6-2020

DNA COMPLEXES OF ONE BOND-EDGE TYPE

Andrew Tyler Lavengood-Ryan
California State University - San Bernardino

DNA Complexes of One Bond-Edge Type

A Thesis

Presented to the

Faculty of

California State University,

San Bernardino

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

in

Mathematics

---

by

Andrew Lavengood-Ryan

June 2020

DNA Complexes of One Bond-Edge Type

A Thesis

Presented to the

Faculty of

California State University,

San Bernardino

---

by

Andrew Lavengood-Ryan

June 2020

Approved by:

Cory Johnson, Committee Chair

Jeffrey S. Meyer, Committee Member

Shawn McMurran, Committee Member

David Maynard, Chair, Department of Mathematics

Corey M. Dunn, Graduate Coordinator

Abstract

DNA self-assembly is an important tool used in the building of nanostructures and targeted virotherapies. We use tools from graph theory and number theory to encode the biological process of DNA self-assembly. The principal component of this process is to examine collections of branched junction molecules, called pots, and study the types of structures that such pots can realize. In this thesis, we restrict our attention to pots which contain identical cohesive-ends, or a single bond-edge type, and we demonstrate the types and sizes of structures that can be built based on a single characteristic of the pot that is easily checked. The results of this thesis classify nearly every type of structure that can be assembled from molecules with one bond-edge type.

Acknowledgements

The author would like to express his sincere gratitude to Dr. Johnson for her support, guidance, and unwavering dedication toward the completion of this research and the writing of the thesis. Without her hours, days, weeks and months of work, this would not have been possible. Sincere thanks also go to Dr. Jeffrey Meyer and Dr. Shawn McMurran for serving on the thesis committee and providing valuable feedback. A special thank you to the entire mathematics department, especially the support staff, for the years of support and guidance. The author dedicates this thesis to all of you.

Table of Contents

List of Figures

Chapter 1

Introduction

DNA self-assembly is an interesting and vital experimental process that is being utilized in labs across the country. The use of DNA self-assembly as a bottom-up technology for creating target nanostructures was invented about 35 years ago by Nadrian C. Seeman. The process relies on the complementary nature of nucleotides that comprise the structure of DNA. DNA self-assembly has applications ranging from the construction of nanostructures to experimental virotherapies (particularly useful for treating various cancers) [EMPB$^+$14, See07]. However, the materials are expensive, and the process has a significant risk of generating a lot of waste. By representing the problem of DNA self-assembly using the tools of graph theory, we can use a combination of graph theoretic and algebraic tools to optimize the assembly process.

Target structures are built by *branched junction molecules*, which are asterisk-shaped molecules that consist of a vertex with several arms extending from the vertex. Each arm consists of a fragment of DNA where the strands are of unequal length, creating a *cohesive-end*. Each cohesive-end will bond with a complementary cohesive-end from another arm. Rather than referring to the precise nature of a cohesive-end (such as the exact nucleotide configuration), we use single alphabet letters to distinguish between cohesive-ends of different types. For example, a and b denote two non-compatible cohesive-ends, but a will bond with $\hat{a}$ while b will bond with $\hat{b}$. We use the term *bond-edge type* to refer to a pair of complementary cohesive-ends.

A collection of branched junction molecules to be used in the self-assembly process is called a pot. Previously, research was primarily concerned with determining the most efficient pot for a target complete complex [EMPB$^+$14, EMJP19]. For our research, we asked the inverse question: given a pot, what are the complete complexes that can be assembled? As is typically the case, the inverse problem proves to be considerably more difficult, and so we restrict our attention to the case where the pot contains one bond-edge type. At this time we reserve our attention to three open questions:

1. What are the sizes of the DNA complexes that can be realized by a specific pot?
2. What types and what distributions of branched junction molecules does a pot use in realizing a target DNA complex?
3. Exactly what types of DNA complexes do we expect a pot to realize? (e.g. disconnected or connected complexes)

These questions are best answered by using tools arising from both graph theory and number theory. We provide some background on these two topics in Chapter 2 before formalizing much of the DNA self-assembly process using our graph-theoretic language in Chapter 3. Chapter 4 is a collection of our results that are related to these three questions, with Chapter 5 providing some insight into future directions.

Chapter 2

Preliminaries

In this chapter we present some standard definitions and results from graph theory and number theory.

2.1 Notation

We use $\{\cdot\}$ to denote a set and $(\cdot)$ to denote a family. Note that a family is an indexed set that may contain duplicates. We reserve s to denote the order of a graph, and q, r will come from the division algorithm.

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** [CZ17].

2.3 Diophantine Equations

We provide some standard definitions and results for number theory that will be referenced later.

Definition 2.1. [Bur12] Given integers m and n, with $n > 0$, there exist unique integers q and r satisfying

$$m = nq + r \qquad 0 \leq r < n.$$

The integers q and r are called, respectively, the *quotient* and *remainder* in the division of m by n.

Lemma 2.2. Let $m, n \in \mathbb{Z}$. Then $\gcd(m, n) = \gcd(m, -n)$.

Definition 2.3. A **Linear Diophantine Equation** is a polynomial $p(x_1, \ldots, x_n) \in \mathbb{Z}[x_1, \ldots, x_n]$ such that $\deg(p(x_1, \ldots, x_n)) = 1$, and if $p(\alpha_1, \ldots, \alpha_n) = 0$ then $(\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}^n$.

Theorem 2.4. [Bur12] The linear Diophantine equation $ax + by = c$ has a solution if and only if $d \mid c$, where $d = \gcd(a, b)$.

Notice that, for a linear Diophantine equation, x and y may be negative. However, in Chapter 4 we will restrict ourselves to those equations where $x \geq 0$ and $y \geq 0$.

Chapter 3

Encoding DNA Complexes using Graph Theory

In this chapter, we encode the structure of DNA complexes into graph-theoretic language. The nanostructures we wish to build are composed of k-armed branched junction molecules, which can be thought of as a vertex with k arms of double-stranded DNA. Two arms can bond only if they have complementary base pairings. A visualization of these molecules can be seen in Figure 3.1.

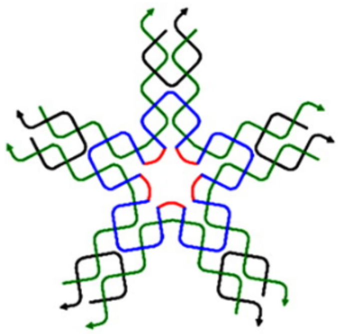


Figure 3.1: Example of a 5-armed Branched Junction Molecule

The following discussion, which can be found in [EMPB$^+$14, EMJP19, JMS11], provides the combinatorial abstraction of DNA self-assembly. See Figure 3.2 below for an example of Definition 3.1.

Definition 3.1. Consider a k-armed branched junction molecule.

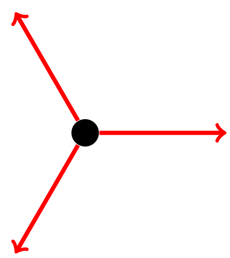
1. A k-armed branched junction molecule is modeled by a *tile*. A tile is a vertex with k half-edges representing the *cohesive-ends* (or arms) of the molecule $a, b, c, \ldots$. We will denote complementary cohesive-ends with $\hat{a}, \hat{b}, \hat{c}, \ldots$.

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.

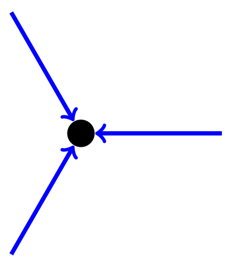
4. A *pot* is a collection of tiles such that for any cohesive-end type that appears on any tile in the pot, its complement also appears on some tile in the pot. We denote a pot by P.

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.



$t_1 = \{a^3\}$

(a) Three a Cohesive-Ends



$t_2 = \{\hat{a}^3\}$

(b) Three $\hat{a}$ Cohesive-Ends

Figure 3.2: Two Examples of Tiles

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. Although in a laboratory setting it is possible for structures to form with unpaired cohesive-ends, for mathematical modeling (and in order to satisfy the definition of a graph) we only consider complete complexes. In other words, each cohesive-end of a tile must match with a complementary cohesive-end to form edges so that no half-edges remain.

Here we develop a tool that is central to our methodology in Chapter 4. This tool captures the characteristic of a pot that we are interested in. Let $P = \{t_1, \ldots, t_p\}$ be a pot with p tile types, and define $A_{i,j}$ to be the number of cohesive ends of type a_i on tile t_j and $\hat{A}_{i,j}$ to be the number of cohesive ends of type $\hat{a}_i$ on tile t_j. Suppose a target graph G of order s is realized by P using R_j tiles of type j. Since we consider only complete complexes, we have the following equations:

$$\sum_j R_j = s \text{ and } \sum_j R_j(A_{i,j} - \hat{A}_{i,j}) = 0 \qquad \text{for all } i. \tag{3.1}$$

Define $z_{i,j} = A_{i,j} - \hat{A}_{i,j}$ and r_j to be the proportion of tile-type t_j used in the construction of G. These with Equation 3.1 give the equations

$$\sum_j r_j = 1 \text{ and } \sum_j r_j z_{i,j} = 0 \qquad \text{for all } i. \tag{3.2}$$

The equations in Equation 3.2 naturally define a matrix associated to P.

Definition 3.2. 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

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

Because, in general, there are infinitely many solutions to the system of equations defined by M_P, it is desirable to concisely express these solutions.

Definition 3.3. The solution space of M_P is called the **spectrum** of P, and is denoted by $\mathcal{S}(P)$.

The following lemma, which may be found in [EMPB$^+$14], connects $\mathcal{S}(P)$ to graphs realized by P.

Lemma 3.4. Let $P = \{t_1, \ldots, t_p\}$. If $\langle r_1, \ldots, r_p\rangle \in \mathcal{S}(P)$, and there exists an $s \in \mathbb{Z}_{\geq 0}$ such that $sr_j \in \mathbb{Z}_{\geq 0}$ for all j, then there is a graph of order s such that $G \in \mathcal{O}(P)$ using sr_j tiles of type t_j.

This thesis will focus exclusively on pots with one bond-edge type, and so M_P will be a $2 \times n$ matrix. Here we provide an example of how to build the construction matrix from a pot P.

Example 3.5. Consider the pot $P = \{t_1, t_2, t_3\}$ where $t_1 = \{a^3\}$, $t_2 = \{\hat{a}\}$, and $t_3 = \{\hat{a}^3\}$. In this case we have

$$t_1 : z_{1,1} = 3$$
$$t_2 : z_{1,2} = -1$$
$$t_3 : z_{1,3} = -3.$$

The construction matrix is

$$M_P = \left[\begin{array}{ccc|c} 3 & -1 & -3 & 0 \\ 1 & 1 & 1 & 1 \end{array}\right].$$

To determine $\mathcal{S}(P)$, row-reduce M_P to obtain

$$\text{rref}(M_P) = \left[\begin{array}{ccc|c} 1 & 0 & \frac{1}{2} & \frac{1}{4} \\ 0 & 1 & \frac{3}{2} & \frac{3}{4} \end{array}\right].$$

Now we write the solutions of this matrix in terms of the variables x, y, z:

$$\begin{aligned} x &= \frac{1}{4} - \frac{1}{2}z \\ &= \frac{1}{4}(1 - 2z), \\ y &= \frac{3}{2} - \frac{3}{4}z \\ &= \frac{1}{4}(6 - 3z), \end{aligned}$$

$$
\begin{aligned}
z &= \frac{4}{4}z \\
&= \frac{1}{4}(4z).
\end{aligned}
$$

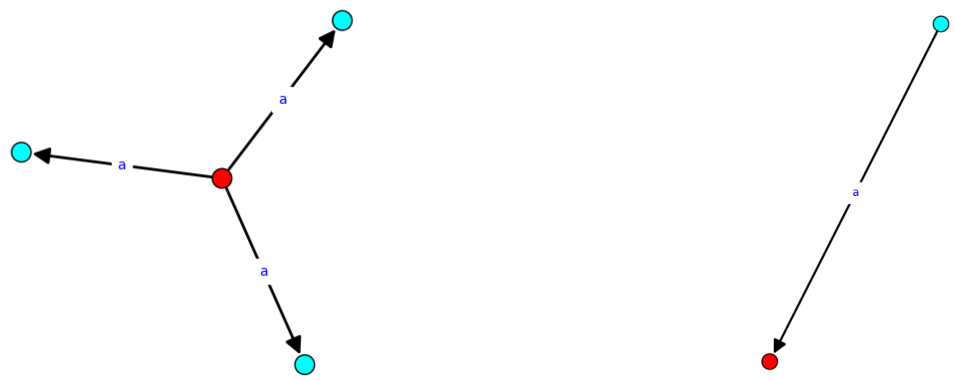
Thus we have

$$\mathcal{S}(P) = \left\{ \frac{1}{4} \langle 1-2z, 6-3z, 4z \rangle \mid z \in \mathbb{Q}_{\geq 0} \right\},$$

and P realizes, for example, a graph of order 4.

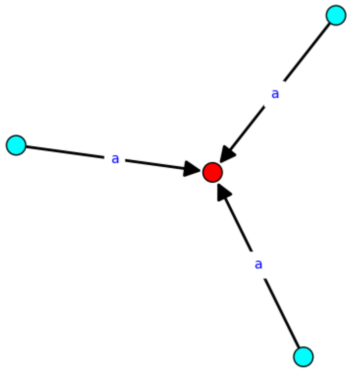
3.1 Modeling DNA Self-Assembly with Sage

For the remainder of the paper, we will use an algorithm we created in Sage to model the DNA self-assembly process. When modeled using Sage, each tile is represented as a star graph on $k+1$ vertices. For clarity in the paper, we color-code the tiles so that the darker vertices will comprise the vertices of the complete DNA complex, and the lighter vertices will vanish by the end of the algorithm for constructing graphs. Figures 3.3 and 3.4 provide four examples of tiles modeled by Sage.

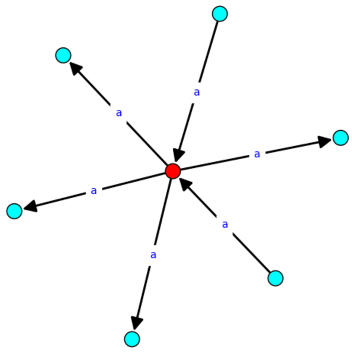


(a) $t_1 = \{a^3\}$ (b) $t_2 = \{\hat{a}\}$

Figure 3.3: Two Tiles Built in Sage



(a) $t_3 = \{\hat{a}^3\}$



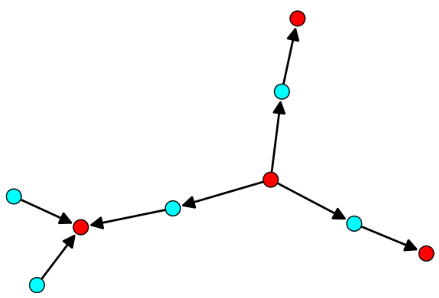
(b) $t_4 = \{a^4, \hat{a}^2\}$

Figure 3.4: Two More Tiles Built in Sage

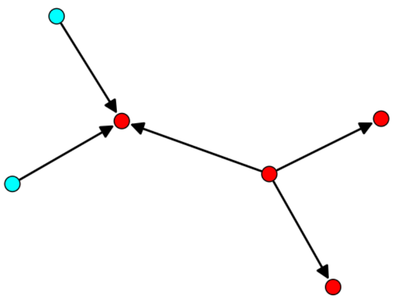
In order to build a graph from a set of tiles, we take the union of tiles at the lighter vertices and then we edge-contract to merge the lighter vertices with the darker vertices while preserving the direction of the edges. An example of this algorithm is given in Example 3.6.

