Exploring Lower Bounds of Self-Assembled DNA Complexes

J. Miller and A. Lavengood

Abstract - DNA self-assembly is a tool that allows for the creation of various nanostructures. We model these nanostructures within the flexible-tile model that omits geometric restrictions on the DNA-stranded arms. To create structures, we use a collection of tiles known as a pot, which we restrict to explore the cases in which the pot realizes a graph of each order greater than some minimum order. We provide an algorithm that calculates the lower bounds and tile distributions of these restricted pots, which allows for discovering important relationships. By utilizing graph and number theoretic tools, we show that a closed formula exists for the minimum order of particular pots.

Keywords : DNA self-assembly; graph theory; number theory; algorithms

Mathematics Subject Classification (2020) **:** 05C90

1 Introduction

DNA has become a focal point in lab experimentation due to its self-assembling nature. Originally introduced in the 1980s in Seeman's laboratories, this DNA-based bottom up assembly used complementary nucleotides to create new genetic structures of all sorts of shapes [1]. The nature of the nucleotide pairing makes it easy to be thought of in graph theoretic terms. We can take a target structure to be a graph, where each vertex is a branched junction molecule that contains base pairings that complement another branched junction molecule to create bonded edges. Due to the direct ties, we can apply known graph theoretic topics to answer questions regarding DNA self-assembly

We refer to branched junction molecules as tiles that are contained within a pot. Due to the behavior of self-assembly, natural questions that may arise are what sort of DNA complexes can be constructed and what are the relationships between a constructed complex and it's corresponding pot? One area of interest that has been researched is determining an optimal pot given a target structure. Our focus is on the inverse problem, expanding on the work done by Johnson and Lavengood-Ryan in [9]. That is, what structures can we create if we are given a specific pot?

When dealing with the inverse problem, we arrive at a crossroads with certain pots having more constraints and a more predictable, well-defined set of structures able to be created versus other pots that have less constraints and a wider range of structures able to be created. The former is discussed in [9]. We narrow our focus to look at the latter

- the pots that are less predictable, but give greater flexibility with the structures they create. We explore one main question:

Given a pot, is there a closed formula for the minimum order of a constructed graph such that a graph with any greater order can be constructed?

We first define and state some important preliminary knowledge in number theory, graph theory, and DNA complexes. We then move to exploring the lower bounds of graph orders in which all greater graph orders are realized. Finally, we establish a conjecture on the first lower bound and the conditions in which it holds.

2 Preliminaries

In this section we present some standard definitions and results from graph theory and number theory, as well as the introductory definitions to DNA complexes.

2.1 Number Theory Preliminaries

Definition 2.1 *Let $a, b \in \mathbb{Z}$, with at least one of a, b not equal to 0. The greatest common divisor of a and b, denoted $\gcd(a,b)$, is the positive integer d satisfying the following:*

- *$d \mid a$ and $d \mid b$.*
- *If $c \mid a$ and $c \mid b$, then $c \leq d$.*

We note the following properties of the greatest common divisor [6]:

1. $\gcd(a,b) = \gcd(b,a)$.
2. $\gcd(a,b) = \gcd(-a,b) = \gcd(a,-b) = \gcd(-a,-b)$.
3. $\gcd(a,b) = \gcd(a, b+an)$ *for all* $n \in \mathbb{Z}$.
4. $\gcd(a, a+1) = 1$.

Let a, b, c be integers $(a, b \neq 0)$. The equation $ax + by = c$ is called a **linear Diophantine equation**. A solution of this equation is of the form $\{(x,y) \mid x, y \in \mathbb{Z}\}$ where (x,y) satisfies the equation. A **homogeneous** linear Diophantine equation is in the form $ax + by = 0$. An important result of homogeneous linear Diophantine equations is given below:

Lemma 2.2 *If $\gcd(a,b) = d$, then the family of solutions to the equation $ax + by = 0$ is $x = \frac{b}{d}k$ and $y = -\frac{a}{d}k$ for $k \in \mathbb{Z}$. [7]*

The **ceiling** function of a real number is the least integer greater than or equal to the given number. We denote the ceiling function as $\lceil \cdot \rceil$.

2.2 Graph Theory Preliminaries

A **graph** G consists of a finite, nonempty set V of objects called **vertices** and a set E of 2-element subsets of V called **edges**. We say G is **connected** if G contains a $u - v$ path for every pair u, v of vertices of G. We call G **disconnected** if it is not connected. A graph H is called a **subgraph** of a graph G, written $H \subseteq G$, if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. A **digraph** (or **directed graph**) D is a finite, nonempty set V of vertices together with a set E of *ordered pairs* of distinct vertices. The elements of E are called **directed edges** or **arcs** [8].

