Rose-Hulman Undergraduate Mathematics Journal

Volume 21
Issue 1

Article 11

DNA Self-Assembly Design for Gear Graphs

Chiara Mattamira
Converse College, cmattamira001@converse.edu

DNA Self-Assembly Design for Gear Graphs

Cover Page Footnote

Support for this research was provided by South Carolina IDeA Networks of Biomedical Research Excellence (SC INBRE) grant P20GM103499 and by South Carolina Independent Colleges and Universities (SCICU). Thanks to Converse College for the use of its facility and to my research advisor Dr. Jessica Williams for her help and support.

ℝose-Hulman Undergraduate Mathematics Journal
Volume 21, Issue 1, 2020

DNA Self-Assembly Design for Gear Graphs

By *Chiara Mattamira*

Abstract. Application of graph theory to the well-known complementary properties of DNA strands has resulted in new insights about more efficient ways to form DNA nanostructures, which have been discovered as useful tools for drug delivery, biomolecular computing, and biosensors. The key concept underlying DNA nanotechnology is the formation of complete DNA complexes out of a given collection of branched junction molecules. These molecules can be modeled in the abstract as portions of graphs made up of vertices and half-edges, where complete edges are representations of double-stranded DNA pieces that have joined together. For efficiency, one aim is to minimize the number of different component molecules needed to build a nanostructure. Previously known flexible strand model results include optimal construction solutions for cycles, trees, complete graphs, and complete bipartite graphs. In this work, we provide results for all sizes of gear graphs within the context of three different restrictive conditions.

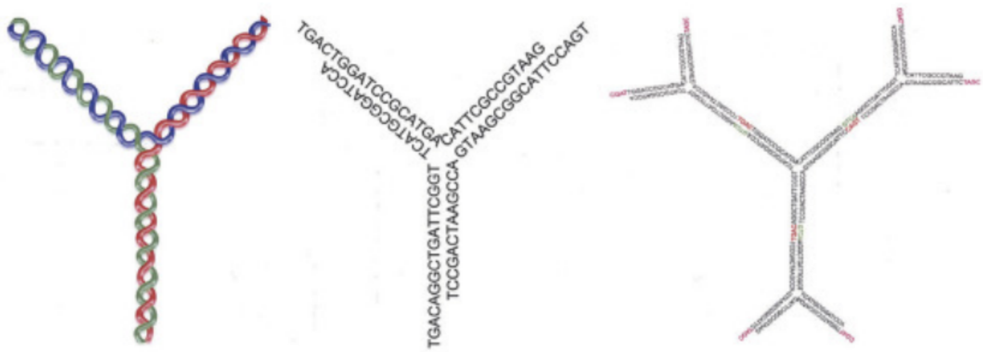
1 Introduction

1.1 Introduction to DNA Self-Assembly

DNA self-assembly is the term used to describe the formation, without external direction, of a collection of DNA molecules into a larger structure. These nanostructures are useful for a variety of applications, including targeted drug delivery, biomolecular computing, biosensors, and the transport and release of nanocargos [3, 4, 9, 5, 6]. Laboratories aim to construct DNA complexes in an efficient and cost-effective way; frequently this translates to using the smallest possible number of different component molecules. One method for modeling DNA nanostructures and the self-assembly processes is to represent the desired structures as discrete graphs. In this work we use a graph theoretical flexible tile model introduced in [2]. In this model, self-assembling DNA structures form from star-shaped molecules with k double-stranded DNA arms called k-armed branched-junction molecules (see Figure 1).

Mathematics Subject Classification. 05C90, 92D20
Keywords. DNA, graph theory

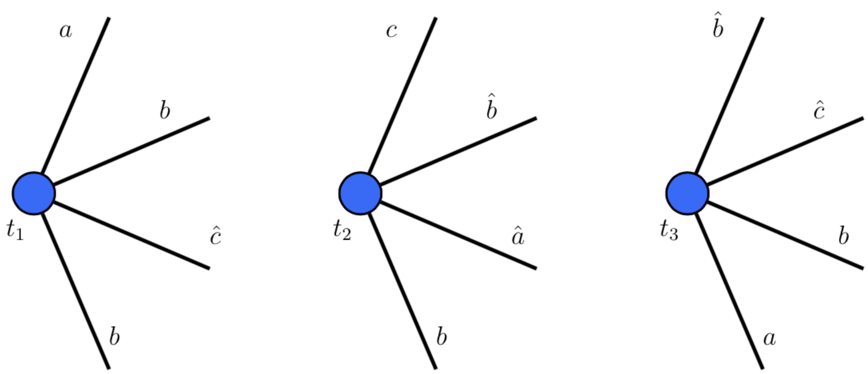
Figure 1: k-armed branched junction molecules [4]



The k arms have cohesive ends which can bond to any other cohesive end with a complementary sequence of Watson-Crick bases and form a variety of DNA nanostructures. It is assumed that the arms are long enough to adhere together. When designing the component molecules needed to construct a DNA complex, we require that the resulting structure is complete, meaning there are no unmatched cohesive ends.