Example 3.6. Recall from Example 3.5 the pot $P = \{t_1, t_2, t_3\}$ where $t_1 = \{a^3\}$, $t_2 = \{\hat{a}\}$, and $t_3 = \{\hat{a}^3\}$.

One graph $G \in \mathcal{O}(P)$ has order 6 and uses the tile distribution $(2, 3, 1)$. First, take the union of one copy of t_1, two copies of t_2, and one copy of t_3; see Figure 3.5a. In the second step, edge contract wherever there is a lighter vertex lying between two darker vertices; see Figure 3.5b.



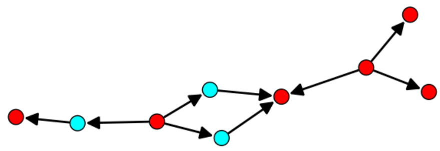
(a) Algorithm Step 1



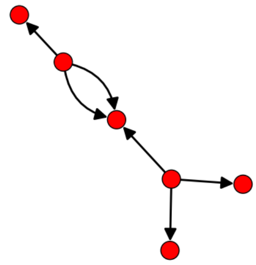
(b) Algorithm Step 2

Figure 3.5: First Steps in Algorithm

The third step is to take the union of the graph in Figure 3.5b with one copy of t_1 and one copy of t_2 to obtain the graph in Figure 3.6a. Finally, for the fourth step, perform an edge contraction wherever there is a teal vertex lying between red vertices to obtain a graph with only red vertices, as in Figure 3.6b.



(a) Algorithm Step 3



(b) Algorithm Step 4

Figure 3.6: Last Steps in Algorithm

The graph modeled in Figure 3.6b is the same graph that can be realized in a laboratory setting.

Chapter 4

Pots with a 1-Armed Tile

This research focuses on pots that contain at least one 1-armed tile. That is, $P_1 = \{\{a^m\}, \{\hat{a}\}, \{\hat{a}^n\}, \ldots\}$ for some $m > 0, n > 1$. In all but one result, which is stated explicitly, we assume $m > n$. Note that all of the results here can be stated identically for $P_1' = \{\{a^m\}, \{a\}, \{\hat{a}^n\}, \ldots\}$ where $m < n$. The pot P_1 has corresponding construction matrix

$$M_{P_1} = \left[\begin{array}{cccc|c} m & -1 & -n & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 1 \end{array}\right].$$

Unless otherwise specified, for the remainder of the chapter we reserve the notation P for the pot $P = \{\{a^m\}, \{\hat{a}\}, \{\hat{a}^n\}\}$ because $\mathcal{S}(P_1)$ has at least one variable. The pot P has construction matrix

$$M_P = \left[\begin{array}{ccc|c} m & -1 & -n & 0 \\ 1 & 1 & 1 & 1 \end{array}\right]. \tag{4.1}$$

The spectrum of P is described in Lemma 4.1.

Lemma 4.1. Consider the pot P with the associated construction matrix M_P. Then

$$\mathcal{S}(P) = \left\{ \frac{1}{m(m+1)} \left\langle m - (m - nm)z, m^2 - (nm + m^2)z, (m^2 + m)z \right\rangle \mid z \in \mathbb{Q}_{\geq 0} \right\}.$$

Proof. Row-reduce M_P to obtain

$$\operatorname{rref}(M_P) = \left[\begin{array}{ccc|c} 1 & 0 & -\frac{n}{m} + \frac{\frac{n}{m}+1}{m(\frac{1}{m}+1)} & \frac{1}{m(\frac{1}{m}+1)} \\ 0 & 1 & \frac{\frac{n}{m}+1}{\frac{1}{m}+1} & \frac{1}{\frac{1}{m}+1} \end{array}\right]. \tag{4.2}$$

Equation 4.2 can be simplified to

$$\operatorname{rref}(M_P) = \left[\begin{array}{ccc|c} 1 & 0 & \frac{m-nm}{m(m+1)} & \frac{1}{m+1} \\ 0 & 1 & \frac{n+m}{m+1} & \frac{m}{m+1} \end{array}\right]. \tag{4.3}$$

From Equation 4.3, we have

$$\begin{aligned} x &= \frac{1}{m+1} - \frac{m-nm}{m(m+1)}z \\ &= \frac{1}{m(m+1)}[m-(m-nm)z], \qquad (4.4) \\ y &= \frac{m}{m+1} - \frac{n+m}{m+1}z \\ &= \frac{m^2}{m(m+1)} - \frac{nm+m^2}{m(m+1)}z \\ &= \frac{1}{m(m+1)}[m^2-(nm+m^2)z], \qquad (4.5) \\ z &= \frac{m(m+1)}{m(m+1)}z \\ &= \frac{1}{m(m+1)}[(m^2+m)z]. \qquad (4.6) \end{aligned}$$

Equations 4.4, 4.5 and 4.6 yield the desired result. □

The immediate concern is if P can realize a graph. The following lemma describes two types of graphs realized by P.

Lemma 4.2. The pot P realizes a graph of order $m+1$ and a graph of order $q+r+1$, where $m = nq+r$, $0 \le r < n$.

Proof. Set $z = 0$. Then we have the particular solution

$$\frac{1}{m(m+1)}\langle m, m^2, 0\rangle = \left\langle \frac{1}{m+1}, \frac{m}{m+1}, 0 \right\rangle.$$

By Lemma 3.4, the graph of order $m+1$ has a tile distribution of $(1, m, 0)$.

Set $z = \frac{q}{q+r+1}$. We will substitute $m = nq+r$ in strategic places in this proof.

Then we have the particular solution

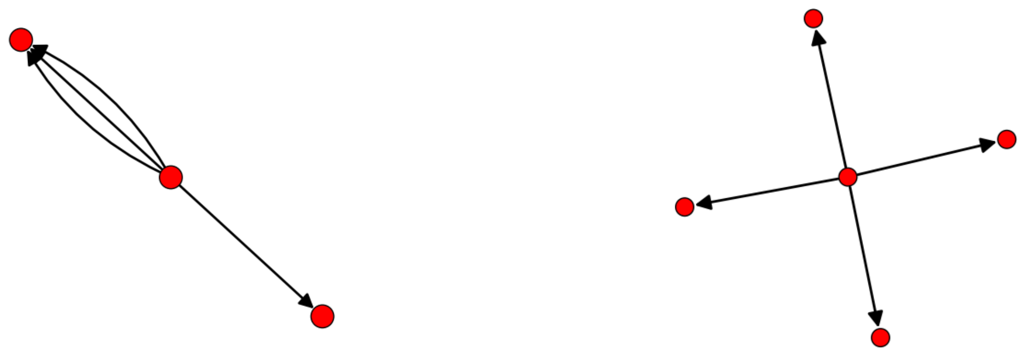
$$\begin{aligned}
&\frac{1}{m(m+1)}\left\langle m-(m-nm)\frac{q}{q+r+1}, m^2-(nm+m^2)\frac{q}{q+r+1}, \frac{(m^2+m)q}{q+r+1}\right\rangle\\
&=\frac{1}{m^2+m}\left\langle \frac{mr+m+mnq}{q+r+1}, \frac{m^2r+m^2-mnq}{q+r+1}, \frac{(m^2+m)q}{q+r+1}\right\rangle\\
&=\frac{1}{m^2+m}\left\langle \frac{m+m(nq+r)}{q+r+1}, \frac{n^2q^2r+2nqr^2+r^3+nqr+r^2}{q+r+1}, \frac{(m^2+m)q}{q+r+1}\right\rangle\\
&=\frac{1}{m^2+m}\left\langle \frac{m^2+m}{q+r+1}, \frac{r((n^2q^2+2nqr+r^2)+(nq+r))}{q+r+1}, \frac{(m^2+m)q}{q+r+1}\right\rangle\\
&=\frac{1}{m^2+m}\left\langle \frac{m^2+m}{q+r+1}, \frac{r(m^2+m)}{q+r+1}, \frac{q(m^2+m)}{q+r+1}\right\rangle\\
&=\left\langle \frac{1}{q+r+1}, \frac{r}{1+q+r}, \frac{q}{q+r+1}\right\rangle.
\end{aligned}$$

Hence the graph of order $q+r+1$ has a tile distribution of $(1, r, q)$. □

The graphs from Lemma 4.2 are important enough to be named. The following definitions are reserved for this paper.

Definition 4.3. Let $P = \{\{a^m\}, \{\hat{a}\}, \{\hat{a}^n\}\}$. If $G \in \mathcal{O}(P)$ and the order of G is $m+1$, then G is called a **fundamental graph** of P, denoted G_F. If $G \in \mathcal{O}(P)$, and the order of G is $q+r+1$, then G is called a **minimal graph** of P, denoted $G_{\min}$.

Example 4.4. Let $P = \{\{a^4\}, \{\hat{a}\}, \{\hat{a}^3\}\}$. Then, according to Definition 4.3, we can construct a fundamental graph of order 5 and a minimal graph of order 3. Each of these graphs is provided below.



(a) $G_{\min}$ (b) G_F

Figure 4.1: Two Graphs for $P = \{a^4, \hat{a}, \hat{a}^3\}$

The order of $G_{\min}$ is significant enough to warrant its own notation. The following proposition from [EMPB$^+$14] is valid for any number of bond-edge types.

Proposition 4.5. [EMPB$^+$14] The order of $G_{\min}$, denoted m_P, is

$$m_P = \min\left\{\operatorname{lcm}\left\{b_j \mid r_j \neq 0 \text{ and } r_j = \frac{a_j}{b_j}\right\} \text{ where } \langle r_1, \ldots, r_p\rangle \in \mathcal{S}(P)\right\},$$

where the minimum is taken over all solutions to $M(P)$ such that $r_j \geq 0$ and $\frac{a_j}{b_j}$ is in reduced form for all j.

This next proposition demonstrates that our notation for $G_{\min}$ is appropriate.

Proposition 4.6. When P contains one bond-edge type, then $m_P = q + r + 1$.

Proof. In constructing the smallest graph realized by P, we must use t_1 exactly once; otherwise, a complete complex cannot be formed since t_2 and t_3 both consist of hatted cohesive-ends. Since t_3 contains more arms than t_2, we maximize the usage of tile-type t_3, and complete the complex by adjoining the appropriate number of tiles of type t_2. This is exactly the process of performing the division algorithm, where $m = nq + r$. Since t_1 was used once, this smallest graph is a graph on $q + r + 1$ vertices, as desired. □

Example 4.4 demonstrates the order of any minimal graph is less than the order of any fundamental graph; this next proposition establishes this fact is always true.

Proposition 4.7. The order of any minimal graph is strictly less than the order of any fundamental graph.

Proof. Recall from the Division Algorithm that $m = nq+r$, and hence $m+1 = nq+r+1$. Now $n > 0$, and so $m + 1 = nq + r + 1 \geq q + r + 1$. In particular, $m + 1 = q + r + 1$ only when $n = 1$, but we have specified $n > 1$ for P. Therefore, $m + 1 > q + r + 1$. □

4.1 Orders of Graphs Realized by P

We begin by studying the orders of graphs realized by P. We show that if $G \in \mathcal{O}(P)$, then the order of G is dependent upon $\gcd(m + 1, -n + 1)$. The most straightforward case is presented in our first theorem.

Theorem 4.8. For the pot P, if $\gcd(m + 1, -n + 1) = d \neq 1$, then P realizes a graph of order s if and only if $s = kd$ where $k \in \mathbb{Z}_{\geq 0}$.

Proof. Let P be a pot of tiles with associated M_P, and suppose $\gcd(m+1, -n+1) = d \neq 1$. From Equation 4.1, we know $\mathcal{S}(P) = \left\{\left\langle \frac{x}{s}, \frac{y}{s}, \frac{z}{s} \right\rangle \mid s \in \mathbb{Z}_{\geq 0}\right\}$, and we have the system

$$\begin{cases} mx - y - nz = 0, \\ x + y + z = s. \end{cases} \tag{4.7}$$

Adding these equations, we obtain

$$(m+1)x + (-n+1)z = s. \tag{4.8}$$

This is a linear Diophantine equation in two variables. Since $\gcd(m+1, -n+1) = d \neq 1$, by Theorem 2.5 a solution to this equation exists if and only if $s = kd$, which establishes our desired result. $\square$

Example 4.9. Let $P = \{\{a^5\}, \{\hat{a}\}, \{\hat{a}^4\}\}$. Then by Theorem 4.8 and 4.8, we examine the family of linear Diophantine equations

$$6x + 3y = s,$$

each of which has a solution if and only if $s = 3k$ for some $k \in \mathbb{Z}_{\geq 0}$. Since $m_P = 3$, there exists a $G \in \mathcal{O}(P)$ for each $k \in \mathbb{Z}_{\geq 0}$. Examples for $s = 3$ and $s = 9$ are shown in Figures 4.2 and 4.3.

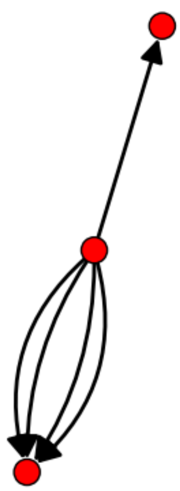


Figure 4.2: $G_{\min}$ for $\{\{a^5\}, \{\hat{a}\}, \{\hat{a}^4\}\}$

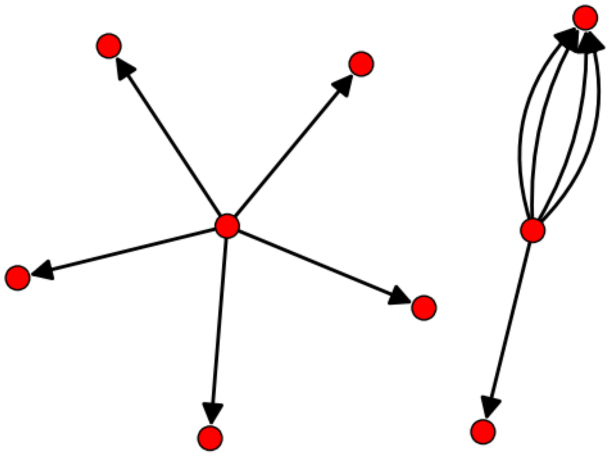


Figure 4.3: Graph of Order 9 for $\{\{a^5\}, \{\hat{a}\}, \{\hat{a}^4\}\}$

Note that in Figure 4.3, we have a disconnected graph on 9 vertices that is the union of $G_{\min}$ and G_F. Although a connected graph exists, we allow disconnected graphs as well.

It need not be the case that P realizes a graph for each $k \in \mathbb{Z}_{\geq 0}$ in the case of Theorem 4.8. For example, the pot $P = \{\{a^{11}\}, \{\hat{a}\}, \{\hat{a}^4\}\}$ has a minimal graph of order 6, so this pot does not realize a graph on 3 vertices. In fact, for the case $\gcd(m+1, -n+1) = 2$, we can realize a graph of order 2 only when $m = n$. Corollary 4.10 generalizes this idea.

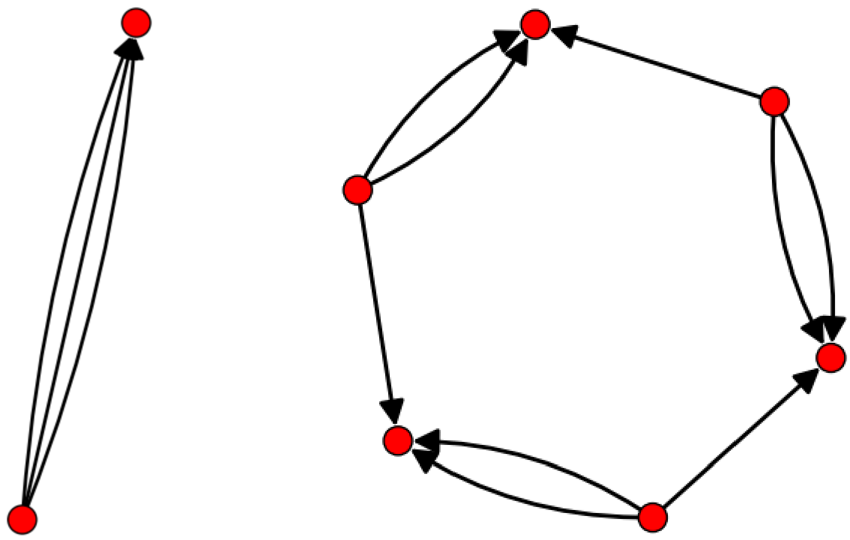
Corollary 4.10. Let P be a pot where $m = 2k+1$ for some $k \in \mathbb{Z}_{\geq 0}$. Then P realizes a graph for all orders s where $s \in 2\mathbb{Z}_{\geq 0}$.

Proof. From Theorem 4.8, it is sufficient to notice

$$\begin{aligned} \gcd(m+1, -m+1) &= \gcd((2k+1)+1, -(2k+1)+1) \\ &= \gcd(2(k+1), -2k). \end{aligned}$$

From Lemma 2.8, we know $\gcd(2(k+1), -2k) = \gcd(2(k+1), 2k) = 2$. Hence all graphs realized by P must have order 2α for $\alpha \in \mathbb{Z}_{\geq 0}$. $\square$

Example 4.11. Let $P = \{\{a^3\}, \{\hat{a}\}, \{\hat{a}^3\}\}$. Then by Corollary 4.10, P realizes a graph on every even number of vertices. Two such graphs are provided below: the minimal graph on 2 vertices, and a graph on 6 vertices.



(a) $G_{\min}$ (b) Graph of Order 6

Figure 4.4: Two Graphs for $P = \{\{a^3\}, \{\hat{a}\}, \{\hat{a}^3\}\}$

One might wonder if a similar corollary exists for the case $m = 2k$. Unfortunately, this case is not as immediate, as we see in this next proposition.

Proposition 4.12. For any $m \in 2\mathbb{Z}_{\geq 0}$, we have $\gcd(m+1, -m+1) = 1$.

Proof. Let $m \in 2\mathbb{Z}$. Suppose $\gcd(m+1, -m+1) = d$. Then

$$\begin{aligned} \gcd(m+1, -m+1) &= \gcd(2k+1, -2k+1) \\ &= \gcd(2k+1, 2k-1). \end{aligned}$$

Consider the linear Diophantine equation $(2k+1)x + (2k-1)y = 1$. Observe $x = k$, $y = k-1$ is a solution, so by Theorem 2.5, $d \mid 1$ and hence $d = 1$. $\square$

The case in which $\gcd(m+1, -n+1) = 1$ is more difficult. We first provide a motivating example.

Example 4.13. Consider the pot $P = \{\{a^6\}, \{\hat{a}\}, \{\hat{a}^4\}\}$. The associated construction matrix is

$$M_P = \left[\begin{array}{ccc|c} 6 & -1 & -4 & 0 \\ 1 & 1 & 1 & 1 \end{array}\right].$$

A minimal graph of P can be found from the division algorithm. Set $6 = 4q + r$ to arrive at $q = 1$ and $r = 2$. Then $m_P = 4$.

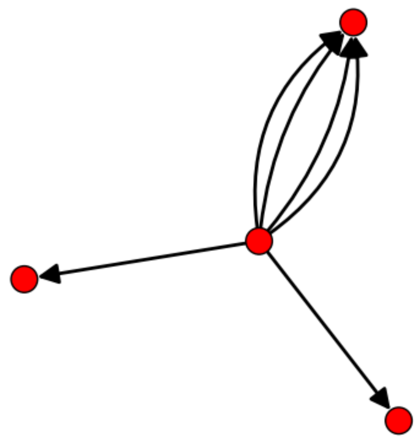


Figure 4.5: $G_{\min}$ for $\{\{a^6\}, \{\hat{a}\}, \{\hat{a}^4\}\}$

P has the associated family of linear Diophantine equations $7x - 3z = s$. Using a program we created in Sage, which uses brute force to find all tile distributions for every s (within some reasonable bound), we determine that this equation has a solution in the nonnegative integers for all s except $s = 1, 2, 3, 6$. That is, P realizes a graph for the orders $\{4, 5, 7, 8, 9, \ldots\}$.

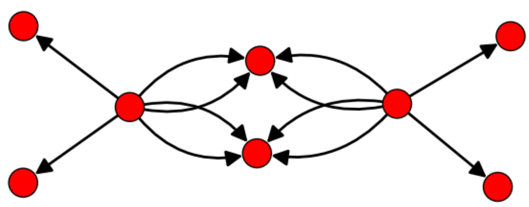


Figure 4.6: Graph of order 8 for $\{\{a^6\}, \{\hat{a}\}, \{\hat{a}^4\}\}$

Two important observations from Example 4.13 are that, after some threshold,

P realizes a graph for every order, and that threshold is not necessarily m_P. We formalize this idea in the following definition.

Definition 4.14. Let $G_P = \{G \mid G \in \mathcal{O}(P)\}$ and consider $S_P = \{s \mid s+n$ is the order of some $G \in G_P$ for every $n \in \mathbb{N}\}$. The **lower density bound** of P is $\zeta = \min(S_P)$.

In general, it is difficult to predict ζ for P. However, we can provide a lower bound that is close to ζ. This is the focus of Theorem 4.15.

Theorem 4.15. Consider the pot P where $\gcd(m+1, -n+1) = 1$. Then P realizes a graph for every s with $s \geq \max\left\{\frac{(m+1)(m+n)}{m}, \frac{(n-1)(m+n)}{n}\right\}$.

Proof. Recall Equation 4.8, where s varies over $\mathbb{Z}_{\geq 0}$. Solving for z and setting $z = f(x)$ gives the function

$$f(x) = \frac{s-(m+1)x}{-n+1}.$$

We now derive the bounds on x and $f(x)$ for which we can guarantee that a solution to Equation 4.8 exists. By Lemma 3.4 since $x \geq 0$ and $z = f(x) \geq 0$, then we proceed in finding the upper bounds. The key observation is to notice $y = s - (x + f(x))$ from Equation 4.7. To find the upper bound on x (correspondingly, $f(x)$), we determine the value for which $y = 0$, since as x (and, correspondingly, $f(x)$) increases, y decreases. Substituting $f(x)$ into the equation $y = s - (x + f(x))$:

$$\begin{aligned} y &= s - \left(x + \frac{s-(m+1)x}{-n+1}\right) \\ &= s - \left(\frac{s+(-n+1)x-(m+1)x}{-n+1}\right) \\ &= \frac{(-n+1)s - s + (m+1)x - (-n+1)x}{-n+1} \\ &= \frac{-ns+(m+n)x}{-n+1}. \end{aligned}$$

Thus, when $y \geq 0$, we have $x \leq \frac{ns}{m+n}$. This provides the bounds

$$\begin{cases} 0 \leq x \leq \frac{ns}{m+n}, \\ 0 \leq f(x) \leq f\left(\frac{ns}{m+n}\right) = \frac{ms}{m+n}. \end{cases}$$

A graph of order s realized by P corresponds to an integer solution to Equation 4.8. Geometrically, we are looking for lattice points in $\mathbb{R}^2$ that lie on $f(x)$ within the rectangle

$$\begin{cases} 0 \le x \le \frac{ns}{m+n}, \\ 0 \le f(x) \le \frac{ms}{m+n}. \end{cases}$$

Now the slope of $f(x)$ is $\frac{m+1}{n-1}$. Notice if $(x, f(x)) \in \mathbb{Z}^2$, then $(x+(n-1), (f(x)+(m+1)) \in \mathbb{Z}^2$ and is a point on $f(x)$. Figure 4.7 below highlights the importance of this observation. The purple leftmost vertical line and blue upper horizontal line represent the bounds on x and $f(x)$, respectively, while the black lines represent $m+1$ and $n-1$. The green line with positive slope is the function $f(x)$.

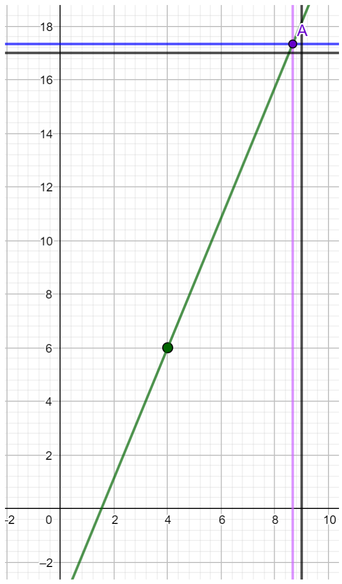


Figure 4.7: Diagram of Bounds on x and $f(x)$

In the worst case scenario, if $(x, f(x)) = (0, 0)$, then $(x+(n-1), f(x)+(m+1)) = (n-1, m+1)$, which must lie inside the bounds rectangle. Thus, we can guarantee at least one desired lattice point lives in the rectangle if the inequalities

$$\begin{cases} m+1 \le \frac{ms}{m+n}, \\ n-1 \le \frac{ns}{m+n}, \end{cases}$$

are both satisfied. Thus, by solving both inequalities for s, we conclude that P realizes a graph for every s with $s \ge \max\left\{\frac{(m+1)(m+n)}{m}, \frac{(n-1)(m+n)}{n}\right\}$. □

Remark 4.16. We denote the lower bound derived in Theorem 4.15 by η. That is, $\eta = \max\left\{\frac{(m+1)(m+n)}{m}, \frac{(n-1)(m+n)}{n}\right\}$.

Although η is an important lower bound, our research has shown $\zeta < \eta$ in general. Example 4.17 demonstrates this difference.

Example 4.17. Recall the pot $P = \{\{a^6\}, \{\hat{a}\}, \{\hat{a}^4\}\}$ from Example 4.13. We saw that $\zeta = 7$ for this pot, but

$$\begin{aligned}\eta &= \max\left\{\frac{(6+1)(6+4)}{6}, \frac{(4-1)(6+4)}{4}\right\}\\ &= \max\left\{\frac{70}{6}, \frac{30}{4}\right\}\\ &\approx \max\{11.67, 7.5\}\\ &\approx 11.67.\end{aligned}$$

We have not found a case where $\zeta = \eta$ thus far in our research, but we are not convinced this case does not occur. Despite the fact that $\zeta \leq \eta$, there are only finitely many orders to check for a pot P to determine the value of ζ. One need only check all orders for $m_P \leq s \leq \eta$.

4.2 Connected and Disconnected Graphs

Knowing orders of the graphs that can be realized by P allows us to address the next question of the types of graphs realized by P. The following theorem demonstrates when we can expect disconnected graphs.

Theorem 4.18. The pot P realizes a disconnected graph for an order s if and only if, for at least one tile distribution $(R_{11}, \ldots, R_{j1})$ associated to s, we can write

$$(R_{11}, \ldots, R_{j1}) = \sum_{i=2}^{k} (R_{1i}, \ldots, R_{ji}),$$

where each j-tuple $(R_{1i}, \ldots, R_{ji})$ is a tile distribution of P that realizes a graph.

Proof. Suppose P realizes a disconnected graph G for an order s and let $(R_{11}, \ldots, R_{j1})$ be the tile distribution used to realize G. Then

$$G = \bigcup_{i=1}^{k} H_i \tag{4.9}$$

where $H_i \cap H_j = \varnothing$ for $i \neq j$. Since each H_i is a graph, and hence a complete complex, there must be some tile distribution, namely $(R_{1i}, \ldots, R_{ji})$, that constructs H_i. Further, since each $H_i \subset G$, it follows that $R_{\ell i} \leq R_{\ell 1}$ for all ℓ. Thus we translate Equation 4.9 into the language of tile distributions to arrive at

$$(R_{11}, \ldots, R_{j1}) = \sum_{i=2}^{k} (R_{1i}, \ldots, R_{ji}),$$

where each j-tuple $(R_{1i}, \ldots, R_{ji})$ is a tile distribution of P that realizes some graph H_i.

Conversely, if we suppose an order s has at least one tile distribution

$$(R_{11}, \ldots, R_{j1}) = \sum_{i=2}^{k} (R_{1i}, \ldots, R_{ji})$$

where each j-tuple is a tile distribution of P that realizes a graph, then it follows immediately that P realizes a disconnected graph of order s. □

The above theorem provides the conditions under which P will realize disconnected graphs, but says nothing about the frequency with which P will realize disconnected graphs. As with the previous section, this question has a more straightforward answer when $\gcd(m+1, -n+1) \neq 1$.

Corollary 4.19. For the pot P, if $\gcd(m+1, -n+1) = d \neq 1$, then P realizes a disconnected graph of order s if and only if

$$s = s_1 + s_2 + \cdots + s_\ell$$

where $\ell \in \mathbb{N}$, $s_i = k_i d$ for some $k_i \in \mathbb{Z}_{\geq 0}$ and $s_i \geq m_P$.

Proof. The proof is immediate from Theorem 4.8 and Theorem 4.18. □

Example 4.20. Consider the pots $P_1 = \{\{a^5\}, \{\hat{a}\}, \{\hat{a}^4\}\}$ and $P_2 = \{\{a^9\}, \{\hat{a}\}, \{\hat{a}^6\}\}$. We have already seen in Example 4.9 that P_1 realizes graphs of orders 3 and 6, which allowed us to produce the graph in Figure 4.3, which is reproduced here.

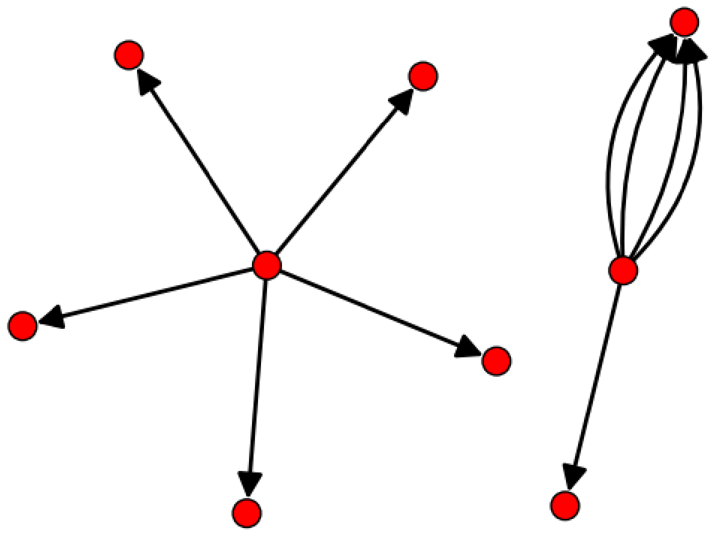


Figure 4.8: Graph of Order 9 for $\{\{a^5\}, \{\hat{a}\}, \{\hat{a}^4\}\}$

For the pot P_2, the associated minimal $G_{\min}$ has order 5. Hence we can realize a disconnected graph on 10 vertices by using two copies of $G_{\min}$, as in Figure 4.9.

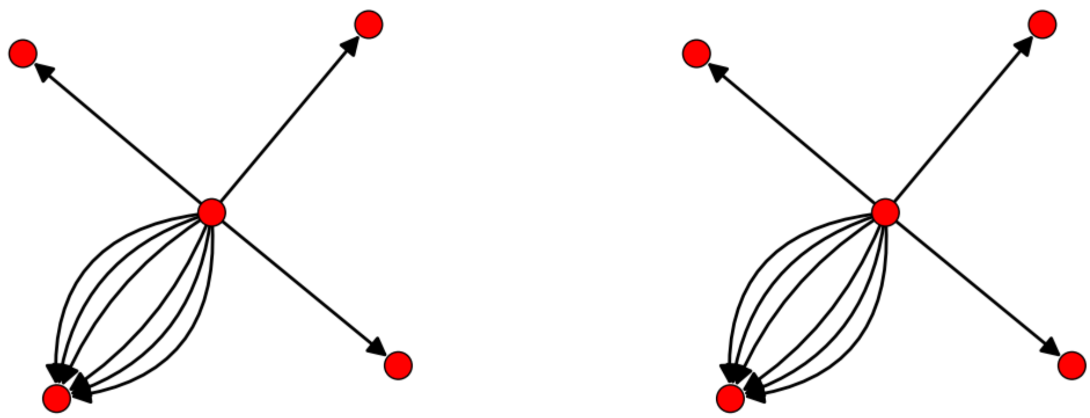


Figure 4.9: Disconnected Graph of Order 10 for $\{\{a^9\}, \{\hat{a}\}, \{\hat{a}^6\}\}$

Surprisingly, the case when $\gcd(m+1, -n+1) = 1$ is not significantly harder to resolve. In this case, there will be a dependency on ζ, and the condition will not be as strong.

Corollary 4.21. For the pot P, if $\gcd(m+1, -n+1) = 1$, then P realizes a disconnected graph of order s if

$$s = s_1 + s_2 + \cdots + s_\ell,$$

where $\ell \in \mathbb{N}$ and each $s_i \geq \zeta$.

Proof. The proof is immediate from Definition 4.14 and Theorem 4.18. □

Example 4.22. Consider the pot $P = \{\{a^7\}, \{\hat{a}\}, \{\hat{a}^4\}\}$. We have $\gcd(8, -3) = 1$, so we first find ζ and η. Again with the assistance of our brute force Sage program, we find $\zeta = 7$ and, from Theorem 4.15, we know $\eta = \lceil 12.57 \rceil = 13$. We are guaranteed a disconnected graph on 15 vertices by using a graph on 7 vertices and a graph on 8 vertices. These graphs are provided in Figure 4.10 below.

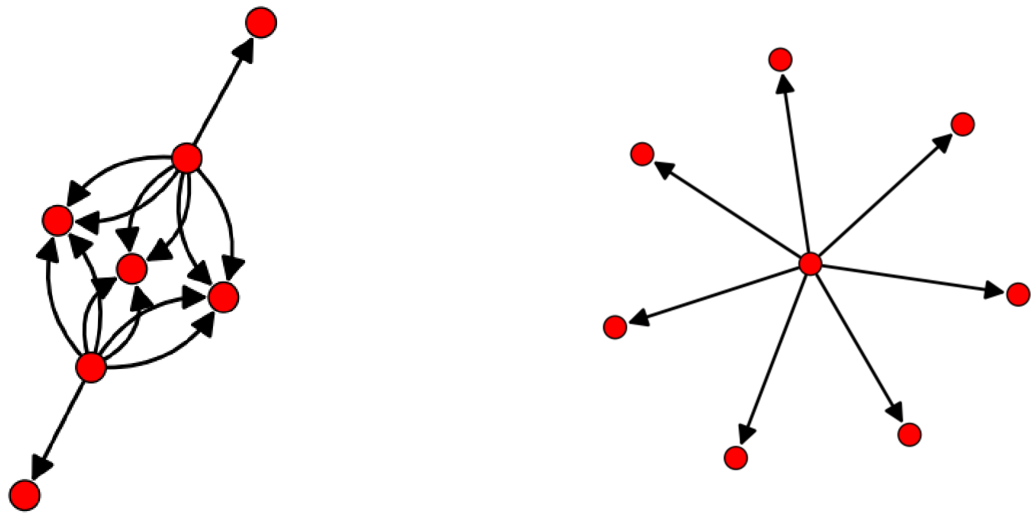


Figure 4.10: Disconnected Graph of Order 15 for $\{\{a^7\}, \{\hat{a}\}, \{\hat{a}^4\}\}$

What is perhaps more interesting is that we can obtain a disconnected graph on 12 vertices. This is because $m_P = 5 < \zeta$. This graph is included in Figure 4.11 below.

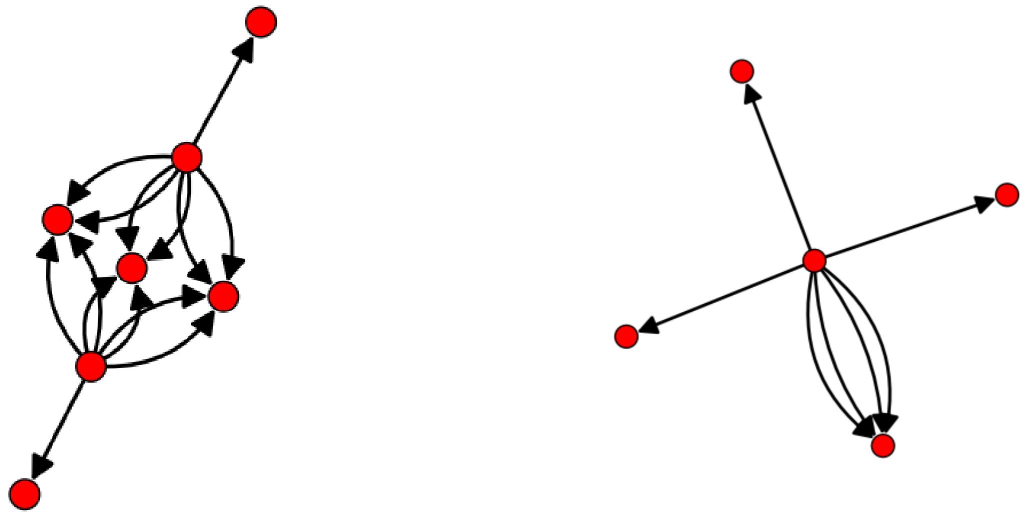


Figure 4.11: Disconnected Graph of Order 12 for $\{\{a^7\}, \{\hat{a}\}, \{\hat{a}^4\}\}$

What remains as an open problem is the question of when connected graphs are realized. The following algorithm provides a partial answer to this question.

Algorithm 4.23. Let $P = \{\{a^m\}, \{\hat{a}\}, \{\hat{a}^n\}\}$ where $m, n \in \mathbb{Z}_{\geq 0}$. By providing an order s, this algorithm can output a connected graph G of order s such that $G \in \mathcal{O}(P)$.

1. Compute the following formula for k:

$$k = \frac{s}{1 + \frac{m}{2} - \frac{n}{2}}.$$

   If $k \in 2\mathbb{Z}_{\geq 0}$, then a graph on s vertices is constructible using the pot P. If $k \notin 2\mathbb{Z}_{\geq 0}$, then a graph on s vertices cannot be constructed using P and this algorithm.

2. Place k vertices, preferably in a circular layout.

3. Assign t_1 to any vertex. To each adjacent vertex, assign t_3. Continue alternating the assignment of t_1 and t_3 in this manner until each vertex is assigned a tile.

4. Join all of the arms of each t_3 with an arm from an adjacent t_1 until every t_3 has no free arms.

5. Each t_1 will have $m - n$ remaining arms. Join these arms with $m - n$ copies of t_2 to form a complete complex.

We provide an example for how to use this algorithm.

Example 4.24. Let $P = \{\{a^6\}, \{\hat{a}\}, \{\hat{a}^4\}\}$. By the algorithm, we can build a connected graph on 12 vertices since

$$k = \frac{12}{1 + \frac{6}{2} - \frac{4}{2}} = 6 \in 2\mathbb{Z}_{\geq 0}.$$

Thus we start by placing 6 vertices in a circular layout (see Figure 4.12).

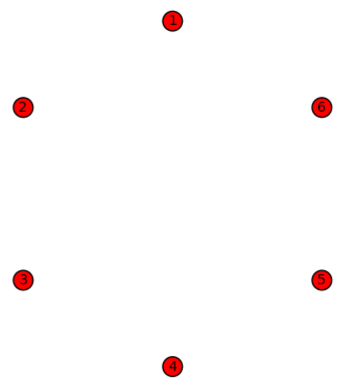


Figure 4.12: Algorithm Step 2

Assign $t_1 = \{a^6\}$ to vertices 1, 3 and 5 and assign $t_3 = \{\hat{a}^4\}$ at vertices 2, 4 and 6 to obtain the graph in Figure 4.13.

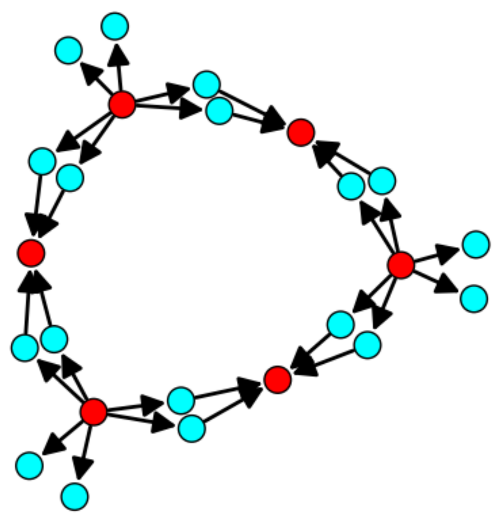


Figure 4.13: Algorithm Step 3

Join all of the t_3 arms with adjacent t_1 arms to obtain the graph in Figure 4.14.

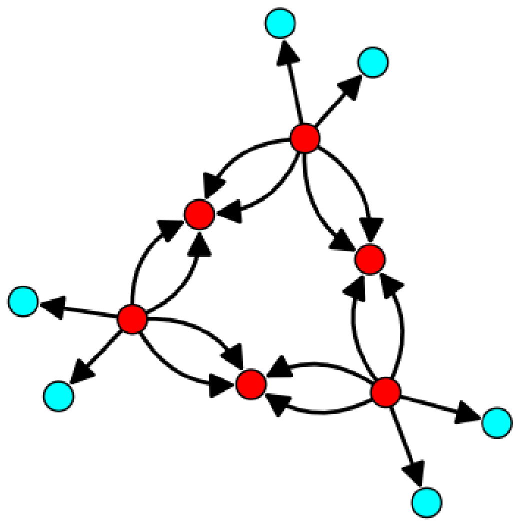


Figure 4.14: Algorithm Step 4

Finally, join any remaining t_1 arms with t_2 tiles to obtain the finished graph in Figure 4.15.

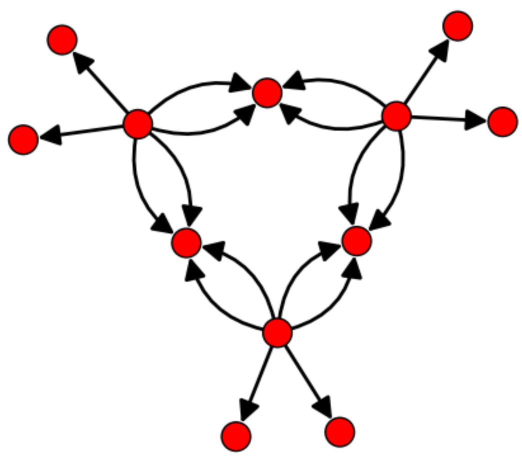


Figure 4.15: Algorithm Step 5

The preceeding discussion and algorithm lead to the following results.

Theorem 4.25. If G is a connected graph of order $k(1+\frac{m}{2}-\frac{n}{2})$ or $k(1+\frac{m}{2}-\frac{n}{2})+(n-1)$, where k comes from Algorithm 4.23, then $G \in \mathcal{O}(P)$.

Proof. The proof is immediate from the algorithm. $\square$

Theorem 4.26. Suppose P realizes a connected graph G of order s. Then for each $j \in \mathbb{Z}_{\geq 0}$, there is a pot $P_j = \{\{a^{m+j}\}, \{\hat{a}\}, \{\hat{a}^{n+j}\}\}$ such that P_j realizes a connected graph G_j of order s with $E(G) \subset E(G_j)$.

Proof. The proof is immediate from the algorithm. $\square$

In general, there can be many distinct connected graphs G of a fixed order where $G \in \mathcal{O}(P)$. Algorithm 4.23 provides a method for finding one such graph.

Chapter 5

Conclusion

We have shown that, given a pot of tiles with one bond-edge type and a 1-armed tile, we can determine the sizes of the complete complexes that can be realized by the pot. To a lesser extent, we can also characterize whether these complete complexes will be disconnected or connected complexes. At this time, the entire case involving a $2 \times n$ construction matrix is close to being completely understood. Two primary questions remain to be explored:

1. What is a formula for ζ in terms of m and n?
2. Do these results extend to pots of the form $P = \{\{a^m\}, \{\hat{a}^j\}, \{\hat{a}^n\}\}$ where $1 < j < m$?

Although the first question remains open, our results provide a lower bound which is "close" to ζ. This means that, for any pot P satisfying the relatively prime condition, there are only finitely many orders to check between the order of the minimal graph of P and the corresponding η.

With the second question, we have some indication that the results in this thesis extend nicely to pots that do not possess a 1-armed tile, but more research is needed in this area. Considering the nice conditions that occur when $\gcd(m+1, -n+1) = d \neq 1$, it would be reasonable to start in this setting rather than the relatively prime setting.

The difficulty of determining $G \in \mathcal{O}(P)$ increases dramatically when moving from pots with one bond-edge type to pots with two bond-edge types. In fact, preliminary research suggests virtually none of the results here generalize to the two bond-edge type case.

Bibliography

[Bur12] David M. Burton. *Elementary Number Theory.* McGraw-Hill, 2012.

[CZ17] Gary Chartrand and Ping Zhang. *A First Course in Graph Theory.* Dover, 2017.

[EMJP19] Joanna Ellis-Monaghan, Nataša Jonoska, and Greta Pangborn. Tile-based DNA nanostructures: mathematical design and problem encoding. In *Algebraic and combinatorial computational biology*, Math. Sci. Eng., pages 35–60. Academic Press, London, 2019.

[EMPB+14] Joanna Ellis-Monaghan, Greta Pangborn, Laura Beaudin, David Miller, Nick Bruno, and Akie Hashimoto. Minimal tile and bond-edge types for self-assembling dna graphs. In *Discrete and Topological Models in Molecular Biology*, pages 241–270. Springer, 2014.

[JMS11] Nataša Jonoska, Gregory L McColm, and Ana Staninska. On stoichiometry for the assembly of flexible tile DNA complexes. *Nat. Comput.*, 10(3):1121–1141, 2011.

[See07] Nadrian C Seeman. An overview of structural dna nanotechnology. *Molecular biotechnology*, 37(3):246, 2007.