Definition 2.3 *We define the star graph S_k to be a tree on k nodes with one node having vertex degree $k - 1$ and the other $n - 1$ having vertex degree 1. [10]*

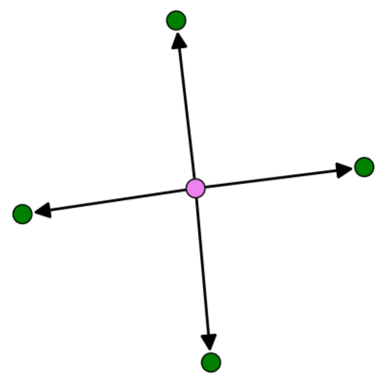

Figure 1: Example of an S_5 graph.

2.3 Preliminaries on DNA Complexes

DNA self-assembly exploits Watson–Crick base pairing to form complex DNA structures, with one of the earliest processes utilizing branched junction molecules as the primary building blocks of the process [4]. A branched junction molecule is an asterisk-shaped molecule with arms that have adhesive ends, allowing for multiple branched junction molecules to combine and form larger structures [5]. This self-assembly process can be translated into graph-theoretic language (see Definition 2.4) and can therefore be modeled using graphs. Previous work has demonstrated how such methods can be used to construct DNA complexes in the form of cubes [3] and octahedrons [2], yielding important biological implications. Other work has been done to understand how efficient the methods are in creating DNA complexes. Several different scenarios exist when determining how to create DNA complexes, however, so the algorithms and optimal solutions often vary depending on the constraints we may apply [4].

Ellis-Monaghan [4] discusses three separate scenarios regarding DNA self-assembly. The first scenario looks at the possibility of creating graphs of smaller size than that of the target graph given a specific set of *tiles* (see Definition 2.4). Scenario two focuses on only allowing the creation of graphs with the same number of vertices as, but not isomorphic to, the target graph. Our main question relates to scenario three: the case in which we only allow complexes isomorphic to the target graph to be created. In scenario 3, the authors study graphs such as complete graphs and trees; analyzing their structure to determine conditions for unique assembly. Throughout the discussion of the scenarios, several proofs and algorithms are demonstrated to showcase how the various restrictions can change the way we must think about the assembly process. An important observation to be made from these scenarios is that our solution sets are highly dependent on the constraints we may apply. As we increase restrictions, we successfully remove extraneous structures, however, require more specific prerequisites which may be more expensive or inaccessible.

Definitions 2.4 and 2.5 are obtained from [4] and [9].

Definition 2.4 *Consider a k-armed branched junction molecule.*

1. *A k-armed branched junction molecule is modeled by a tile. A tile is a vertex with k half-edges representing the cohesive-ends (or arms) of the molecule $a, b, c, \ldots$. We will denote complementary cohesive-ends with $\hat{a}, \hat{b}, \hat{c}, \ldots$.*

2. *A bond-edge type is a classification of the cohesive-ends of tiles (without regard to hatted and unhatted letters). For example, a and $\hat{a}$ will bond to form bond-edge type a.*

3. *We denote tiles by t_j, where $t_j = \{a_1^{e_1}, \hat{a}_1^{e_2}, \ldots, a_k^{e_{2k-1}}, \hat{a}_k^{e_{2k}}, \ldots\}$. The exponent on a_i indicates the quantity of cohesive-ends of type a_i present on the tile. We denote the number of tiles of type j used to realize a graph with R_j.*

4. *A pot is a collection of tiles such that for any cohesive-end type that appears on any tile in the pot, its complement also appears on some tile in the pot. We denote a pot by P.*

5. *It is our convention to think of bonded arms (that is, where an a_i has been matched with an $\hat{a}_i$) as edges on a graph, and we think of the bond-edge type as providing direction and compatibility of edges. Unhatted cohesive-ends will denote half-edges directed away from the vertex, and hatted cohesive-ends will denote a half-edge directed toward the vertex. When cohesive-ends are matched, this will result in a directed edge pointing away from the tile that had an unhatted cohesive end and toward the vertex that had a hatted cohesive end.*

6. *We say that a graph G is realized by a pot P if some collection of tiles $t_i \in P$ can be self-assembled to form G without any unpaired half-edges, and we write $G \in \mathcal{O}(P)$, the set of all graphs realized by P.*

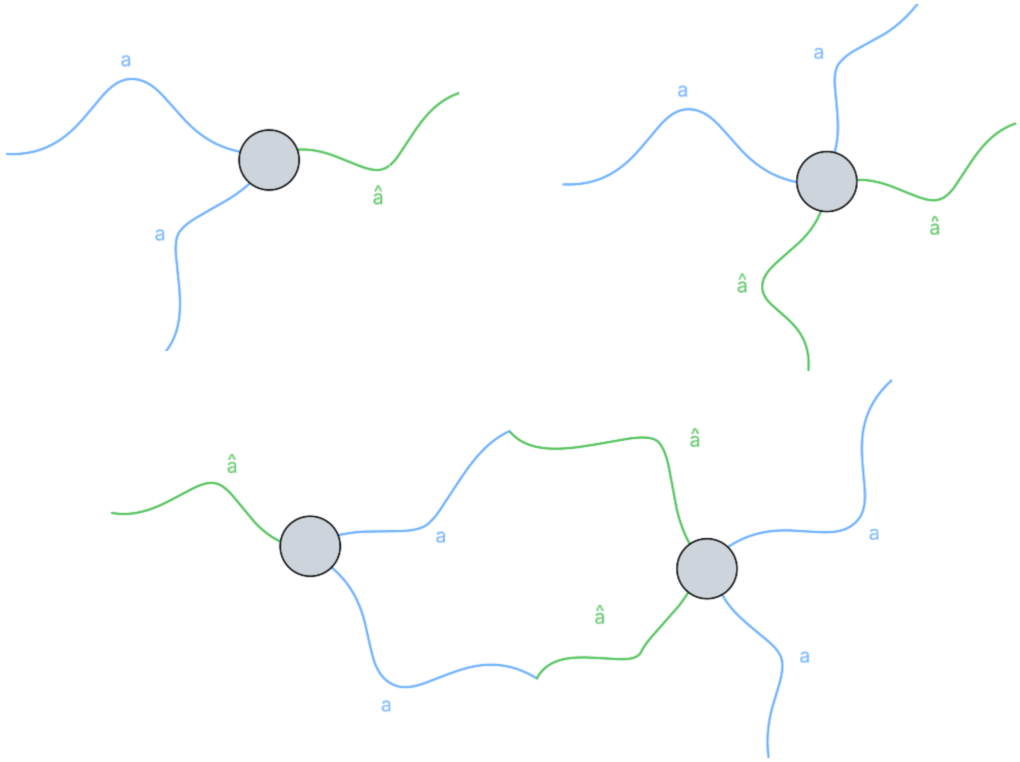

Figure 2: Example of two flexible tiles and how they can form a branched junction molecule.

As discussed in [4], we require all of our assembled complexes to be complete. Let $P = \{t_1, \ldots, t_2\}$ be a pot. Since we require our complexes to be complete, we know that the total number of unhatted cohesive ends must equal the total number of hatted cohesive ends. Further, each constructed graph must have a nonnegative integer, R_i, representing the number of each tile type t_i used in the construction.

Definition 2.5 *Let P be a pot with tiles t_i for $i \in \{1, \ldots, n-1\}$. Then the construction matrix of P is given by*

$$M_P = \begin{array}{c} \\ a_1 \\ a_2 \\ \vdots \\ a_m \\ \\ \end{array} \begin{array}{c} \begin{array}{cccc} t_1 & t_2 & \cdots & t_{n-1} \end{array} \\ \left[\begin{array}{cccc|c} z_{1,1} & z_{1,2} & \cdots & z_{1,n-1} & 0 \\ z_{2,1} & z_{2,2} & \cdots & z_{2,n-1} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ z_{m,1} & z_{m,2} & \cdots & z_{m,n-1} & 0 \\ 1 & 1 & \cdots & 1 & 1 \end{array}\right] \end{array},$$

where $z_{i,j} = A_{i,j} - \hat{A}_{i,j}$ such that $A_{i,j}$ is the number of cohesive ends of type a_i on tile t_j and $\hat{A}_{i,j}$ is the number of cohesive ends of type $\hat{a}_i$ on tile t_j. The solution space of M_P

is called the ***spectrum*** *of* P*, and is denoted by* $\mathcal{S}(P)$*. We denote the minimal element within* $\mathcal{S}(P)$ *as* m_p*.*

We reserve P for the pot $P = \{t_1 = \{a^{e_1}\}, t_2 = \{\hat{a}^{e_2}\}, t_3 = \{\hat{a}\}\}$, where $e_1, e_2 \geq 0$, and n for the order of a graph.

3 Exploring the Lower Bounds of S_P

In this section, we discuss important theorems derived from [9] and further explore the lower bounds of S_P. We restrict our pot due to how quickly DNA complexes escalate in computational and theoretical difficulty.

We let $G_P = \{G \mid G \in \mathcal{O}(P)\}$ and consider $S_P = \{n \mid n + x$ is the order of some $G \in G_P$ for every $x \in \mathbb{N}\}$. The **lower density bound** of P is $\zeta = \min(S_P)$ [9].

It is important to note that the construction matrix of P plays a crucial role for the logic in our code to manually compute ζ values (see pseudocode below). The matrix is given by

$$M_P = \left[\begin{array}{ccc|c} e_1 & -e_2 & -1 & 0 \\ 1 & 1 & 1 & 1 \end{array}\right],$$

which gives us the system of equations

$$\begin{cases} e_1x - e_2y - z = 0, \\ x + y + z = n, \end{cases}$$

which we add together to get

$$(e_1 + 1)x + (-e_2 + 1)y = n.$$

We note that from here on out, we will always consider the case such that $\gcd(e_1+1, -e_2+1) = 1$ and that $e_1 > e_2$. The case where $\gcd(e_1 + 1, -e_2 + 1) = d \neq 1$ does not result in the existence of a lower density bound so it is of no interest to us, but a full treatment can be found in [9].

The values of ζ for a given pot prove difficult to predict. Alternatively, we can find a lower bound that is finitely greater than ζ, allowing us to use computational methods to determine the value of ζ.

Theorem 3.1 *[9] Consider the pot* P *where* $gcd(e_1 + 1, -e_2 + 1) = 1$*. Then* P *realizes a graph for every* n *with* $n \geq max\left\{\frac{(e_1+1)(e_1+e_2)}{e_1}, \frac{(e_2-1)(e_1+e_2)}{e_2}\right\}$*.*

As discussed in [9], we will denote

$$\eta = \max\left\{\frac{(e_1 + 1)(e_1 + e_2)}{e_1}, \frac{(e_2 - 1)(e_1 + e_2)}{e_2}\right\}.$$

Our main area of interest is in further exploring η, as well as how it relates to ζ, since $\eta \geq \zeta$. To gain a better understanding of ζ, we also wanted to explore how ζ compared to other variables, such as e_1, e_2, and the tile distribution of P (that is, R_1, R_2, R_3). To explore the relationship of all these variables, we created a Sage program to manually compute the $\eta, \zeta, R_1, R_2, R_3$ calculations for a given e_1 and e_2. A pseudocode snippet of some of the logic used to generate a list of ζ values, given a threshold, is included in Section 5.

Table 1 is a sample of output data received after iterating for several values of e_1 and e_2, using an upper bound of $e_2 \leq 100$:

e_1	e_2	$\lceil\eta\rceil$	η	ζ	R_1	R_2	R_3
3	2	7	20/3	3	1	1	1
4	2	8	15/2	3	1	2	0
4	3	9	35/4	5	1	0	4
6	3	11	21/2	5	1	1	3
6	4	12	35/3	7	1	0	6
7	4	13	88/7	7	2	3	2
6	5	13	77/6	9	3	3	3
8	5	15	117/8	9	1	0	8

Table 1: Sample data for specific e_2 values from our algorithm.

The data table we constructed allowed us to make valuable observations on the values of η and certain values of ζ. The values of η will be discussed in Theorem 3.3. We classify these certain values of ζ with our main definition.

Definition 3.2 *Let P_{e_2} denote a family of pots with a fixed value for e_2, and let $P_1 \in P_{e_2}$ denote the pot with the minimum e_1 value such that (1) $e_1 > e_2$ and (2) P_1 has a lower density bound. We denote the lower density bound of P_1 by ζ_1.*

Our objective is to determine the value of ζ_1 for any given e_2. This is a long journey that begins with refining the value of η.

Since n is always a natural number, we know that $n \geq \eta$ is the same as $n \geq \lceil\eta\rceil$. Therefore we can adjust Theorem 3.1 to be stated as:

Theorem 3.3 *Consider the pot P where $gcd(e_1 + 1, -e_2 + 1) = 1$. Then P realizes a graph for every order n with $n \geq (e_1 + 1) + (e_2 + 1)$.*

Proof. It is sufficient to show $\lceil\eta\rceil = (e_1 + 1) + (e_2 + 1)$. Recall that

$$\eta = \max\left\{\frac{(e_1+1)(e_1+e_2)}{e_1}, \frac{(e_2-1)(e_1+e_2)}{e_2}\right\}$$

We begin by showing that

$$\frac{(e_1+1)(e_1+e_2)}{e_1} > \frac{(e_2-1)(e_1+e_2)}{e_2}$$

for all $e_1 > e_2$.

We know that $0 < \frac{e_2}{e_1} < 1$, so we know $\frac{e_2}{e_1} > -1$. We establish the desired inequality through a series of algebraic manipulations stemming from this fact:

$$\begin{aligned} \frac{e_2}{e_1} &> -1 \\ \implies e_2 + \frac{e_2}{e_1} &> e_2 - 1 \\ \implies \frac{e_2}{e_1}(e_1+1) &> e_2 - 1 \\ \implies \frac{e_1+1}{e_1} &> \frac{e_2-1}{e_2} \\ \implies \frac{(e_1+1)(e_1+e_2)}{e_1} &> \frac{(e_2-1)(e_1+e_2)}{e_2} \end{aligned}$$

Therefore, we know that $\eta = \frac{(e_1+1)(e_1+e_2)}{e_1}$.