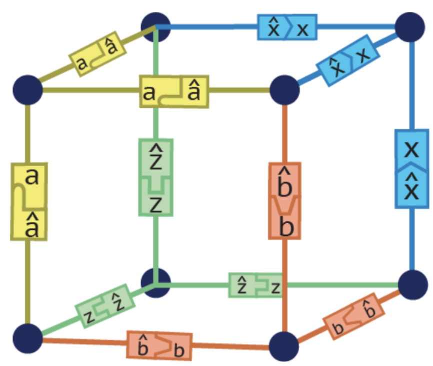
The work done in [2] also introduces the notions of tile and pot on which we base our work. The graph theoretical abstraction of a k-armed branched junction molecule is called a *tile*. A collection of tiles is called a *pot* (see Figure 2).

Figure 2: Pot of three tiles



The abstraction of complementary cohesive ends bonded together is called a *bond-edge* as defined in [1]. We represent the bond-edge types by letters, such that a cohesive end labeled with an "unhatted" letter can join to a cohesive end labeled with its complementary "hatted" label (see Figure 3). The *tile type* is the multiset of letters corresponding to the cohesive-end types for the tile [1]. Since we use the flexible tile model, we do not distinguish between permutations of cohesive-end labels about a vertex. For this reason we can identify a tile with the multiset of its cohesive-end labels. For example, the pot of tiles shown in Figure 2 is denoted as $P = \{t_1 = \{a, b^2, \hat{c}\}, t_2 = \{\hat{a}, b, \hat{b}, c\}, t_3 = \{a, b, \hat{b}, \hat{c}\}\}$, where the exponent indicates a repeat in bond-edge type.

Figure 3: Labeling of a complete graph



In order to efficiently construct a given target graph G, we must find the minimum number of tiles types and bond-edge types that must be included. We consider this question in three scenarios, representing three levels of restriction [1]:

- Scenario 1. The incidental construction of a graph smaller than G is acceptable.
- Scenario 2. The incidental construction of a graph smaller than G is not acceptable, but a graph with the same size as G is acceptable.
- Scenario 3. Any graph incidentally constructed must be larger than G.

We let $T_i(G)$ for $i = 1, 2, 3$ denote the minimum number of tiles required to construct a complex in each of the scenarios above. Similarly, $B_i(G)$ denotes the minimum number of bond-edge types needed for each scenario. Currently known results include optimal solutions for cycles, trees, complete graphs and complete bipartite graphs for all scenarios [1].

Definition. Given a pot P, we define C(P) to be the set of graphs that can be constructed from P. The set of graphs of minimum size that may be constructed from P is denoted $C_{min}(P)$ [1].

In order to prove that a pot P does not create graphs smaller than the target graph, we often use a matrix derived from a system of equations that any pot of tiles forming complete complexes must satisfy.

We follow the notation from [1]. Given a pot $P = \{t_1, \ldots t_p\}$, we 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. Let $z_{i,j} = \mathrm{A}_{i,j} - \hat{\mathrm{A}}_{i,j}$ and let r_i be the proportion of tile type t_i used in the assembly process. The relationship between $z_{i,j}$ and r_i can be described with the following equations:

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

From these equations we can write the *construction matrix* M(P) for the pot P:

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

Note that this is the augmented matrix formed from the system of equations above. Solutions to this matrix give the proportion of tile types needed to construct a complete graph G, which allow us to determine the size of the smallest graph that may be constructed from P [1].

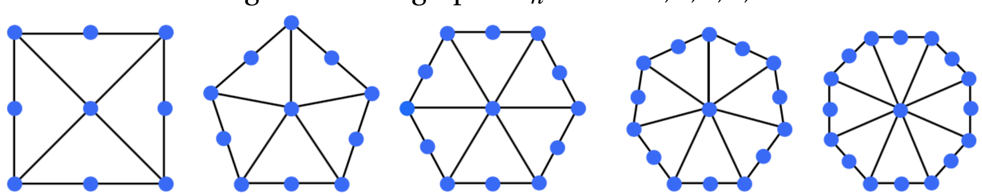
1.2 Gear Graphs

In this work we build upon previously known graph theoretical methods for optimizing the self-assembly process to find minimum numbers of tile and bond-edge types for the family of gear graphs.

Definition. A *wheel graph*, denoted W_n, is a graph formed by connecting a single universal vertex to all vertices of the cycle C_n.

Definition. A *gear graph*, denoted G_n, is a graph obtained by inserting an extra vertex between each pair of adjacent vertices on the perimeter of a wheel graph W_n.

Figure 4: Gear graphs G_n for $n = 4, 5, 6, 7, 8$



The gear graph G_n consists of $2n+1$ vertices and $3n$ edges. All gear graphs are bipartite. One partition consists of all degree 2 tiles together with the central tile of degree n, and the other partition consists of all degree 3 tiles. Throughout this paper, we will refer to an edge incident to both a degree 3 tile and the degree n tile as a *spoke*. We will refer to the largest cycle in the graph as the *perimeter* of the graph.

Gear graphs are examples of squaregraphs, which are graphs that can be drawn in the plane in such a way that every bounded face is a quadrilateral and every vertex with three or fewer neighbors is incident to an unbounded face. Squaregraphs share many similarities with lattice graphs. The primary difference is that a tiling formed by squaregraphs is not made of regular polygons (with the exception of the the one made by squares). Squaregraphs and lattice graphs can be considered graphical representations of meshes, which are potentially useful nanostructures. In particular, a DNA mesh or lattice could be wrapped to form a DNA nanotube, a sought-after structure in DNA self-assembly [8] [5].

2 Scenario 1

Recall that in scenario 1 the incidental construction of graph-theoretical complexes of smaller size than the target graph is allowed.

Since $B_1(G) = 1$ for any graph G [1], $B_1(G_n) = 1$ for all n.

Lemma 2.1. $T_1(G_3) = 3$.

Proof. Theorem 1 in [1] states that $av(G) \leq T_1(G) \leq ev(G) + 2ov(G)$. Here, $av(G)$ denotes the number of different degrees of vertices in G, $ev(G)$ denotes the number of different even degrees of vertices in G, and $ov(G)$ denotes the number of different odd degrees of vertices in G.

By inspection, $2 \leq T_1(G_3) \leq 3$. When $n = 3$ there are only two different vertex degrees, 2 and 3. However, at least three tile types are still needed. Let $P = \{t_1, t_2\}$ be a pot with t_1 a degree 3 tile and t_2 a degree 2 tile. In order for P to realize G_3 the proportion of t_1 must be $r_1 = 4/7$ and the proportion of t_2 must be $r_2 = 3/7$. Thus the pot must satisfy the linear Diophantine equation(s)

$$3z_{i,1} + 4z_{i,2} = 0$$

where $z_{i,j}$ is the net number of cohesive ends of type a_i on tile t_j. All integer solutions to this equation are of the form $(z_{i,1}, z_{i,2}) = (-4x, 3x)$ with $x \in \mathbb{Z}$. This is impossible since t_1 is a degree 3 tile and t_2 is a degree 2 tile. □

Proposition. For any gear graph G_n, $T_1(G_n) = 3$.

Proof. Theorem 1 in [1] states that $av(G) \leq T_1(G) \leq ev(G) + 2ov(G)$.

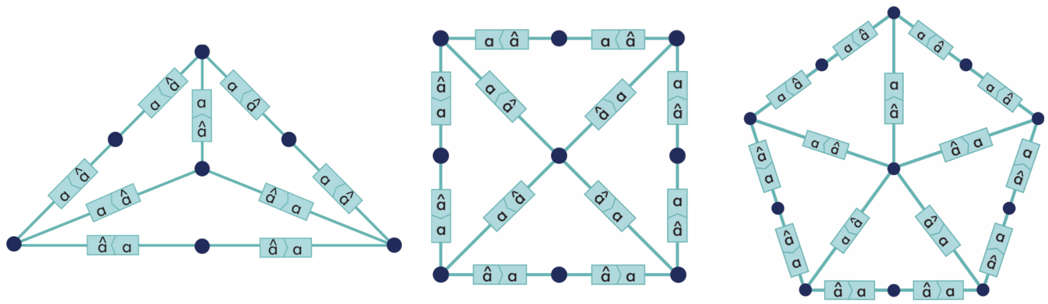
By inspection, $3 \leq T_1(G_n) \leq 5$ when $n > 3$ and n is odd, and $3 \leq T_1(G_n) \leq 4$ when $n > 3$ and n is even. Lemma 2.1 accounts for the case of $n = 3$.

The following pot realizes the lower bound for all n.

$$\mathrm{P} = \{t_1 = \{\hat{a}^n\}, t_2 = \{a^2, \hat{a}\}, t_3 = \{a, \hat{a}\}\}$$

Figure 5 below illustrates a complete labeling of G_3, G_4, and G_5 using pot P. □

Figure 5: Labeling of G_3, G_4, and G_5 in scenario 1



3 Scenario 2

Recall that in scenario 2 the incidental construction of graph-theoretical complexes of smaller size than the target graph is not allowed. However, a graph with the same size as G is acceptable.

Proposition. For any gear graph G_n, $\mathrm{T}_2(\mathrm{G}_n) = 3$.

Proof. For any graph G, $\mathrm{T}_1(\mathrm{G}) \le \mathrm{T}_2(\mathrm{G})$ by Proposition 1 in [1], then $\mathrm{T}_2(\mathrm{G}_n) \ge 3$.

The following pot P realizes G_n using 3 tile types: $\mathrm{P} = \{t_1 = \{\hat{a}_1^n\}, t_2 = \{a_1^2, \hat{a}_2\}, t_3 = \{\hat{a}_1, a_2\}\}$ The construction matrix M(P) and its reduced row echelon form follow.

$$\mathrm{M(P)} = \begin{bmatrix} -n & 2 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 1/(2n+1) \\ 0 & 1 & 0 & n/(2n+1) \\ 0 & 0 & 1 & n/(2n+1) \end{bmatrix}$$

M(P) has unique solution $\langle 1/(2n+1), n/(2n+1), n/(2n+1)\rangle$. Thus P realizes no graphs smaller than $2n+1$ vertices by Proposition 3 in [1]. □

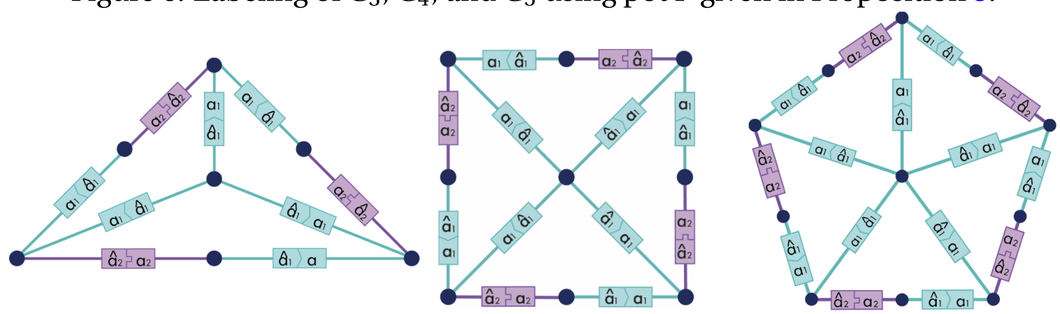
Proposition. For any gear graph G_n, $\mathrm{B}_2(\mathrm{G}_n) = 2$.

Proof. $\mathrm{B}_2(\mathrm{G}) + 1 \le \mathrm{T}_2(\mathrm{G})$ by Theorem 2 in [1]. Since $\mathrm{T}_2(\mathrm{G}_n) = 3$ by Proposition 3, it suffices to prove that $\mathrm{B}_2(\mathrm{G}_n) \neq 1$. Assume P is a pot of tiles with only one bond-edge type constructing G_n. The tiles of degree 2 in the pot P can only be labeled with a^2 or $\hat{a}^2$, since otherwise a single loop graph can form. Moreover, all tiles of degree 2 must be of the same tile type; otherwise they can combine with each other and create a graph with two vertices and a double-edge between them.

Without loss of generality, assume all tiles of degree 2 are the same tile type $t_1 = \{a^2\}$. It follows that all perimeter tiles of degree 3 have two half-edges labeled with $\hat{a}$. The un-labeled half-edge can either be labeled as a or $\hat{a}$, resulting in two possible second tile types of $t_2 = \{\hat{a}^3\}$ or $t_2 = \{\hat{a}^2, a\}$.

If $t_2 = \{\hat{a}^3\}$, then a graph of size 5 can be constructed using two of t_2 and three of t_1. If $t_2 = \{\hat{a}^2, a\}$, then $t_3 = \{\hat{a}^n\}$. If n is even, then t_1 and t_3 can combine and create a graph H of size $n/2+1$. If n is odd, then t_1 and t_3 can combine and create a graph J of size $n+2$. Since G_n has size $2n+1$ with $n \geq 3$, both H and J have fewer vertices than G_n. □

Figure 6: Labeling of G_3, G_4, and G_5 using pot P given in Proposition 3.



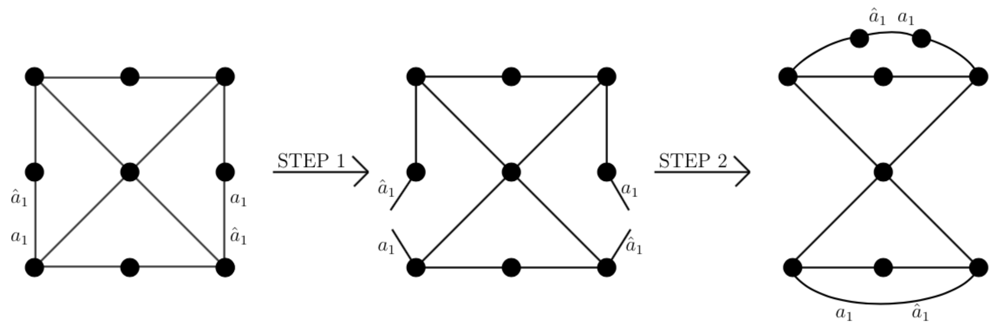
4 Scenario 3

Recall that in scenario 3 the incidental constructions of a graph smaller than G or a graph with the same size as G but not isomorphic to G are not acceptable.

