb{P}(A^i_{n+j+1} \mid \bigcap_{i=1}^j A^i_n, G_n \text{ is } k_n\text{-paired}) \\
\geq \frac{1}{2} \frac{(\begin{array}{c} n \\end{array}) - 2j}{(\begin{array}{c} n \\end{array}) - 2j} \quad \text{(by considering our choices as detailed above)} \\
\geq \frac{1}{2} \frac{(\begin{array}{c} n \\end{array}) - 2n}{(\begin{array}{c} n \\end{array}) - 2n} \quad \text{(since } j \leq n \text{)} \\
= \frac{(n^2 - 5n)(n^2 - 9n + 6)}{(n^2 + 3n + 2)(n^2 + 3n)} \geq \frac{n^2 - 5n)(n^2 - 9n)}{(n^2 + 4n)(n^2 + 3n)} \\
= \frac{n^2 - 14n + 45}{n^2 + 7n + 12} = 1 - \frac{21n - 33}{n^2 + 7n + 12} \\
\geq 1 - \frac{21}{n},
\]
and where the 1/2 in the first term accounts for the symmetry between the selected complexes.

Therefore, for \( n \) large enough, we have

\[
\mathbb{P}(A_n | G_n \text{ is } k_n\text{-paired}) = \prod_{j=0}^{k_n-1} \mathbb{P}(A_{n+1}^j | \cap_{i=1}^j A_i, G_n \text{ is } k_n\text{-paired}) \geq \left( 1 - \frac{21}{n} \right)^{k_n} \geq 1 - \frac{21k_n}{n} \tag{7}
\]

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

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

and the proof is complete. \( \Box \)

**Lemma 5.4.** Suppose that \( k_n \ll n \). Then we have

\[
\lim_{n \to \infty} \mathbb{P}(\delta G_n = 0 | G_n \text{ is } k_n\text{-paired}) = 1. \tag{8}
\]

**Proof.** Let \( G_n \) be a \( k_n\)-paired reaction network, where \( k_n \ll n \). The deficiency of \( G_n \) is given by

\[
\delta_{G_n} = |C| - l - s = 2k_n - k_n - s = k_n - s.
\]

Thus \( \delta_{G_n} = 0 \iff s = k_n \). In other words, \( G_n \) has deficiency zero if and only if all \( k_n \) reaction vectors are linearly independent. Let \( I_n \) be the event that all \( k_n \) reaction vectors are linearly independent.

Similar to Lemma 5.3, denote the \( k_n \) reaction vectors by \( \{v_i\}_{i=1}^{k_n} \in \mathbb{Z}^n \) and denote by \( A_n \) the event that all reactions have exactly 4 species. We have

\[
\mathbb{P}(\delta_{G_n} = 0 | G_n \text{ is } k_n\text{-paired}) = \mathbb{P}(I_n | G_n \text{ is } k_n\text{-paired}) \geq \mathbb{P}(I_n | A_n, G_n \text{ is } k_n\text{-paired}) \mathbb{P}(A_n | G_n \text{ is } k_n\text{-paired}). \tag{9}
\]

Using Lemma 5.3, it suffices to show that

\[
\lim_{n \to \infty} \mathbb{P}(I_n | A_n, G_n \text{ is } k_n\text{-paired}) = 1,
\]

or

\[
\lim_{n \to \infty} \mathbb{P}(I_{n}^c | A_n, G_n \text{ is } k_n\text{-paired}) = 0.
\]

We say a set of vectors is **minimally dependent** if any of its proper subsets are linearly independent. For any set of indices of reaction vectors \( T \subseteq \{1, 2, \ldots, k_n\} \) we denote \( V_n^T = \{ v_i : i \in T \} \). By noting that

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

We have

\[
\mathbb{P}(I_{n}^c | A_n, G_n \text{ is } k_n\text{-paired}) \leq \sum_{\ell=2}^{k_n} \sum_{|T|=\ell} \mathbb{P}(V_n^T \text{ is minimally dependent} | A_n, G_n \text{ is } k_n\text{-paired})
\]

\[
= \sum_{\ell=2}^{k_n} \binom{k_n}{\ell} \mathbb{P}(B_\ell | A_n, G_n \text{ is } k_n\text{-paired}) \tag{10}
\]
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 $T = \{1, 2, \ldots, \ell\}$. Consider a matrix $M_\ell$ whose columns are the vectors in $V_n^T$ conditioned on (i) $A_n$ and (ii) $G_n$ is $k_n$-paired. Not 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

$$
\mathbb{P}(B_\ell|A_n, G_n \text{ is } k_n\text{-paired}) \leq \mathbb{P}(M_\ell \text{ has at least } n - 2\ell \text{ zero rows}|A_n, G_n \text{ is } k_n\text{-paired}). \quad (11)
$$

We denote the row vectors of $M_\ell$ by $\{w^n_i\}_{i=1}^n$. For a subset of indices of species $R \subseteq \{1, 2, \ldots, n\}$ we denote $W^R_n = \{w^n_i : i \in R\}$. We say that $W^R_n = 0$ if all the vectors in the set are the zero vector.

We have

$$
\mathbb{P}(M_\ell \text{ has at least } n - 2\ell \text{ zero rows}|A_n, G_n \text{ is } k_n\text{-paired}) \leq \sum_{|R| = n-2\ell} \mathbb{P}(W^R_n = 0|A_n, G_n \text{ is } k_n\text{-paired}) \quad (12)
$$

$$
= \binom{n}{n - 2\ell} \mathbb{P}(C_\ell|A_n, G_n \text{ is } k_n\text{-paired})
$$

where $C_\ell$ is the event that $W^R_n = 0$ for a particular $R$ satisfying $|R| = n - 2\ell$.

Now fix a set $R$ with $|R| = n - 2\ell$. Without loss of generality, let $R = \{2\ell + 1, \ldots, n\}$. Then the event $C_\ell$ involves picking $\ell$ column vectors: $V_n^T = \{v_1, \ldots, v_\ell\}$ where the last $n - 2\ell$ elements of each column vector are zero. Recall that conditioned on $A_n$, each vector has two elements being 1 and two elements being $-1$. Suppose we have already picked $j$ such column vectors. Conditioned on $A_n$ and $G_n$ is $k_n$-paired, the number of ways we can pick the $j + 1$-st vector is at least $(\binom{n}{2} - 2j)(\binom{n-2}{2} - 2j)$ (this follows from the same argument as in the proof of Lemma 5.3).

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

$$
\mathbb{P}(C_\ell|A_n, G_n \text{ is } k_n\text{-paired}) \leq \prod_{j=0}^{\ell-1} \frac{\binom{2\ell}{2} \binom{2\ell - 2}{2}}{(\binom{n}{2} - 2j)(\binom{n-2}{2} - 2j)} \leq \prod_{j=0}^{\ell-1} \frac{\binom{2\ell}{2} \binom{2\ell - 2}{2}}{\frac{1}{4} \binom{n}{2} \binom{n-2}{2}} \leq 4 \left(\frac{2\ell}{n}\right)^4.
$$

Plugging the above into (12), we see

$$
\mathbb{P}(M_\ell \text{ has at least } n - 2\ell \text{ zero rows}|A_n, G_n \text{ is } k_n\text{-paired}) \leq \binom{n}{n - 2\ell} 4 \left(\frac{2\ell}{n}\right)^{4\ell}
$$

$$
\leq \frac{n^{2\ell}}{(2\ell)!} 4 \left(\frac{2\ell}{n}\right)^{4\ell}
$$

$$
\leq \frac{4n^{2\ell}}{\sqrt{2\pi}(2\ell/e)^{2\ell}} \left(\frac{2\ell}{n}\right)^{4\ell} = \frac{4}{\sqrt{2\pi}} \left(\frac{2\ell e}{n}\right)^{2\ell}.
$$