Now we show that $\lceil \eta \rceil = (e_1+1) + (e_2+1)$:

$$\eta = \frac{(e_1+1)(e_1+e_2)}{e_1} = \frac{e_1^2 + e_1e_2 + e_1 + e_2}{e_1} = e_1 + e_2 + 1 + \frac{e_2}{e_1}.$$

Since e_1, e_2 are integers, we know $e_1 + e_2 + 1$ is an integer and that $0 < \frac{e_2}{e_1} < 1$, therefore $\lceil \eta \rceil = \lceil (e_1 + e_2 + 1 + \frac{e_2}{e_1} \rceil = e_1 + e_2 + 1 + 1 = (e_1+1) + (e_2+1)$. □

This theorem gives us a simpler version of η which will always be a natural number. While the importance of this is shown later, our main interest is in finding a closed-form of ζ. After analyzing output, such as from Table 1, we created Conjecture 3.4 to begin answering this question.

Conjecture 3.4 For any $e_2 \geq 2$, $\zeta_1 = 2e_2 - 1$.

In order to prove Conjecture 3.4, we aim to prove the following three statements:

1. A graph is realized of order ζ_1.

2. P realizes a graph for every order n such that such that $\zeta_1 < n < \eta$.

3. A graph of order $n = \zeta_1 - 1$ is not realized.

We split our hypothesis into two cases: when $e_2 = 2$ and when $e_2 > 2$. The case when $e_2 = 2$ will be brute forced and the case when $e_2 > 2$ will utilize several lemmas and theorems.

We begin by proving that a graph of order $2e_2 - 1$ is realized by P. We first establish a property of the greatest common divisor.

Lemma 3.5 $\gcd(2e_2 - 1, -e_2 + 1) = 1$

Proof. We will prove that $\gcd(2e_2 - 1, -e_2 + 1) = 1$. Using the gcd properties in Definition 2.1, we have:

$$\begin{aligned}
\gcd(2e_2 - 1, -e_2 + 1) &= \gcd(-e_2 + 1, 2e_2 - 1) && \text{Property 1}\\
&= \gcd(-e_2 + 1, (2e_2 - 1) + (-e_2 + 1)) && \text{Property 3}\\
&= \gcd(-e_2 + 1, e_2)\\
&= \gcd(e_2 - 1, e_2) && \text{Property 2}
\end{aligned}$$

By property 4 of Definition 2.1, we know that $\gcd(2e_2 - 1, -e_2 + 1) = \gcd(e_2 - 1, e_2) = 1$. □

With Lemma 3.5 established, we proceed by showing a graph of order $2e_2 - 1$ can be realized.

Theorem 3.6 *If $e_2 > 2$, a star graph of order $2e_2 - 1$ can be realized by P.*

Proof. Assume that $e_2 > 2$. We want to show that a star graph can always be realized of order $2e_2 - 1$. From Definition 2.3, we know that a star graph is a tree that has one central vertex. Consider the tile distribution $R_1 = 1, R_2 = 0$ and $R_3 = e_1$, noting that R_1 would be our central vertex and R_3 would be our leaves. Note that $2e_2 - 2 = 2(e_2 - 1) > e_2$ because $e_2 > 2$. By Lemma 3.5, we know that $e_1 = 2e_2 - 2$ satisfies the criterion for the existence of ζ. We let the order of our graph be n, so

$$\begin{aligned}
&R_1 + R_2 + R_3 = n\\
\implies &1 + 0 + e_1 = n\\
\implies &1 + 2e_2 - 2 = n\\
\implies &2e_2 - 1 = n.
\end{aligned}$$

Therefore we have realized a star graph that has order $2e_2 - 1$ for an arbitrary $e_2 > 2$. □

Figure 3 is an example of a star graph we can create:

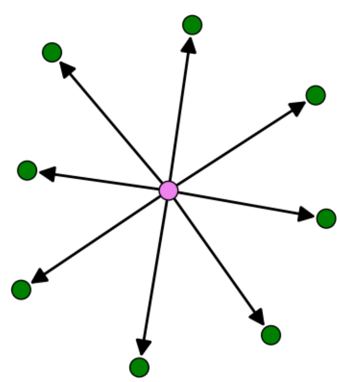

Figure 3: Star graph of order 9 for $P = \{\{a^8\}, \{\hat{a}^5\}, \{\hat{a}\}\}$.

An important observation from Theorem 3.6 is that $e_1 = 2e_2 - 2$ satisfies our greatest common divisor requirement for the existence of a lower density bound, which we will utilize to prove that every order n such that $2e_2 - 1 < n < \eta$ is realized by a graph.

Corollary 3.7 *For a pot that realizes a graph of order $2e_2 - 1$ and has $e_1 = 2e_2 - 2$, the difference between $2e_2 - 1$ and η is $e_2 + 1$.*

Proof. By Theorem 3.6, we know $e_1 = 2e_2 - 2$ is a value of e_1 that will realize a graph of order $2e_2 - 1$. By Theorem 3.3, we know that $\eta = e_1 + e_2 + 2$. We consider the difference $\eta - (2e_2 - 1)$:

$$\eta - (2e_2 - 1) = (e_1 + e_2 + 2) - (2e_2 - 1) = e_2 + 1.$$

□

This new notion of difference between our η and ζ values will allow us to utilize proof by cases when proving condition 2 of Conjecture 3.2. We prove this condition in Lemma 3.8.

Lemma 3.8 *If a pot P realizes a graph of order $2e_2 - 1$ with $e_2 > 2$ and $e_1 = 2e_2 - 2$, then every order n such that $2e_2 - 1 < n < \eta$ is realized by some $G \in \mathcal{O}(P)$.*

Proof. By corollary 3.7, we know that the difference between $2e_2 - 1$ and η is $e_2 + 1$. Therefore, the orders we must check are $2e_2 - 1 + x$ for $x \in [1, e_2]$, noting that adding $e_2 + 1$ to $2e_2 - 1$ gives η which we already know can be realized. We proceed by splitting this proof into two cases: (1) $x \in [1, e_2 - 1]$, and (2) $x = e_2$.

We create a system of equations that generates the tile proportions for each order $2e_2 - 1 + x$, where $x \in [1, e_2 - 1]$.

$$\begin{cases} R_{1x} &= 1 + x \\ R_{2x} &= 2x \\ R_{3x} &= 2e_2 - 2x - 2 \end{cases}$$

We know that $R_{1x}, R_{2x}, R_{3x} \geq 0$. For this tile distribution to realize a graph, we know that the net number of cohesive-end types must be 0, so we must satisfy the following equation:

$$R_1 e_1 = R_2 e_2 + R_3 \tag{1}$$

We note that we are using the same pot to realize graphs of all orders, therefore we hold e_1 and e_2 to be constant. We also note that we are considering a pot that realizes a graph of order $2e_2 - 1$, therefore by the proof of Theorem 3.6, we know $e_1 = 2e_2 - 2$ is a valid e_1. Substituting into Equation 1 we get:

$$\begin{aligned} R_1 e_1 &= R_2 e_2 + R_3, \\ (1 + x)e_1 &= 2xe_2 + 2e_2 - 2x - 2, \\ e_1 + e_1 x &= 2xe_2 + 2e_2 - 2x - 2, \\ 2e_2 - 2 + 2e_2 x - 2x &= 2xe_2 + 2e_2 - 2x - 2, \\ 2e_2 - 2 &= 2e_2 - 2. \end{aligned}$$

Equation 1 holds so we know that graphs are realized for orders $2e_2-1+x$ for $x \in [1, e_2-1]$.

We now need to show that the graph of order $2e_2 - 1 + e_2$ can be realized. Consider the tile distribution $R_1 = 2, R_2 = 1$ and $R_3 = 2e_2 - 1 + e_2 - 3 = 3e_2 - 4$. Since $e_2 > 2$, we know that $R_1, R_2, R_3 > 0$. Once again, we must show that Equation 1 holds:

$$\begin{aligned} R_1e_1 &= R_2e_2 + R_3, \\ 2e_1 &= e_2 + 3e_2 - 4, \\ 2(2e_2 - 2) &= 4e_2 - 4, \\ 4e_2 - 4 &= 4e_2 - 4. \end{aligned}$$

The equation holds, so we know this tile distribution realizes a graph of order $2e_2 - 1 + e_2$. Therefore, every graph of order n such that $2e_2 - 1 < n < \eta$ can be realized. □

We have shown that graphs of all necessary orders can be realized. It remains to show that $2e_2 - 1$ is the smallest order of a graph that can be realized; we accomplish this by showing that a graph of order $2e_2 - 2$ cannot be realized.

Lemma 3.9 *Let P be a pot with $e_1 > e_2 > 2$. Then there is no $G \in \mathcal{O}(P)$ of order $2e_2 - 2$.*

Proof. Let P be a pot where $e_1 > e_2 > 2$. We proceed by way of contradiction by assuming that there exists $G \in \mathcal{O}(P)$ with order $2e_2 - 2$. Therefore, there exists some $R_1, R_2, R_3 \geq 0$ such that $R_1 + R_2 + R_3 = 2e_2 - 2$. Since a graph is realized, we also have Equation 1, implying $R_3 = R_1e_1 - R_2e_2$. Using substitution:

$$\begin{aligned} R_1 + R_2 + R_3 &= 2e_2 - 2, \\ R_1 + R_2 + (R_1e_1 - R_2e_2) &= 2e_2 - 2, \\ R_1(e_1 + 1) + R_2(-e_2 + 1) &= -2(-e_2 + 1), \\ R_1(e_1 + 1) + (R_2 + 2)(-e_2 + 1) &= 0. \end{aligned}$$

We note that this is a homogeneous linear Diophantine equation. Recall that the $\gcd(e_1 + 1, -e_2 + 1) = 1$, so the family of solutions for our equation is

$$R_1 = (-e_2 + 1)k \quad \text{and} \quad R_2 = -(e_1 + 1)k - 2 \quad \text{for } k \in \mathbb{Z}.$$

