



Journal of Mathematical Problems, Equations and Statistics

E-ISSN: 2709-9407
P-ISSN: : 2709-9393
JMPES 2021; 2(2): 08-16
© 2020 JMPES
www.mathematicaljournal.com

Received: 16-04-2021
Accepted: 05-06-2021

Mohd Altaf Mir
M.Phil Scholar, Mathematics,
Rabindranath Tagore
University, Raisen, Madhya
Pradesh, India

Dr. Bhawna Agrawal
Associate Professor,
Mathematics, Rabindranath
Tagore University, Raisen,
Madhya Pradesh, India

A mathematical study of DNA complexes of one-bond edge type

Mohd Altaf Mir and Dr. Bhawna Agrawal

Abstract

The discovery of DNA structure 55 years ago marked the beginning of a process that has transformed the foundations of biology and medicine, and accelerated the development of new fields, such as molecular biology or genetic engineering. Today, we know much about DNA, its properties, and function. We can determine the structure of short DNA fragments with picometer precision, find majority of the genes encoded in DNA, and we can manipulate, stretch and twist individual DNA molecules. We can utilize our knowledge of gene regulatory apparatus encoded in DNA to produce new microorganisms with unexpected properties. Yet, there are aspects of DNA function that defy our understanding, mostly because the molecule is just one, albeit essential, component of a complex cellular machinery.

Keywords: Mathematical study, DNA complexes, one-bond edge etc.

Introduction

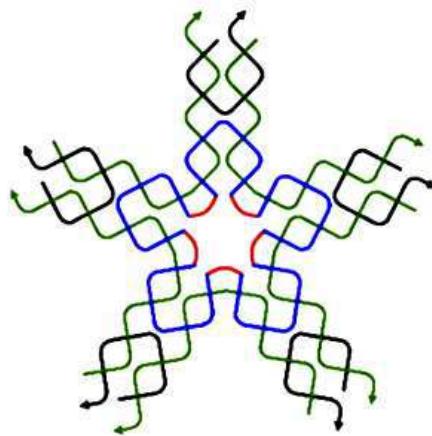
DNA self-assembly is a fascinating and important experimental technique that is being used in labs all around the United States. Nadrian C. Seeman pioneered the utilisation of DNA self-assembly as a bottom-up technology for producing target nanostructures approximately 35 years ago. The method is based on the complementary nature of the nucleotides that make up DNA's structure. DNA self-assembly offers a wide range of applications, from nanostructure creation to experimental virotherapies (especially helpful in the treatment of malignancies) [EMPB+14, See07]. However, the materials are costly, and the process has a high danger of producing a large amount of waste. We may use a mix of graph theoretic and algebraic tools to improve the assembly process by modelling the issue of DNA self-assembly using graph theory tools.

Branched junction molecules, which are asterisk-shaped molecules with multiple arms extending from the vertices, are used to construct target structures. Each arm is made up of a DNA fragment with uneven length strands, resulting in a cohesive conclusion. Each arm's cohesive-end will link with a comparable cohesive-end from the opposite arm. We utilise single alphabet letters to distinguish between cohesive-ends of different sorts, rather than referring to the particular nature of a cohesive-end (such as the exact nucleotide arrangement). For example, a and b represent two non-compatible cohesive ends, yet a will bind with a and b with b. A pair of complimentary cohesive-ends is referred to as a bond-edge type.

Mathematical Structure of DNA

Scientific articles by a group of Brazilian researchers from the University of Campinas (Unicamp) and the University of São Paulo (USP) show that genetic sequences can have the same mathematical structure as the Error Correcting Codes (ECC) used in both broadcast and digital recording systems. ECCs are a set of commands built into the software installed in computer chips, telecommunications equipment, televisions and smartphones to correct digital information defects in such processes as telephone conversations or the storage of data on a computer's hard disk.

Correspondence
Mohd Altaf Mir
M.Phil Scholar, Mathematics,
Rabindranath Tagore
University, Raisen, Madhya
Pradesh, India

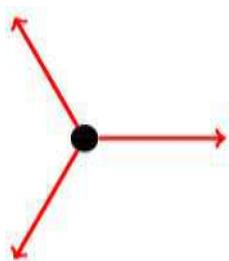
**Fig 1:** Example of a 5-armed Branched Junction Molecule

The same mathematical logic, say the researchers, is found in the formation of DNA—the deoxyribonucleic acid whose cells carry the genes and all instructions for development and survival of living beings. In the study, they compared algebraic equations of error-correcting codes with certain DNA sequences, attributing a numerical logic to the nucleotides that make up the genome: thymine (T), guanine (G), cytosine (C) and adenine (A). In doing so, they discovered that there are patterns that link the nucleotide to a number. Thus, depending on the type of sequence, A is represented by 0, C is 2, G is 1 and T is 3. In digital language, which consists of bits, the information is translated into 0s and 1s. “We have shown that DNA has sequences that follow the same mathematical structures and rules as digital communication,” says Márcio de Castro Silva Filho, from the Genetics Department of the Luiz de Queiroz School of Agriculture (ESALQ) at USP. “The DNA sequence is not random; it follows a pattern,” he says.