Lemma 4.1. *For any gear graph* G_n*, if* P *is a pot such that* $\{G_n\} = C_{min}(P)$*, then no bond-edge type present on the tiles in* P *used in the construction of* G_n *may appear more than twice on the perimeter of* G_n*.*

Proof. Assume a bond-edge type appears three times on the perimeter of a gear graph G_n. Then, at least two of the three edges have the same orientation around the perimeter. These two edges may detach and break the perimeter cycle of size $2n$ as shown in step 1 of figure 7. In this event, the middle tile becomes a cut vertex. To see this, note that if the middle vertex is deleted the perimeter cycle decomposes into two disjoint parts, which we will now refer to as "potential components". Each potential component has two unpaired half-edges of the same bond-edge type, one hatted and one unhatted. It follows that each pair of complementary half-edges can join together within each potential component as shown in step 2 of figure 7. If the unpaired edges rejoin in this way, no edge is formed between the two components. Then, the middle tile remains a cut vertex even after the two edges rejoin. Since any two vertices of a gear graph lay on a common cycle, gear graphs have no cut vertices ([7], Theorem 4.2.4). Therefore, the resulting graph is not isomorphic to G_n. □

Figure 7: Edges with the same orientation and bond-edge type break and join to create a graph that is not isomorphic to G_n.



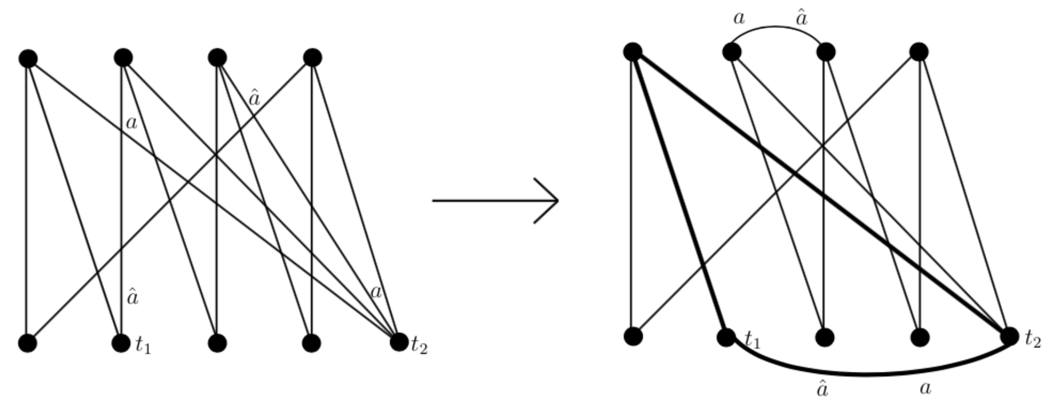
Lemma 4.2. *For any gear graph* G_n*, if* P *is a pot such that* $\{G_n\} = C_{min}(P)$*, then the half-edges of the tiles in* P *used in the construction of* G_n *that are in the same partition must have the same version (hatted or unhatted) of any given bond-edge type.*

Proof. Let X be the partition with all degree 2 tiles and the degree n tile and Y be the partition with all perimeter degree 3 tiles. By Lemma 2 of [1] no tile used in the construction of G_n may have both the hatted and unhatted version of a bond-edge type.

If any two tiles t_1 and t_2 in X each have a half-edge labeled with opposite versions of the same bond-edge type, then these two edges may be detached and re-join so that t_1 and t_2 are adjacent. If the degree of both t_1 and t_2 is 2, then this forms a non-isomorphic graph since no degree 2 tiles in G_n are adjacent to one another. Now consider if t_1 is a degree 2 tile and t_2 is the degree n tile. If t_1 and t_2 each have a half-edge labeled with opposite versions of the same bond-edge type, then a cycle of length 3 can form between the degree 2 tile, the degree n tile, and the tile neighboring both of those tiles (see Figure 8). G_n has no odd cycles, so this also produces a non-isomorphic graph. If any two tiles in Y each have a half-edge labeled with opposite versions of the same bond-edge type, the these two edges can detach and re-join the two degree 3 tiles to one another. Since no degree 3 tiles are adjacent in G_n, this produces a non-isomorphic graph.

Hence, if a bond-edge type appears on any two (or more) tiles in the same partition, it must appear as the same version on each, otherwise the tiles can reconfigure to form a graph not isomorphic to G_n. Therefore, all half-edges in either partition must be labeled with the same version of any given bond-edge type. □

Figure 8: A degree 2 tile and the degree n tile of G_4 are labeled with the opposite version of the same bond type. A length three cycle is formed.



Lemma 4.3. *If* P *is a pot such that* $\{\mathrm{G}_n\} = \mathrm{C}_{min}(\mathrm{P})$*, and a perimeter edge and spoke edge of* G_n *use the same bond-edge type, then these two edges are incident to the same vertex.*

Proof. Assume that a spoke s and an edge e on the perimeter have the same bond-edge type a and that they are not incident to the same vertex. Let t_1 be the degree 3 tile containing a half-edge of e. Then, the spoke adjacent to t_1 is not s since it is assumed that s and e are not incident to the same vertex. Without loss of generality, let's label t_1's spoke half-edge with the bond-edge type x (this could be any bond-edge type including a). Then, t_1 has one half-edge labeled a (or $\hat{a}$), and one half-edge labeled with x (or $\hat{x}$). The central tile has two half-edges labeled with the same bond-edge types as t_1. The middle tile and t_1 are in different partitions, so Lemma 4.2 guarantees that if a bond-edge type appears in one version (hatted/unhatted) on one tile, it will appear in the opposite version on the other tile. It follows that t_1 and the middle tile can combine and create a double edge, which creates a graph not isomorphic to G_n. □

Lemma 4.4. *For any gear graph* G_n*, if* P *is a pot such that* $\{\mathrm{G}_n\} = \mathrm{C}_{min}(\mathrm{P})$*, then no two half-edges of a tile type used twice in the construction of* G_n *can be labeled with the same bond-edge type.*

Proof. If two half-edges of two tiles of the same tile type are labeled with the same bond-edge type, then there are four edges in G_n labeled with that bond-edge type. At most two of these four edges can be on the perimeter by Lemma 4.1, so two edges must be spoke edges.

Since no two spoke edges and a perimeter edge are incident with the same vertex, Lemma 4.3 dictates that these two spoke edges must be labeled with a different bond-edge type, making this construction impossible. □

Lemma 4.5. *For any gear graph* G_n*, if* P *is a pot such that* $\{\mathrm{G}_n\} = \mathrm{C}_{min}(\mathrm{P})$*, then no three tiles in* P *used in the construction of* G_n *can be of the same tile type.*

Proof. In any gear graph, if three tiles are the same then they are either all degree 2 tiles or all degree 3 tiles. If three degree 2 tiles are the same, then those three tiles are on the perimeter and non-adjacent. In this case, a bond-edge type used on the tile would appear at least three times around the perimeter, a violation of Lemma 4.1.

Assume three degree 3 tiles have the same tile type. Each arm of the tile type will have a different bond-edge type, as dictated by Lemma 4.4. Label any one of the spokes with bond-edge type a. The spokes of the other two tiles must be labeled with a, since by Lemma 4.3 no two spokes and one perimeter edge can be labeled using the same bond-edge type and no two non-incident perimeter edges and a spoke can be labeled with the same bond-edge type. The six perimeter edges (two per tile) must be labeled with the remaining two bond-edge types. It follows that at least one bond-edge type is repeated more than twice on the perimeter, a violation of Lemma 4.1. □

Lemma 4.6. *If* P *is a pot such that* $\{\mathrm{G}_n\} = \mathrm{C}_{min}(\mathrm{P})$ *and three edges of* G *use the same bond-edge type, then these three edges must be incident with a common vertex.*

Proof. No three edges on the perimeter can have the same bond-edge type by Lemma 4.1. Lemma 4.3 dictates that no two spoke edges and one perimeter edge can have the same bond-edge type. Thus, three edges with the same bond-edge type must be either two perimeter edges and one spoke edge, all adjacent to a single vertex, or three spoke edges. □

Proposition. For any gear graph G_n, $T_3(G_n) = n+2$.

Proof. By Lemma 4.5, if $\{G_n\} = C_{min}(P)$ then P must consist of at least $\lceil \frac{n}{2} \rceil$ degree 2 tile types and at least $\lceil \frac{n}{2} \rceil$ degree 3 tile types. A distinct tile type is needed for the middle tile since it has degree n, which is a greater degree than all other tiles in G_n for $n > 3$. In the case of $n = 3$ a distinct tile type is still needed to label the middle tile, otherwise Lemma 4.4 is violated. This results in a minimum of $\lceil \frac{n}{2} \rceil + \lceil \frac{n}{2} \rceil + 1 = n+1$ distinct tile types when n is even and $\lceil \frac{n}{2} \rceil + \lceil \frac{n}{2} \rceil + 1 = n+2$ distinct tile types when n is odd.

Assume by way of contradiction that $\{G_n\} = C_{min}(P)$ where n is even and P is a pot with $n+1$ tile types. Then, by Lemma 4.5 each tile type must appear exactly twice on the perimeter. It follows that each bond-edge type on the perimeter is used twice, which implies no spoke can be labeled with those same bond-edge types by Lemma 4.3.

Consider an arbitrary degree 2 tile, which we will call t_1, on the perimeter of G_n. Suppose t_1's half-edges are labeled with $\hat{a}$ and $\hat{b}$. Label the matching half-edges connecting t_1 to the two adjacent degree 3 tiles. Next, consider the degree 3 tile with one half-edge labeled with b. Its other perimeter half-edge cannot be labeled with b by Lemma 4.4, it cannot be labeled with $\hat{a}$ or $\hat{b}$ by Lemma 4.2, and it cannot be labeled with a because the labeling of that edge would have the same bond-edge type and direction as one of the edges of tile t_1. Then, label that edge with a new bond-edge type c. Label the matching half-edge. Consider the degree 2 tile with a half-edge label $\hat{c}$. Its other edge has to be labeled with a new bond-edge type which we call $\hat{d}$. Label the matching half-edge. Consider the degree 3 tile with one half-edge labeled with d. Its other perimeter half-edge cannot be labeled with any of the bond-edge types already used because of Lemmas 4.2, 4.4, and the fact that two perimeter edges with the same bond-edge type must have opposite orientation around the perimeter as explained above. Since a new bond-edge type is required to label any additional tile on the perimeter, the initial degree 2 tile t_1 can never be repeated in G_n. This necessarily results in two unmatched degree 2 tile types that each appear only once on the perimeter of G_n. This leaves $2n-2$ other tiles on the perimeter, n of which are degree 3 and $n-2$ of which are degree 2. All other tile types on the perimeter can be used exactly twice, resulting in $n/2$ degree 3 tile types and $(n-2)/2$ degree 2 tile types that are repeated. In sum, $n+1$ tile types must be used on the perimeter of G_n.

The following two pots, one for n even and one for n odd, achieve this bound:

$P_{even} = \{t_1 = \{a_1^2, a_{n+1}\}, t_i = \{\hat{a}_{i-1}, a_i\}$ for $i = 2,4...n$, $t_i = \{\hat{a}_{i-1}, a_i, a_{n+1}\}$ for $i = 3,5...n-1$, $t_{n+1} = \{\hat{a}_n^2, a_{n+1}\}$, $t_{n+2} = \{\hat{a}_{n+1}^n\}\}$

$P_{odd} = \{t_1 = \{a_1^2, a_{n+1}\}, t_i = \{\hat{a}_{i-1}, a_i\}$ for $i = 2,4...n-1$, $t_i = \{\hat{a}_{i-1}, a_i, a_{n+1}\}$ for $i = 3,5...n$, $t_{n+1} = \{\hat{a}_n^2,\}$, $t_{n+2} = \{\hat{a}_{n+1}^n\}\}$

To see that these two pots do not realize any graphs of smaller size than G_n, consider the construction matrix for each.

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

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

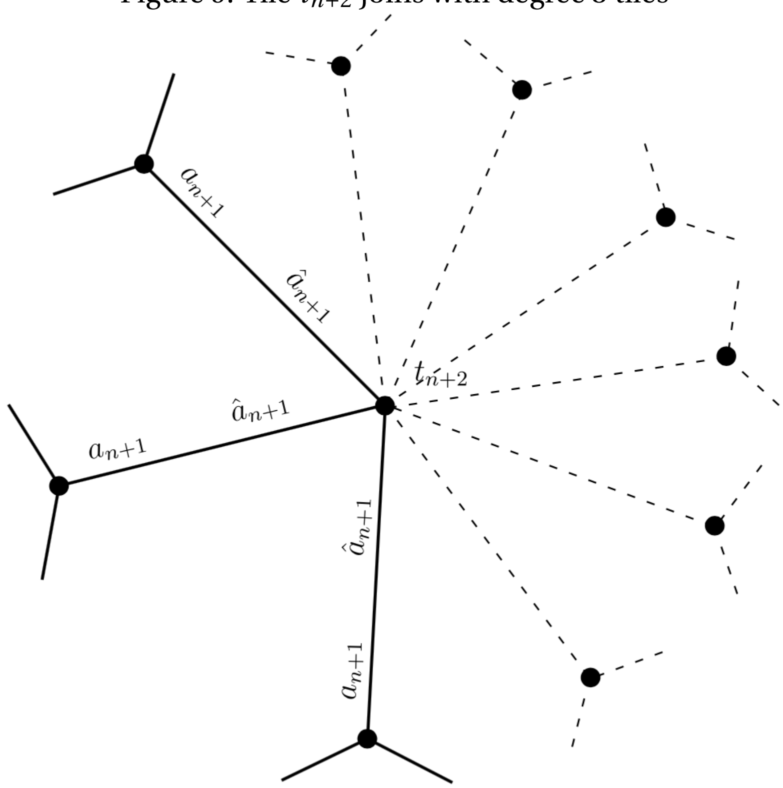
Both matrices have the following reduced row echelon form.

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 1/(2n+1) \\ 0 & 1 & 0 & 0 & \cdots & 0 & 2/(2n+1) \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \ddots & \ddots & 0 & 0 & 2/(2n+1) \\ 0 & 0 & 0 & \ddots & 1 & 0 & 1/(2n+1) \\ 0 & 0 & 0 & \cdots & 0 & 1 & 1/(2n+1) \end{bmatrix}$$

Hence, both matrices have the unique solution $\langle 1/(2n+1), 2/(2n+1), ..., 2/(2n+1), 1/(2n+1), 1/(2n+1)\rangle$.

To see that no graph of size $2n+1$ not isomorphic to G_n may be realized by this pot(s), consider $G' \in C_{min}(P)$. Note that G' must be constructed with the exactly the same number of each tile type as G_n, since the construction matrix has a unique solution. The tile t_{n+2} has n half-edges labeled $\hat{a}_{n+1}$. Notice that no other tiles have half-edges labeled $\hat{a}_{n+1}$. Also notice that there are exactly n half-edges labeled a_{n+1}, each appearing on a distinct degree 3 tile type. It follows that t_{n+2} must join with all n degree 3 tile types via one half-edge each. This is shown in the figure below.

Figure 9: Tile t_{n+2} joins with degree 3 tiles



The only tile types remaining in the pot are those of degree 2. Moreover, no other half-edge can combine at this point since half-edges of all degree 3 tiles have the same version of any given bond-edge type as dictated by Lemma 4.2. The same is true for all degree 2 tiles, and therefore they cannot form any bond between each other. It follows that each degree 2 tile must bond together with two degree 3 tiles. By doing so G' becomes a graph isomorphic to G_n. □

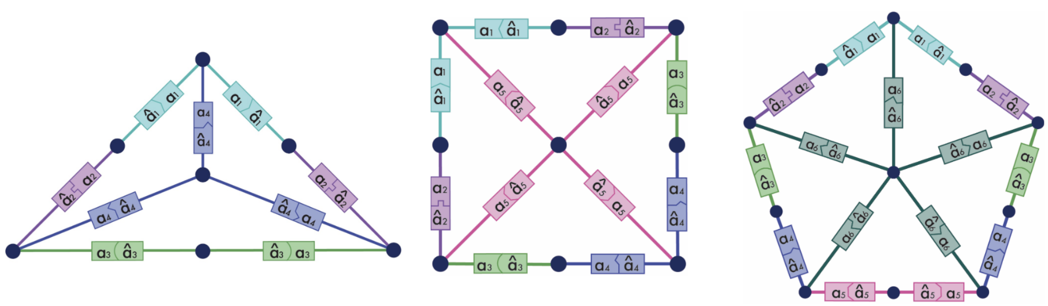
Proposition. For any gear graph G_n, $B_3(G_n) = n+1$.

Proof. First we show $B_3(G_n) \geq n+1$. Suppose P is a pot such that $\{G_n\} = C_{min}(P)$. There are two possible cases in the labeling of G_n using the tiles of P: (1) At least one spoke is labeled with a bond-edge type different from all bond-edge types of the perimeter edges; (2) No spoke is labeled with a bond-edge type different from all bond-edge types of the perimeter edges.

In the first case, Lemma 4.1 requires at least $n+1$ bond-edge types be used to label the perimeter and spokes. Let's now consider case 2. If spokes are of the same bond-edge types as perimeter edges, then all n spokes must be of distinct bond-edge types by Corollary 4.3. Assume that only those n bond-edge types are used to label all tile types in P. This results in each bond-edge type being used exactly three times when labeling G_n, twice for perimeter edges and once for a spoke edge. By Lemma 4.6 this forces a labeling in which degree 3 tiles have half-edges labeled with the same version of the same bond-edge type. Let t_1 and t_2 be two degree 3 tiles adjacent to the same degree 2 tile, t_3, in this gear graph labeling. Then, a graph G' of size 5 can be constructed from the pot $\{t_1, t_2, t_3\}$. This violates the conditions for scenario 3.

The pots given in Proposition 4 achieve this bound. A complete labeling of G_3, G_4, and G_5 using these pots is shown below in Figure 10. □

Figure 10: Labeling of G_3, G_4, and G_5 in scenario 3



5 Conclusion

This work built upon previously known graph theoretical methods for optimizing the self-assembly process to find minimum numbers of tile and bond-edge types for the family of gear graphs. This task was accomplished for three different scenarios, with scenario 3 being the most restrictive. Optimal results found for scenarios 1 and 2 are constant and do not depend on the size of G_n. However, minimum numbers of tile and bond-edge types for scenario 3 are size dependent and thus increase as n increases.

We anticipate that the results on gear graphs presented in this work could aid others in finding optimal solutions for grid graphs, as the gear graph G_n has a grid-like structure of n squares around a central vertex. Future work includes exploration of tilings of repeated gear graphs to form grid-like DNA meshes.

References

[1] 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*, chapter 13, pages 241–270. Springer, 2014.

[2] Nataša Jonoska, Gregory L McColm, and Ana Staninska. Spectrum of a pot for dna complexes. In *International Workshop on DNA-Based Computers*, pages 83–94. Springer, 2006.

[3] Thomas H. LaBean and Hanying Li. Constructing novel materials with dna. *Nano Today*, 2(2):26–35, 2007.

[4] Dan Luo. The road from biology to materials. *Materials Today*, 6(11):38–43, 2003.

[5] Abdul M. Mohammed, Petr Šulc, and Rebecca Schulman. Self-assembling dna nanotubes to connect molecular landmarks. *Nature Nanotechnology*, 12:312–316, 2017.

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

[7] Douglas West. *Introduction to Graph Theory*, chapter 4, page 162. Pearson, New York, New York, 2018.

[8] Ofer I. Wilner, Ron Orbach, Anja Henning, Carsten Teller, Omer Yehezkeli, Michael Mertig, Daniel Harries, and Itamara Willner. Self-assembly of dna nanotubes with controllable diameters. *Nature Communications*, 2:540, 2011.

[9] Hao Yan, Sung Ha Park, Gleb Finkelstein, John H. Reif, and Thomas H. LaBean. Dna-templated self-assembly of protein arrays and highly conductive nanowires. *Science*, 301(5641):1882–1884, 2003.

Chiara Mattamira
Converse College
cmattamira001@converse.edu