Since $e_2 > 2$, we know that $(-e_2 + 1) < 0$ and we know that $-(e_1 + 1) - 2 < 0$. Since $R_1, R_2 \geq 0$, we know that k must be negative. However, we know

$$\begin{aligned} R_3 &= R_1e_1 - R_2e_2 \\ &= (-e_2 + 1)ke_1 - (-(e_1 + 1)k - 2)e_2 \\ &= -ke_1e_2 + ke_1 + 2e_2 + ke_1e_2 + ke_2 \\ &= 2e_2 + k(e_1 + e_2). \end{aligned}$$

We know $e_1 > e_2$, so $e_1 + e_2 > 2e_2$. Since $k < 0$, $R_3 < 0$, which contradicts the requirement $R_3 \geq 0$. Therefore, we cannot realize a graph of order $2e_2 - 2$. □

With all of our theorems and lemmas established, we complete our long journey by proving conjecture 3.4.

Theorem 3.10 *For any given pot with* $e_1 > e_2 \geq 2$, $\zeta_1 = 2e_2 - 1$.

Proof. Let P be a pot such that $e_1 > e_2 \geq 2$. We proceed by cases: when $e_2 = 2$, and when $e_2 > 2$.

We first consider the case when $e_2 = 2$. Our proposed formula would yield $\zeta_1 = 3$, so we must show that a graph of order 2 is not realized, and that graphs of every order $n \in [\zeta_1, \eta)$ are realized. Since $e_1 > e_2$, we select the minimum value of e_1 to correspond with ζ_1, which is $e_1 = 3$ (since $e_1 > e_2$ and $\gcd(4, -1) = 1$). Therefore, by Theorem 3.3, we know that $\eta = e_1 + e_2 + 2 = 7$. We summarize the cases for $e_1 = 3$ in Table 2.

Order n	(R_1, R_2, R_3)	Equation 1
3	$(1, 1, 1)$	$e_1 = e_2 + 1$
4	$(1, 0, 3)$	$e_1 = 3$
5	$(2, 3, 0)$	$2e_1 = 3e_2$
6	$(2, 2, 2)$	$2e_2 = 2e_2 + 2$

Table 2: Summary of the order, tile distribution, and the result of Equation 1 for $e_1 = 3$, $e_2 = 2$.

By Table 2 and Theorem 3.1, we know a graph is realized for every $n \geq \eta$, so we see that we can realize a graph for every $n \geq \zeta_1$.

We now proceed by contradiction and assume P can realize a graph of order 2. From P, we know that $R_1 \geq 1$, or else P would not be able to realize a graph. Therefore, for a graph of order 2, we have $(1, 1, 0)$ or $(1, 0, 1)$. For $(1, 1, 0)$, we see $e_1 = e_2$, which contradicts our assumption of $e_1 > e_2$. For $(1, 0, 1)$, we see $e_1 = 1$, which also contradicts that $e_1 > e_2 = 2$. Therefore, a graph of order 2 cannot be realized and our ζ_1 for $e_2 = 2$ is $2e_2 - 1 = 3$.

Now assume $e_2 > 2$. We first note that a graph of order ζ_1 exists by Theorem 3.6.

Next, we must show that P can realize a graph for every order n such that $n \geq \zeta_1$. From Theorem 3.1, we have $G \in \mathcal{O}(P)$ for all $n \geq \eta$, and by Lemma 3.8 and Theorem 3.6, we have $G \in \mathcal{O}(P)$ for all $n \in [\zeta_1, \eta)$. Together, this implies $G \in \mathcal{O}(P)$ for all $n \geq \zeta_1$.

Finally, the order $\zeta_1 - 1$ is not realized from Lemma 3.9. This completes the proof. □

4 Conclusion

We have shown that, for the pot $P = \{t_1 = \{a^{e_1}\}, t_2 = \{\hat{a}^{e_2}\}, t_1 = \{\hat{a}\}\}$, where $e_1 > e_2 > 0$ and $\gcd(e_1 + 1, -e_2 + 1) = 1$, we can find the minimum order graph in which all greater graph orders can be constructed by a simple formula only relying on e_2. We note that we have specific conditions for this formula to hold, which gives rise to some new outstanding questions:

1. Can we find a closed form of the other ζ values that are not ζ_1?

2. What results do we see when $e_1 < e_2$?

3. What other graphs are realized by the pot corresponding to ζ_1?

For the first question, preliminary research shows that the values of ζ seem to follow a sequential pattern as we increase e_1. Determining this sequence and proving that every value satisfies the requirements for ζ will give a closed form for every ζ given $e_1 > e_2$.

Preliminary data collection has also been done on the second question, but we left it unexplored. However, one important observation we made is that there are instances where $\zeta = \eta$. Further research and more advanced mathematical tools will be required to tackle complexes that have more tiles or do not contain a single-armed tile (as required in our choice of e_3, see entry $a_{1,3}$ in construction matrix M_P in Section 3).

The third question can be related to the graph isomorphism problem. We already know that a star graph can be realized for ζ_1 when $e_1 > e_2 > 2$, but there are different graphs that can also be realized of the same order. Figure 4 is an example of an alternate realized graph of Figure 3 that has the same ζ_1.

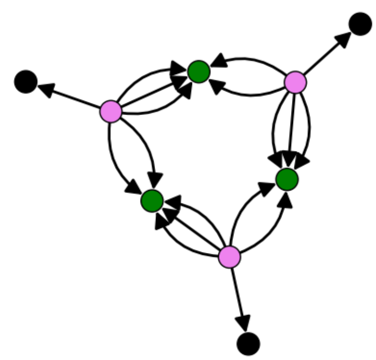

Figure 4: Graph of order 9 for $P = \{\{a^8\}, \{\hat{a}^5\}, \{\hat{a}\}\}$

Although we have a way to determine all possible tile distributions for a given order n, we do not have a way to generate all non-isomorphic graphs, and little progress has been made in this direction thus far.

5 Appendix A

Input: $e_1, e_2, thres$ for $e_1 \in \mathbb{Z}_{>0}$, $e_2 \in \mathbb{Z}_{<0}$, $thres \in \mathbb{Z}$ and $|e_1| > |e_2|$
Output: $\lceil \eta \rceil, \eta, minimal\_orders = (\zeta, R_1, R_2, R_3)$
Algorithm: order2Vars

$graphsize, soln \leftarrow list, dictionary$

$g \leftarrow \gcd(e_1 + 1, e_2 + 1)$

```
for x = 1 to thres + 1
    for y = 0 to thres + 1
        for c = 1 to e_1 · x + x + 1
            e_3 ← e_1 · x + e_2 · y
            if x + y + e_3 = c and e_3 ≥ 0
                n ← x + y + e_3
                Append n : (x, y, e_3) to soln
                Append (n, x, y, e_3) to graphsize
        endfor
    endfor
endfor
Sort graphsize
η ← max{ (e_1+1)(e_1+|e_2|)/e_1 , (|e2|−1)(e_1+|e_2|)/|e_2| }
Reverse graphsize
minimal_orders ← list
if length of graphsize = 1
    Append graphsize[0] to minimal_orders
else
    for i = 0 to length of graphsize − 1
        if (graphsize[i][0] > η
        or graphsize[i][0] − graphsize[i + 1][0] = 1)
        and i = length of graphsize − 2
            Append graphsize[i + 1] to minimal_orders
        elseif graphsize[i][0] − graphsize[i + 1][0] > 1
            Append graphsize[i] to minimal_orders
            break
    endfor
```

Acknowledgments

The authors are grateful for the thoughtful and constructive feedback provided by the reviewers. Their time and effort in reviewing this paper are sincerely appreciated.

References

[1] N.C. Seeman, An overview of structural DNA nanotechnology, *Molecular Biotechnology*, **37** (2007), 246.

[2] Y. Zhang, N.C. Seeman, The construction of a DNA truncated octahedron, *J. Amer. Chem. Soc.*, **116** (1994), 1661-1669.

[3] J.H. Chen, N.C. Seeman, Synthesis from DNA of a molecule with the connectivity of a cube, *Nature*, **350** (1991), 631-633.

[4] J. Ellis-Monaghan, G. Pangborn, L. Beaudin, D. Miller, N. Bruno, A. Hashimoto, Minimal tile and bond-edge types for self-assembling DNA graphs, *Discrete and Topological Models in Molecular Biology*, Springer, 2014, 241–270.

[5] J. Ellis-Monaghan, G. Pangborn, N. Jonoska, Tile-based DNA nanostructures: mathematical design and problem encoding, *Algebraic and combinatorial computational biology*, Academic Press, London, 2019, 35–60.

[6] D.M. Burton, *Elementary Number Theory*, McGraw-Hill, 2012.

[7] G.E. Andrews, *Number Theory*, Dover Publications, 1994.

[8] G. Chartrand, P. Zhang, *A First Course in Graph Theory*, Dover, 2017.

[9] C. Johnson, A. Lavengood, Analysis and Algorithmic Construction of Self-Assembled DNA Complexes, available online at the URL: `https://arxiv.org/abs/2310.04398`

[10] E.W. Weisstein, Star Graph, available online at the URL: `https://mathworld.wolfram.com/StarGraph.html`

Joshua Miller
Nevada State University
1300 Nevada State Drive
Henderson, Nevada
E-mail: `joshmiller5472@gmail.com`

Andrew Lavengood
College of Southern Nevada
3200 E Cheyenne Ave
North Las Vegas, Nevada
E-mail: `andrew.lavengood@csn.edu`

Received: April 8, 2025 **Accepted**: January 30, 2026
Communicated by Mike Krebs