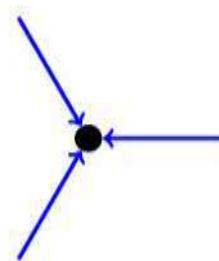
Definitions

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, \dots . We will denote complementary cohesive-ends with $\hat{a}, \hat{b}, \hat{c}, \dots$
2. A bond-edge type is a classification of the cohesive-ends of tiles (without regard to hatted and unhattled letters). For example, a and \hat{a} will bond to form bond-edge type a .
3. We denote tiles by t_j , where $t_j = \{a^{e1}, a_1^{e2}, \dots, a_k^{e2k-1}, a_k^{e2k}, \dots\}$ 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**Fig 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.

This tool captures the characteristic of a pot that we are interested in. Let $P = \{t_1, \dots, t_p\}$ 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:

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

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

$$\sum_j r_j = 1 \text{ and } \sum_j r_j Z_{i,j} = 0 \quad \text{for all } i \quad (3.2)$$

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

Definition 2: Let P be a pot with tiles t_i for $i \in \{1, \dots, n-1\}$.

Then the construction matrix of P is given by

$$M_P = \begin{bmatrix} & t_1 & t_2 & \cdots & t_{n-1} & \\ a_1 & Z_{1,1} & Z_{1,2} & \cdots & Z_{1,n-1} & 0 \\ a_2 & Z_{2,1} & Z_{2,2} & \cdots & Z_{2,n-1} & 0 \\ \vdots & \vdots & \vdots & & & \vdots \\ z_{m,1} & \cdots & z_{m,2} & \cdots & z_{m,n-1} & 0 \\ 1 & \cdots & 1 & \cdots & 1 & 1 \end{bmatrix}$$

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: The solution space of M_P is called the spectrum of P , and is denoted by $S(P)$. The following lemma, which may be found in [EMPB⁺14], connects $S(P)$ to graphs realized by P .

Lemma 4: Let $P = \{t_1, \dots, t_p\}$. If $(r_1, \dots, r_p) \in S(P)$, 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 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

$$\begin{aligned} t_1 : z_{1,1} &= 3 \\ t_2 : z_{1,2} &= -1 \\ t_3 : z_{1,3} &= -3. \end{aligned}$$

The construction matrix is

$$M_P = \begin{bmatrix} 3 & -1 & -3 & | & 0 \\ 1 & 1 & 1 & | & 1 \end{bmatrix}$$

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

$$\text{rref}(M_P) = \begin{bmatrix} 1 & 0 & \frac{1}{2} & \frac{1}{4} \\ 0 & 1 & \frac{3}{2} & \frac{3}{4} \end{bmatrix}$$

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

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

Thus we have

$$\mathcal{S}(P) = \left\{ \frac{1}{4}(1 - 2z, 6 - 3z, 4z) \mid z \in \mathbb{Q}_{\geq 0} \right\} \text{ and } P \text{ realizes, for example, a graph of order 4.}$$

Pots with a 1-Armed Tile

This research focuses on pots that contain at least one 1-armed tile. That is, $P_1 = \{\{a^m\}, \{\hat{a}\}, \{a^n\}, \dots\}$ 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 $P'_1 = \{\{a^m\}, \{a\}, \{a^n\}, \dots\}$ where $m < n$. The pot p_1 has corresponding construction matrix $M_{P1} = \begin{bmatrix} m & -1 & -n & \dots & 0 \\ 1 & 1 & 1 & \dots & 1 \end{bmatrix}$.

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 $S(P1)$ has at least one variable. The pot P has construction matrix $M_{P1} = \begin{bmatrix} m & -1 & -n & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$

The spectrum of p is described in Lemma 4.1.

Lemma 1: Consider the pot p with the associated construction matrix M_P .

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

Proof: Row-reduce M_P to obtain'

$$\text{ref}(M_P) = \begin{bmatrix} 1 & 0 & -\frac{n}{m} + \frac{\frac{n}{m} + 1}{m(\frac{1}{m} + 1)} & \frac{1}{m(\frac{1}{m} + 1)} \\ 0 & 1 & \frac{\frac{n}{m} + 1}{\frac{1}{m} + 1} & \frac{1}{\frac{1}{m} + 1} \end{bmatrix}$$

Equation 4.2 can be simplified to

$$\text{rref}(M_P) = \begin{bmatrix} 1 & 0 & \frac{m - nm}{m(m+1)} & \frac{1}{m+1} \\ 0 & 1 & \frac{n+m}{m+1} & \frac{m}{m+1} \end{bmatrix} \quad (4.3)$$

From Equation 4.3, we have

$$\begin{aligned} x &= \frac{1}{m+1} - \frac{m - nm}{m(m+1)} z \\ &= \frac{1}{m(m+1)} [m - (m - nm)z] \end{aligned} \quad 4.3$$

$$\begin{aligned} y &= \frac{m}{m+1} - \frac{n+m}{m+1} z \\ &= \frac{m^2}{m(m+1)} - \frac{nm + m^2}{m(m+1)} z \end{aligned} \quad 4.4$$

$$\begin{aligned} z &= \frac{m(m+1)}{m(m+1)} z \\ &= \frac{1}{m(m+1)} [(m^2 + m)z] \end{aligned} \quad 4.6$$

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

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

$$\text{Set. } