8-2023

DNA SELF-ASSEMBLY OF TRAPEZOHEDRAL GRAPHS

Hytham Abdelkarim
California State University - San Bernardino

DNA Self-Assembly of Trapezohedral Graphs

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

Hytham Abdelkarim

August 2023

DNA Self-Assembly of Trapezohedral Graphs

A Thesis

Presented to the

Faculty of

California State University,

San Bernardino

---

by

Hytham Abdelkarim

August 2023

Approved by:

Dr. Cory Johnson, Committee Chair

Dr. Jeremy Aikin, Committee Member

Dr. Min-Lin Lo, Committee Member

Dr. Madeleine Jetter, Chair, Department of Mathematics

Dr. Corey Dunn, Graduate Coordinator

Abstract

Self-assembly is the process of a collection of components combining to form an organized structure without external direction. DNA self-assembly uses multi-armed DNA molecules as the component building blocks. It is desirable to minimize the material used and to minimize genetic waste in the assembly process. We will be using graph theory as a tool to find optimal solutions to problems in DNA self-assembly. The goal of this research is to develop a method or algorithm that will produce optimal tile sets which will self-assemble into a target DNA complex. We will minimize the number of tile and bond-edge types needed to assemble a DNA complex with the same structure as the trapezohedral graphs under different laboratory constraints.

Acknowledgements

First, I would like to express my gratitude to all of the professors who have taught me during this incredible journey. I would like to thank my committee members, Dr. Jeremy Aikin and Dr. Min-Lin Lo for their time and making sure this paper was perfectly written. Also, thank you Dr. Aikin for introducing me to graph theory. My advisor, Dr. Cory Johnson, thank you for helping me throughout this journey. I will forever be grateful for your support. I would like to thank my family for supporting me since I first staring my journey as a student. To my wife, thank you for being my number one fan and cheering me up when I felt down. Thank you for sacrificing your time and believing I could achieve this degree. To my son, thank you for being the reason for my success.

Table of Contents

List of Figures

Chapter 1

Introduction

1.1 Introduction

Self-assembly is the process of a collection of components combining to form an organized structure without external direction. DNA self-assembly uses multi-armed DNA molecules as the component building blocks and this process has evolved since Nadrian Seeman's lab developed the process 40 years ago. It is desirable to minimize the material used and to minimize genetic waste in the assembly process. In this paper, we will be using graph theory as a tool to find optimal solutions to problems in DNA self-assembly. The goal of this research is to develop a method or algorithm that will produce optimal tile sets which will self-assemble into a target DNA complex. Since producing synthetic DNA is costly, the goal is to minimize the number of tile and bond-edge types needed to assemble a DNA complex with the same structure as trapezohedral graphs. We will be studying trapezohedral graphs for our research to find minimal optimal tile sets in different scenarios which will be described in Section 1.5. Figure 1.1 shows the trapezohedral graph when $n = 4$ and $n = 5$, respectively.

1.2 DNA Self-Assembly

As mentioned in [EMPB+14], DNA self-assembly has several applications in the area of nanotechnology, including for drug delivery and biomolecular computing. An essential step in building the self-assembling nanostructure is designing the component molecular building blocks, and moreover determining where in the final structure they

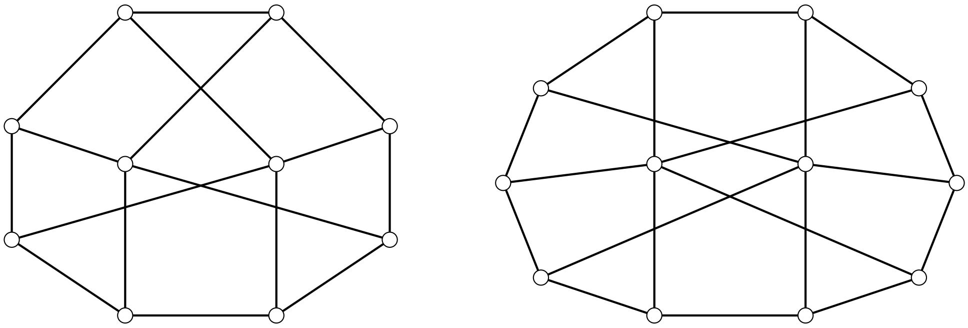

Figure 1.1: An n-trapezohedral graph when $n = 4$ (left) and $n = 5$ (right).

will appear. Many structures being built are essentially graphs, so these design strategy problems fall naturally into the realm of graph theory [EMP11].

1.3 Tools for DNA Self-Assembly

DNA is a molecule twisted into a shape that is known as a double helix. This double helix consists of four nitrogen bases, adenine (A), thymine (T), guanine (G), cytosine (C). Using Watson-Crick pairing, A forms a base pair with T and C forms a base pair with G. Figure 1.2 shows how these bases pair with each other. Different techniques

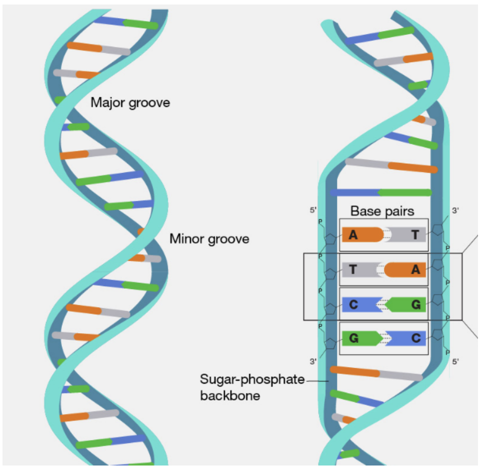

Figure 1.2: Watson-Crick pairing [Bat].

have been developed using the Watson-Crick complementarity properties of DNA strands to achieve self-assembly [EMP11]. There are many DNA self-assembly models but for our research we will be using the flexible tile model. A construction method for DNA self-assembly uses k-armed branched junction molecules which are star-shaped molecules whose centers form the vertices of the structure and the arms are multi strands of DNA with one strand extending beyond the other [EMP11]. There is a longer strand which forms a *cohesive-end* at the end of the arm that can bond to any other cohesive-end with complementary Watson-Crick bases [EMP11]. In the flexible tile model, arms can bond and reach in different directions; whereas rigid tiles have geometric constraints. The cohesive-ends that have bonded with each other then form the edges of our graph. We can see an 8-arm DNA branched junction molecule in Figure 1.3. A DNA complex is said to be complete if it has no unmatched cohesive-ends.

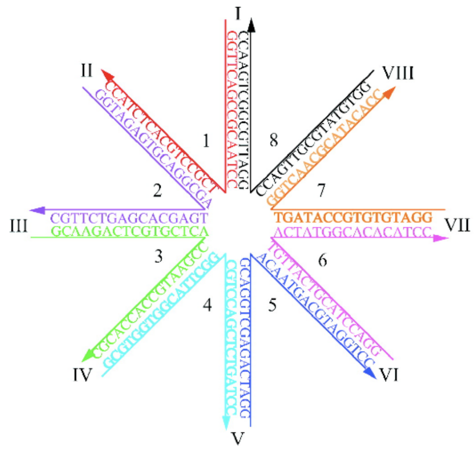

Figure 1.3: 8-arm DNA branched junction molecule [WS07].

1.4 Definitions

The following definitions are consistent with those found in [EMPB+14], [EMJP19], and [AEMH+].

Definition 1.1. *Tile: A tile is a graph-theoretical representation of a flexible k-armed branched DNA molecule as a vertex with k half-edges representing cohesive-ends.*

Definition 1.2. *Cohesive-end type: Cohesive-ends are distinguished by cohesive-end types that are letter labels such that a cohesive-end labeled with an unhatted letter can adjoin to a cohesive-end labeled with its complementary hatted label.*

Definition 1.3. *Bond-edge type: The abstraction of a cohesive-end together with its complementary cohesive-end.*

Definition 1.4. *Pot: A pot P is a collection of tile types such that for each cohesive-end type that appears in any tile, its complement appears in some tile.*

An example of a tile is seen in Figure 1.4. A tile contains a vertex with half edges representing cohesive-ends. We can see in Figure 1.4 a 5-armed tile with five half-edges, (i.e. $\hat{a}, a, a, b, \hat{c}$). A pot P is a set of tiles. If we look at the graph G on the left side in Figure 1.5, there exist two different tile types. We represent these tiles as, $t_1 = \{a^4\}, t_2 = \{\hat{a}^2, c, \hat{c}\}$. In the assembly process, a tile type may be used more than once. The pot P that realizes G is $P = \{t_1 = \{a^4\}, t_2 = \{\hat{a}^2, c, \hat{c}\}\}$.

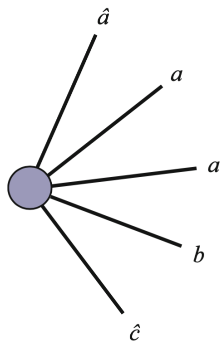

Figure 1.4: A representation of a 5-armed tile [EMP11].

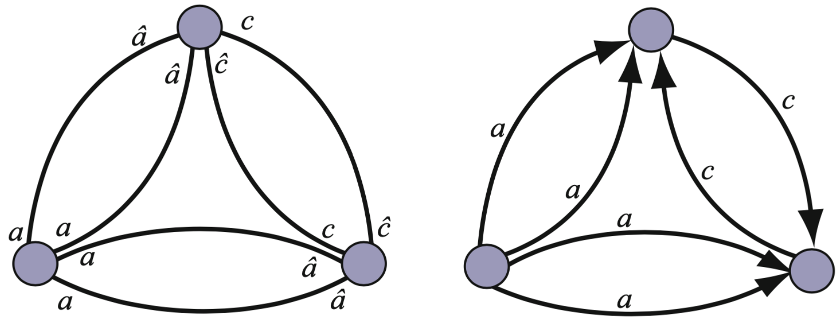

Figure 1.5: Graphs G and H [EMP11].

If we look at graph H on the right side in Figure 1.5, we see that both graphs are similar. We can represent G into H by using a directed graph. The cohesive-end types will be edges going from a to $\hat{a}$ and c to $\hat{c}$. We used two bond-edge types, a and c, and we used two tiles, t_1 and t_2, to construct G.

1.5 The Three Scenarios

The goal of this research is to assist the laboratories in being efficient in the design of self-assembling DNA. Our problem becomes the following: Given a graph G, what is the minimum number of tiles and bond-edge types that must be designed to construct the target graph? We take into account the three scenarios below. These scenarios correspond to how much genetic waste is permitted in the self-assembly process [EMPB+14].

1. *Scenario 1.* We allow the possibility that graph-theoretical complexes of smaller order (that is, complexes representing graphs with fewer vertices) than the target graph could be created from the pot used to build the target graph.

2. *Scenario 2.* We allow the possibility that graph-theoretical complexes with the same number of vertices as, but not isomorphic to, the target graph could be created from the pot that builds the target graph, but require that no complexes with fewer vertices can be created from the pot used to build the target graph.

3. *Scenario 3.* We require that no non-isomorphic complexes with a number of vertices less than or equal to that of the target graph can be created from the pot used to build the target graph.

We will denote $T_i(G)$ where $i = 1, 2, 3$, representing the three scenarios to be the minimum number of tile types needed to construct a graph G. Also, we will denote $B_i(G)$ where $i = 1, 2, 3$, representing the three scenarios to be minimum number of bond-edge types needed to construct G.

1.6 Properties of the Trapezohedral Graph

Recall that a *graph* G is a pair of sets (V, E), where V is a finite non-empty set of elements called *vertices*, and E is a finite set of elements called *edges*, each of which has two associated vertices. We can also write $V(G)$ and $E(G)$ to represent the vertex and edge set of a graph G. The *order*, n, of a graph is the number of vertices and the *size*, m, is the number of edges in the graph. Two vertices that are joined by an edge are said to be *adjacent*. Two graphs, G and H, are said to be *isomorphic* if there exists a bijection $f : V(G) \to V(H)$ such that u and v are adjacent in G if and only if $f(u)$ and $f(v)$ are adjacent in H. An n-trapezohedron is a solid composed of interleaved symmetric quadrilateral kites, n of which meet in a top vertex and n in a bottom vertex. A n-trapezohedral graph is a skeleton of an n-trapezohedron. Since we are using the family of trapezohedral graphs for this research, we need to understand a few of its graph theoretical properties. The trapezohedral graphs are bipartite, Hamiltonian, perfect, planar, and traceable [Wei]. The order of an n-trapezohedral graph is $2(n + 1)$ and the size is $4n$. Trapezohedral graphs can be drawn in a way so that there is an "outer" cycle with two "central" vertices, (see Figure 1.1), so we may use the cycle graph properties for this research. A *cycle graph* is a graph on n vertices containing a single cycle through all vertices. Since trapezohedral graphs are bipartite, the bipartite graph properties may be used as well. A *bipartite graph* is a graph G whose vertices can be divided into two disjoint sets U and W, such that every edge of G joins a vertex of U to a vertex of W. It is worth noting that a graph is bipartite if and only if all cycles are even length.

1.7 Goal for this Research

This research will oversee Scenarios 1 and 2 for the trapezohedral graph. As mentioned, the results for the cycle graph may be used in this research since a cycle of length $2n$ does exist in the trapezohedral graph, see Figure 1.1. A goal of this research is to study the open question of whether there exist relationships between the pots of a graph and known pots of subgarphs. Since all trapezohedral graphs are bipartite graphs, results from bipartite graphs may be used as well. We will use a program from [AEMH$^+$] to check the condition of Scenario 2. The goal is to find an optimal solution for each of the two scenarios. Since this area of research is still new, all results are original.

Chapter 2 describes some known results we will use to answer our research questions. Chapters 3 and 4 describe optimal pots for trapezohedral graphs in Scenarios 1 and 2, respectively.

Chapter 2

Information about DNA Self-Assembly

2.1 Preliminaries

The following propositions and definitions are from [EMPB$^+$14] and [EMJP19].

Proposition 2.1. *For every graph G, $B_1(G) \leq B_2(G) \leq B_3(G)$ and $T_1(G) \leq T_2(G) \leq T_3(G)$*

Proposition 2.1 establishes bounds between the three scenarios for the minimum number of bond-edge types and tile types.

Definition 2.2. *The set of graphs realized by a pot P is called the output of P and is denoted by $\mathcal{O}(P)$.*

We will use tools from linear algebra to determine the smallest order graph realized by a pot. Given a pot $P = \{t_1, \ldots, t_p\}$, let $A_{i,j}$ be the number of cohesive-ends of type a_i on tile t_j and $\hat{A}_{i,j}$ be the number of cohesive-ends of type $\hat{a}_i$ on tile t_j. Since we are requiring complexes to be complete, the following proposition is a result from this requirement.

Proposition 2.3. *Let $P = \{t_1, \ldots, t_p\}$ be a pot. Then:*

1. *The total number of hatted cohesive-end types must equal the total number of unhatted cohesive-end types in a complete complex.*

2. *If a graph G with n vertices may be constructed from the pot P, then there are non-negative integers N_j where $j = 1, \ldots, p$ with $\sum_j N_j = n$ and such that $\sum_j N_j(A_{i,j} - \hat{A}_{i,j}) = 0$ for all i. That is, the number of hatted cohesive-ends of each type used in the construction of G must equal the number of unhatted cohesive-ends of the same type that appear in the construction.*

Now we can define the construction matrix and see how it is applied in Scenario 2.

Definition 2.4. *Let P be a pot with p tile types labeled $t_1, \ldots, t_p$ and let $z_{i,j}$ be the net number of cohesive-ends of type a_i on tile t_j, i.e., $z_{i,j} = A_{i,j} - \hat{A}_{i,j}$. Let r_i be the proportion of tile type t_i used in the assembly process. We can make the following system of equations:*

$$\begin{aligned} r_1 z_{1,1} + r_2 z_{1,2} + \cdots + r_p z_{1,p} &= 0 \\ &\vdots \\ r_1 z_{m,1} + r_2 z_{m,2} + \cdots + r_p z_{m,p} &= 0 \\ r_1 + r_2 + \cdots + r_p &= 1 \end{aligned}$$

The construction matrix of P, denoted by $M(P)$, is the corresponding augmented matrix:

$$M(P) = \begin{bmatrix} z_{1,1} & z_{1,2} & \ldots & z_{1,p} & 0 \\ \vdots & \vdots & & \vdots & \\ z_{m,1} & z_{m,2} & \ldots & z_{m,p} & 0 \\ 1 & 1 & \ldots & 1 & 1 \end{bmatrix}.$$

The following proposition is a result of the construction matrix and Proposition 2.3 [EMPB+14]:

Proposition 2.5. *Let $P = \{t_1, \ldots, t_p\}$ be a pot. Then:*

1. *If a graph G of order n may be constructed from P using N_j tiles of type t_j, then $(1/n)\langle N_1, \ldots, N_p\rangle$ is a solution of the construction matrix $M(P)$.*

2. *If* $\langle r_1, \ldots, r_p \rangle$ *is a solution to the construction matrix* $M(P)$*, and there is a positive integer* n *such that* $nr_j \in \mathbb{Z} \geq 0$ *for all* j*, then there is a graph of order* n *that may be constructed from* P *using* nr_j *tiles of type* t_j*.*

3. *The smallest order of a graph in* $\mathcal{O}(P)$ *is* $m_p = min\{lcm\{b_j | r_j \neq 0$ *and* $r_j = a_j/b_j\}$*, where* $\langle r_1, \ldots, r_p \rangle$ *is a solution to* $M(P)\}$*, and where the minimum is taken over all solutions to* $M(P)$ *such that* $r_j \geq 0$ *and* a_j/b_j *is in reduced from for all* j*.*

Example 2.6. *If we look at the pot in Figure 1.5,* $P = \{t_1 = \{a^4\}, t_2 = \{\hat{a}^2, c, \hat{c}\}\}$*. In this case we have:*

$$z_{1,1} = 4, z_{1,2} = -2, z_{2,1} = 0, z_{2,2} = 0$$

and

$$M(P) = \begin{bmatrix} 4 & -2 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}.$$

By doing reduced row echelon form, we get the following matrix:

$$\begin{bmatrix} 1 & 0 & 1/3 \\ 0 & 1 & 2/3 \\ 0 & 0 & 0 \end{bmatrix}.$$

The solution for this matrix is $\langle \frac{1}{3}, \frac{2}{3} \rangle$*. As mentioned in [EMJP19], we can use the spectrum, which is the solution space of* $M(P)$ *and is denoted by* $\mathcal{S}(P)$*. We can write the set of all solutions to* $M(P)$ *as* $\mathcal{S}(P) = \{\frac{1}{r}\langle \frac{r}{3}, \frac{2r}{3} \rangle | r \in \mathbb{Z}^+\}$*. This means the smallest graph we can build is a graph on three vertices, one vertex labeled* t_1 *and two vertices labeled* t_2*. Hence one graph realized by* P *is in Figure 1.5 but we can also get a non-isomorphic graph in which the cohesive-ends* c *and* $\hat{c}$ *on* t_2 *form a loop. Notice the pot* P *will also realize graphs of order* $3r$ *which may or may not be connected.*

2.2 Background Information for Scenario 1

In Scenario 1, we allow the possibility for complexes of smaller order to be created from a pot that builds our target graph. We will discover pots with the fewest number of tile and bond-edge types. Now we will define a few terms that will help us in Scenario 1. The *valency sequence* of G is the sequence of vertex degrees of G without repeats and the length of the sequence is denoted by $av(G)$. The *even-valency sequence* is the sequence of even degrees and the length of the sequence is denoted by $ev(G)$. The *odd-valency sequence* is the sequence of odd degrees and the length of the sequence is denoted by $ov(G)$. The following theorem, corollary, and propositions are from [EMPB$^+$14]. The following theorem gives us bounds on the minimum number of tile types in Scenario 1.

Theorem 2.7. *For all G, $av(G) \leq T_1(G) \leq ev(G) + 2ov(G)$.*

The following corollary tells us how many bond-edge types are needed for any graph G in Scenario 1,

Corollary 2.8. *$B_1(G) = 1$, for all G.*

2.3 Background Information for Scenario 2

In Scenario 2, we allow the possibility that complexes the same order as, but not isomorphic to, the target graph could be created from the pot that builds the target graph G, but require that no smaller complexes can be built from the pot. We will look for pots P with the least tile or bond-edge types such that $G \in \mathcal{O}(P)$. The following theorem is from [EMPB$^+$14]. The theorem describes a relationship between the minimum number of tiles and bond-edge types in a graph G in Scenario 2.

Theorem 2.9. *If G is a graph with $n > 2$ vertices, then $B_2(G) + 1 \leq T_2(G)$.*

In Scenario 2, we explored the cycle graph properties in the hopes of finding a minimal optimal pot for the trapezohedral graph. From [EMPB$^+$14], we find $B_2(C_n) = \left\lceil \frac{n}{2} \right\rceil$ and $T_2(C_n) = \left\lceil \frac{n}{2} \right\rceil + 1$. It is currently an open question if there are any kinds of relationships of pots of graphs G and pots of subgraphs H.

In the trapezohedral graph, only an even cycle appears in the graph. The pot of for an even cycle is:

$$P_{\text{even}} = \left\{ t_1 = \{a_1^2\}, t_i = \{\hat{a}_{i-1}, a_i\} \text{ for } i = 2, \ldots, \left\lceil \frac{n}{2} \right\rceil, t_{\lceil n/2 \rceil + 1} = \{\hat{a}^2_{\lceil n/2 \rceil}\} \right\} \tag{2.1}$$

Example 2.10. *In Figure 2.1, we have a cycle graph on 8 vertices.*

If we were to apply Equation 2.1 on 8 vertices, we will get the following pot,

$$P = \left\{ t_1 = \{a_1^2\}, t_2 = \{\hat{a}_1, a_2\}, t_3 = \{\hat{a}_2, a_3\}, t_4 = \{\hat{a}_3, a_4\}, t_5 = \{\hat{a}_4^2\} \right\}.$$

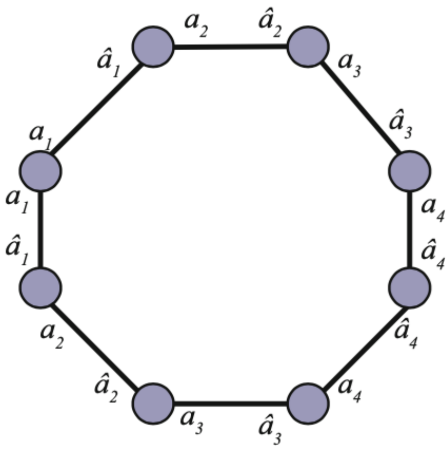

Figure 2.1: A cycle graph on 8 vertices [EMPB$^+$14].

Now by using Definition 2.4, we are able to build the construction matrix:

$$M(P) = \begin{bmatrix} 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -2 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

By using the program from [AEMH$^+$], we get a unique matrix solution, $\langle \frac{1}{8}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{8} \rangle$. The spectrum is $\mathcal{S}(P) = \{\frac{1}{r}\langle \frac{r}{8}, \frac{r}{4}, \frac{r}{4}, \frac{r}{4}, \frac{r}{8} \rangle | r \in \mathbb{Z}^+\}$. This means the smallest graph we can build is a graph of order eight, one vertex labeled t_1, two vertices labeled t_2, two vertices labeled t_3, two vertices labeled t_4, and one vertex labeled t_5. Hence, we get the graph in Figure 2.1.

Chapter 3

Trapezohedral Graph Under Scenario 1

Before we start, there are a few things to point out. In both scenarios, we will consider the trapezohedral graph when $n \geq 4$. Looking at the n-trapezohedron in a three-dimensional space we realize n must be at least equal to three. For $n = 3$, this is the cubic graph and a pot P has been found in all scenarios [AEMH$^+$]. In Scenario 1, we will be using corollaries and theorems from Section 2.2 to help build our pot. As a reminder, we allow the possibility that complexes of smaller order than the target graph could be created from the pot used to build the target graph.

Lemma 3.1. *If G is a trapezohedral graph, then $B_1(G) = 1$.*

Proof. $B_1(G) = 1$ follows directly from Corollary 2.8. □

Theorem 3.2. *If G is a trapezohedral graph, then $T_1(G) = 2$.*

Proof. Theorem 2.7 determines the minimum and maximum number of tile types needed. Since $av(G) = 2$, then $T_1(G) \geq 2$ for all n. Let $P = \left\{ t_1 = \{a, \hat{a}^2\}, t_2 = \{a^n\} \right\}$. Then $2n(A_{1,1} - \hat{A}_{1,1}) + 2(A_{1,2} - \hat{A}_{1,2}) = 0$; that is, $2n(-1) + 2(n) = 0$. All of the cohesive-ends are matched and the number of unhatted cohesive-ends is equal to the number of hatted cohesive-ends. From Figure 3.1, we see $G \in \mathcal{O}(P)$. Thus, $T_1(G) = 2$. □

Below is a directed trapezohedral graph using the pot P from Scenario 1 for any $n \geq 4$.

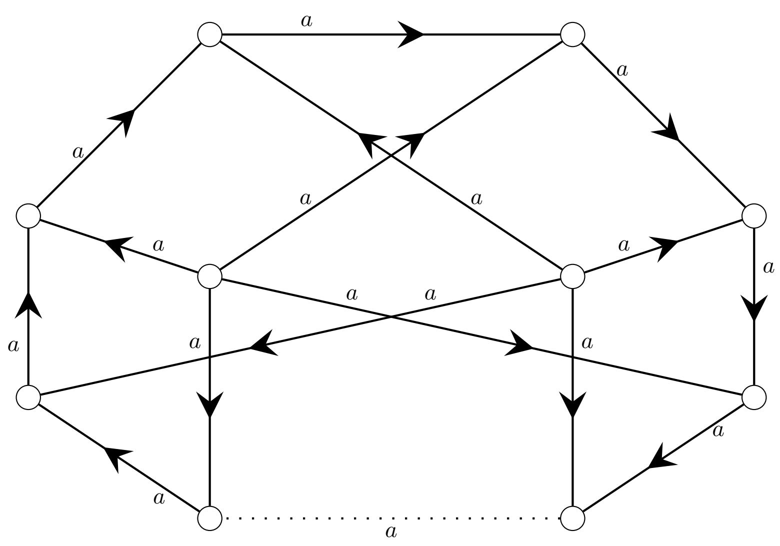

Figure 3.1: An n-trapezohedral graph using P from Scenario 1.

Chapter 4

Trapezohedral Graph Under Scenario 2

In Scenario 2, we allow the possibility that non-isomorphic complexes of the same order but no smaller complexes than the target graph G may be realized by the pot. We will look for pots P with the least tile or bond-edge types such that $G \in \mathcal{O}(P)$. We will use theorems and propositions from Section 2.3 to help us with our pot in Scenario 2. Looking at the trapezohedral graph, we can see a subgraph within our graph is the cycle graph. We explore the idea of using the cycle subgraph to find a pot in Scenario 2.

If we look at a trapezohedral graph when $n = 4$, then we obtain a cycle subgraph on 8 vertices. By using pot P from Equation 2.1, we have the following pot for the cycle subgraph,

$$P' = \left\{ t_1 = \{a^2\}, t_2 = \{\hat{a}, b\}, t_3 = \{\hat{b}, c\}, t_4 = \{\hat{c}, d\}, t_5 = \{\hat{d^2}\} \right\}.$$

Now, adding a tile type to P' that represents the "central" vertices, and adding edges from the 8-cycle to the central vertices, we have the following pot,

$$P = \left\{ t_1 = \{a^2, \hat{a}\}, t_2 = \{\hat{a}^2, b\}, t_3 = \{\hat{a}, \hat{b}, c\}, t_4 = \{\hat{a}, \hat{c}, d\}, t_5 = \{\hat{a}, \hat{d^2}\}, t_6 = \{a^4\} \right\}.$$

Note the pot P still uses only four bond-edge types. Using Definition 2.4, we are able to build the construction matrix:

$$M(P) = \begin{bmatrix} 1 & -2 & -1 & -1 & -1 & 4 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -2 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

The spectrum is $\mathcal{S}(P) = \{\frac{1}{16r}\langle 9r - 37t, 2r + 6t, 2r + 6t, 2r + 6t, r + 3t, t\rangle | r \in \mathbb{Z}^+, t \in \mathbb{Z} \cap [\frac{-r}{3}, \frac{9r}{37}]\}$. By using the program from [AEMH$^+$], the smallest order graph is 10 and the ratio of tile types is $\langle \frac{1}{10}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{10}, \frac{1}{5}\rangle$. This means the smallest graph we can build is a graph on ten vertices, one vertex labeled t_1, two vertices labeled t_2, two vertices labeled t_3, two vertices labeled t_4, one vertex labeled t_5, and two vertices labeled t_6. Hence, the trapezohedral graph in Figure 4.1 is realized by P.

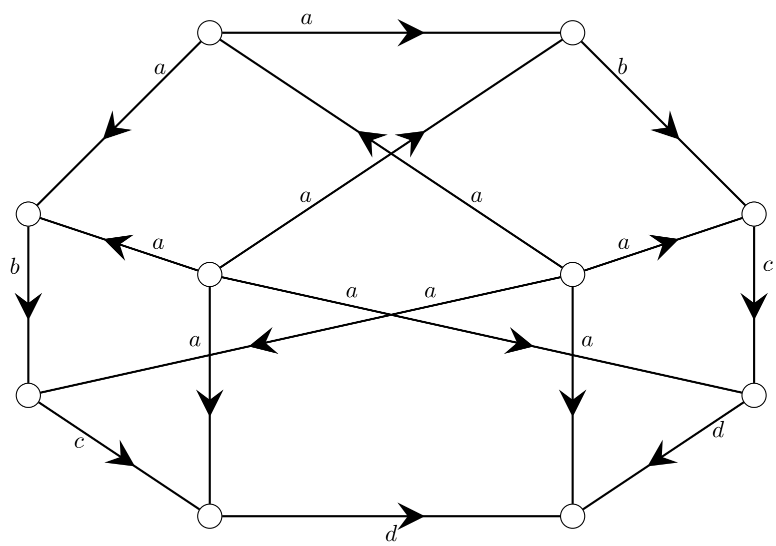

Figure 4.1: An 4-trapezohedral graph using P from Scenario 2.

Although the pot P from above does realize G, we will show it is not the optimal pot in Scenario 2. In fact, the following theorem shows that two bond-edge types is sufficient for all n-trapezohedral graphs when $n \geq 4$.

Theorem 4.1. *Let n be odd. If G is a n-trapezohedral graph, then $B_2(G) = 2$.*

Proof. Assume $B_2(G) = 1$ and let P be a pot with exactly one bond-edge type where $G \in \mathcal{O}(P)$. If we look at the "outer" cycle of a trapezohedral graph, we notice all vertices are of degree three. Here, there are only four possible tile types which are $t_1 = \{a^3\}, t_2 = \{a^2, \hat{a}\}, t_3 = \{a, \hat{a}^2\}$, and $t_4 = \{\hat{a}^3\}$. Notice t_1 and t_4 are not both in P since they are complements of each other as well as t_2 and t_3; that is, the two tile types may realize a graph of order two. If using t_1 and t_3 in P for the "outer" cycle then P can realize a graph of order four which is not permitted in Scenario 2. We can say the same if t_2 and t_4 are both in P. Since the case in which $t_4 \in P$ is analogous, we may assume without loss of generality that $t_1 \in P$. Then either $t_2 \in P$ or $t_3 \in P$ since an edge on the "outer" cycle must be labeled by a and $\hat{a}$. Without loss of generality, since we eliminated the case where $t_1, t_3 \in P$ then we can assume $t_2 \in P$. The "central" vertex must be labeled by $t_3 = \{a^j, \hat{a}^k\} \in P$ where $0 \leq j < k \leq n$ and $j + k = n$ in order for there to be no unmatched cohesive-ends. But then a graph of order $(k - j) + 1$ can be realized using $(k - j)$ copies of t_2 and one copy of t_3. Since $k \leq n$, then $k - j + 1 \leq n + 1 < 2(n + 1)$. So the graph has a smaller order than G which is not permitted in Scenario 2. Thus, $B_2(G) \geq 2$. The following pot can realize a graph G in Scenario 2 using two bond-edge types:

$$P_{odd} = \{t_1 = \{a^2, \hat{b}\}, t_2 = \{\hat{a}^2, \hat{b}\}, t_3 = \{b^n\}\}. \tag{4.1}$$

To see that nothing smaller can be made from this pot, note that

$$M(P_{odd}) = \begin{bmatrix} 2 & -2 & 0 & 0 \\ -1 & -1 & n & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

has a unique solution of the form $\langle \frac{n}{2(1+n)}, \frac{n}{2(1+n)}, \frac{1}{1+n} \rangle$. Since n is odd then $\frac{n}{2(1+n)}$ is in reduced form. By Proposition 2.5, the smallest order graph is $2(1 + n)$ which is the order of G. □

Corollary 4.2. *Let n be odd. If G is a n-trapezohedral graph, then $T_2(G) = 3$.*

Proof. Since $B_2(G) = 2$ then $B_2(G) + 1 \leq T_2(G)$. So $3 \leq T_2(G)$. The pot in Theorem 4.1 has exactly three tile types, therefore $T_2(G) = 3$. □

The pot in Equation 4.1 will not work when n is even because a graph of a smaller order will be realized by P_{odd}. For example, let $n = 4$, then the solution to the construction matrix will be $\langle \frac{2}{5}, \frac{2}{5}, \frac{1}{5} \rangle$. A graph of order five can be realized when $n = 4$ but our target graph is of order ten. As a result, we will need a new pot when n is even.

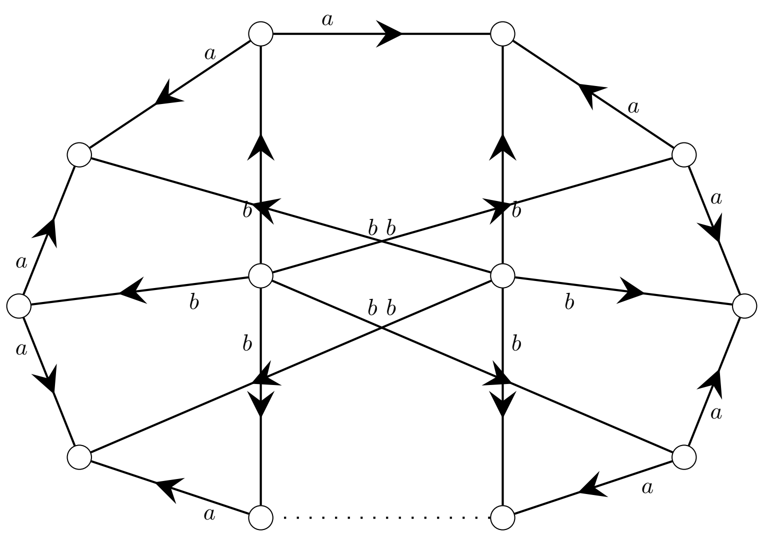

Figure 4.2: An n-trapezohedral graph using P_{odd} from Scenario 2.

Theorem 4.3. *Let n be even. If G is a n-trapezohedral graph, then $B_2(G) \leq 3$ and $T_2(G) \leq 4$.*

Proof. The following pot P can realize a graph G in Scenario 2 using three bond-edge types and four tile types:

$$P_{even} = \{t_1 = \{a^2, b\}, t_2 = \{\hat{a}^2, c\}, t_3 = \{\hat{b}^n\}, t_4 = \{\hat{c}^n\}\}. \tag{4.2}$$

To see that nothing smaller can be made from this pot, note that

$$M(P_{even}) = \begin{bmatrix} 2 & -2 & 0 & 0 & 0 \\ 1 & 0 & -n & 0 & 0 \\ 0 & 1 & 0 & -n & 0 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

has a unique solution of the form $\langle \frac{n}{2(1+n)}, \frac{n}{2(1+n)}, \frac{1}{2(1+n)}, \frac{1}{2(1+n)} \rangle$. By Proposition 2.5, the smallest order graph is $2(1+n)$ which is of the order G. Therefore $B_2(G) \leq 3$ and $T_2(G) \leq 4$. □

When n is even then $B_2(G) \geq 2$ by the proof of Corollary 4.2 and thus, $T_2(G) \geq 3$ by Theorem 2.9. Therefore, when n is even, $2 \leq B_2(G) \leq 3$ and $3 \leq T_2(G) \leq 4$. The question of equality for $B_2(G)$ and $T_2(G)$ in the even case remains open.

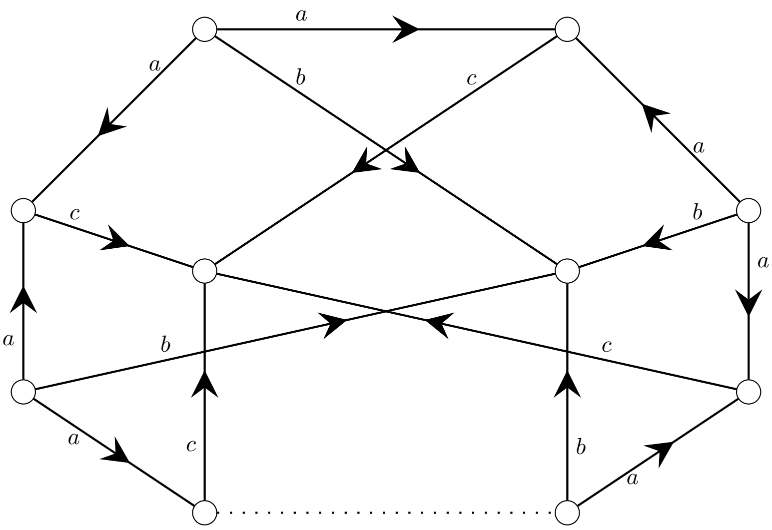

Figure 4.3: An n-trapezohedral graph using P_{even} from Scenario 2.

In both the case when n is even and n is odd, we again notice that the optimal pot for the trapezohedral graph is independent from the optimal pot for a cycle graph. While some results use cycle subgraphs to help find optimal pots [GS23], this strategy does not work for all graphs with a cycle subgraph. Thus, in general, there may be no relationship between optimal pots for a graph G and a subgraph H.

Chapter 5

Conclusion

This paper explored modeling the DNA self-assembley of complexes with the same structure as the n-trapezohedral graph using flexible k-armed tiles. With a goal of minimizing lab costs, we found optimal pots of tiles in Scenarios 1 and 2. The difference between these scenarios is a question of how much byproduct would a lab want to allow in the assembly process. In Scenario 2, we found an optimal pot of tiles in the case that n is odd, but this pot can construct a smaller order graph if n is even. In the case that n is even, we have a pot of tiles that will realize the n-trapezohedral graph but it remains to be shown if the number of bond-edge types and tile types can each be reduced by one. In general, relationships between optimal pots for subgraphs and graphs is an open question, but our work found that is it possible for $T_2(G) < T_2(H)$ and $B_2(G) < B_2(H)$ where H is a proper subgraph of G. We leave finding an optimal pot in Scenario 3 as an open question.

Bibliography

[AEMH+] Leyda Almodóvar, Jo Ellis-Monaghan, Amanda Harsy, Cory Johnson, and Jessica Sorrells. Computational complexity and pragmatic solutions for flexible tile-based dna self-assembly. Under review, http://arxiv.org/abs/2108.00035.

[Bat] Sarah A. Bates. Deoxyribonucleic acid (dna). https://www.genome.gov/genetics-glossary/Deoxyribonucleic-Acid. Accessed: 2023-03-12.

[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.

[EMP11] J. Ellis-Monaghan and G. Pangborn. Using DNA self-assembly design strategies to motivate graph theory concepts. *Math. Model. Nat. Phenom.*, 6(6):96–107, 2011.

[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*, Nat. Comput. Ser., pages 241–270. Springer, Heidelberg, 2014.

[GS23] Chloe Griffin and Jessica Sorrells. Tile-based modeling of dna self-assembly for two graph families with appended paths. *Involve, a Journal of Mathematics*, 16(1):69–106, 2023.

[Wei] Eric W. Weisstein. Trapezohedral graph. https://mathworld.wolfram.com/TrapezohedralGraph.html. Accessed: 2022-08-30.

[WS07] Xing Wang and Nadrian C. Seeman. Assembly and characterization of 8-arm and 12-arm dna branched junctions. *Journal of the American Chemical Society*, 129(26):8169–8176, 2007. PMID: 17564446.