16
Now combining (10), (11) and (13), we have

\[
P(T_n^c|A_n, G_n \text{ is } k_n\text{-paired}) \leq \sum_{\ell=2}^{k_n} \binom{k_n}{\ell} \frac{4}{\sqrt{2\pi}} \left( \frac{2\ell e}{n} \right)^{2\ell} \\
\leq \sum_{\ell=2}^{k_n} \frac{k_n^\ell}{\ell!} \frac{4}{\sqrt{2\pi}} \left( \frac{2\ell e}{n} \right)^{2\ell} \\
\leq \sum_{\ell=2}^{k_n} \frac{k_n^\ell}{\sqrt{2\pi}(\ell/e)^\ell} \frac{4}{\sqrt{2\pi}} \left( \frac{2\ell e}{n} \right)^{2\ell} \\
= \sum_{\ell=2}^{k_n} \frac{2}{\pi} \left( \frac{4e^3k_n}{n^2} \right)^\ell \\
\leq \sum_{\ell=2}^{\infty} \frac{2}{\pi} \left( \frac{4e^3k_n}{n^2} \right)^\ell \\
\leq \frac{k_n^4}{n^4}. \tag{14}
\]

for some constant \( c > 0 \), since \( k_n \ll n \). Thus using (7), (9) and (14), we must have

\[
P(\delta G_n = 0| G_n \text{ is } k_n\text{-paired}) \geq \left( 1 - \frac{k_n^4}{n^4} \right) \left( 1 - \frac{21k_n}{n} \right), \tag{15}
\]

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

Combining the three Lemmas above, we are ready to state the main theorem for this section.

**Theorem 5.2.** Suppose \( p_n \ll \frac{1}{n^3} \), then

\[
\lim_{n \to \infty} P(\delta G_n = 0) = 1.
\]

**Proof.** We have

\[
P(\delta G_n = 0) = P(\delta G_n = 0, G_n \text{ is paired}) + P(\delta G_n = 0, G_n \text{ is not paired}).
\]

Since

\[
P(\delta G_n = 0, G_n \text{ is not paired}) \leq P(G_n \text{ is not paired}),
\]

we must have

\[
\lim_{n \to \infty} P(\delta G_n = 0, G_n \text{ is not paired}) = 0
\]

due to Lemma 5.2. Therefore it suffices to show

\[
\lim_{n \to \infty} P(\delta G_n = 0, G_n \text{ is paired}) = 1.
\]

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

\[
\mathbb{P}(\delta_{G_n} = 0, G_n \text{ is paired}) = \sum_{i=1}^{n} \mathbb{P}(\delta_{G_n} = 0, G_n \text{ is } i\text{-paired})
\]

\[
= \sum_{i=1}^{n} \mathbb{P}(\delta_{G_n} = 0|G_n \text{ is } i\text{-paired})\mathbb{P}(G_n \text{ is } i\text{-paired})
\]

\[
= \sum_{i=1}^{n} \mathbb{P}(\delta_{G_n} = 0|G_n \text{ is } i\text{-paired}) \frac{N_n!}{i!2^i(N_n-2i)!} p_n^i (1 - p_n)^{N_n(N_n-1)/2-i}
\]

\[
\geq \sum_{i=1}^{n} \mathbb{P}(\delta_{G_n} = 0|G_n \text{ is } i\text{-paired}) \frac{(N_n-2i)^2}{i!2^i} p_n^i (1 - p_n)^{N_n(N_n-1)/2-i}
\]

(16)

where the third equality uses that the number of \( i\)-paired graphs is \((N_n/2)(N_n-2)\ldots(N_n-2i+2)/2\), with the repetition of the graphs accounted for by division by \( i! \).

Let \( k_n \) satisfy \( \lim_{n \to \infty} k_n = \infty \), \( k_n \ll n \), and \( k_n \gg N_n^2 p_n \). This is possible because \( p_n \ll 1/n^3 \) and \( N_n \sim n^2 \). Cutting off the last \( n - k_n \) terms from (16), yields

\[
\mathbb{P}(\delta_{G_n} = 0, G_n \text{ is paired}) \geq \sum_{i=1}^{k_n} \mathbb{P}(\delta_{G_n} = 0|G_n \text{ is } i\text{-paired}) \frac{(N_n-2i)^2}{i!2^i} p_n^i (1 - p_n)^{N_n(N_n-1)/2-i}
\]

\[
\geq \left(1 - c\frac{k_n^4}{n^4}\right) \left(1 - \frac{21k_n}{n}\right) \frac{(N_n-2i)^2}{i!2^i} p_n^i (1 - p_n)^{N_n(N_n-1)/2-i}
\]

\[
\geq \left(1 - c\frac{k_n^4}{n^4}\right) \left(1 - \frac{21k_n}{n}\right) \frac{(N_n-2k_n)^2}{i!2^i} p_n^i.
\]

where the second inequality is obtained from (15) in Lemma 5.4.

Let \( \lambda_n = \frac{(N_n-2k_n)^2}{2} p_n \), 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

\[
\sum_{i=1}^{k_n} \frac{\lambda_n^i}{i!} \geq e^{\lambda_n} - e^{\lambda_n} \frac{\lambda_n^{k_n+1}}{(k_n+1)!} \geq e^{\lambda_n} \left(1 - \frac{\lambda_n^{k_n+1}}{\sqrt{2\pi}(k_n+1)k_n+1e^{-k_n+1}}\right) = e^{\lambda_n} \left(1 - \frac{1}{\sqrt{2\pi}} \left(\frac{\lambda_n e}{k_n+1}\right)^{k_n+1}\right).
\]

Thus we have

\[
\mathbb{P}(\delta_{G_n} = 0, G_n \text{ is paired}) \geq \left(1 - c\frac{k_n^4}{n^4}\right) \left(1 - \frac{21k_n}{n}\right) \left(1 - p_n\right)^{N_n^2/2} e^{\lambda_n} \left(1 - \frac{1}{\sqrt{2\pi}} \left(\frac{\lambda_n e}{k_n+1}\right)^{k_n+1}\right).
\]

Since \( \lambda_n \ll k_n \ll n \), the first, second, and last terms converge to one. Hence, it suffices to show

\[
\lim_{n \to \infty} (1 - p_n)^{N_n^2/2} e^{\lambda_n} = 1,
\]

or

\[
\lim_{n \to \infty} \frac{N_n^2}{2} \ln(1 - p_n) + \lambda_n = 0.
\]
Since $p_n \ll 1$, we have $-p_n - p_n^2 \leq \ln(1 - p_n) \leq -p_n$. Thus
\[
\frac{N_n^2}{2} \ln(1 - p_n) + \lambda_n \leq -\frac{N_n^2}{2} p_n + \lambda_n = \frac{p_n}{2}((N_n - 2k_n)^2 - N_n^2) = \frac{p_n}{2}(-4k_n N_n + 4k_n^2).
\]
On the other hand, and using the equality above,
\[
\frac{N_n^2}{2} \ln(1 - p_n) + \lambda_n \geq -\frac{N_n^2}{2} (p_n + p_n^2) + \lambda_n = \frac{p_n}{2}(-4k_n N_n + 4k_n^2) - \frac{N_n^2 p_n^2}{2}.
\]
Since $k_n \ll n$, $N_n \sim n^2$ and $p_n \ll \frac{1}{n^2}$, we have
\[
\lim_{n \to \infty} \frac{p_n}{2}(-4k_n N_n + 4k_n^2) = 0, \quad \text{and,} \quad \lim_{n \to \infty} \frac{N_n^2 p_n^2}{2} = 0.
\]
Thus
\[
\lim_{n \to \infty} \frac{N_n^2}{2} \ln(1 - p_n) + \lambda_n = 0,
\]
which concludes the proof of the Theorem.

6 Discussion

We view the current work not only as an attempt to make a connection between reaction network theory and network science and network biology, but also as an opening to a potentially deep line of subsequent inquiry, as there are numerous directions we can expand from here. As mentioned in the introduction, instead of the Erdős-Rényi framework, we can study models where additional structure, or certain prior information is known. For example, in open networks where inflows and outflows are likely or even guaranteed, we can consider a weighted Erdős-Rényi framework. In particular, we can put higher edge probabilities on inflow and outflow reactions ($\emptyset \leftrightarrow S_i$). Another example is from the setting of molecular biology, where some proteins may be more active and interact with many other proteins while some proteins may be 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 or complexes can be assigned different weights [11]. From another direction, we can utilize a directed version of the Erdős-Rényi framework to study weak reversibility in addition to deficiency zero.

Another interesting direction is to study the probability of a “sub-network” being deficiency zero. This may be relevant in control theory, where a group of species may be chemostated and the original network is reduced to a smaller “sub-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 according to the stochastic model [4].

Finally, 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 [21], endotactic, strongly endotactic, and asyphonic as in [1, 2, 17].

Acknowledgements

We thank Robin Pemantle for influencing the formulation of the question posed in this paper. We gratefully acknowledge grant support from the Army Research Office via grant W911NF-18-1-0324.
References


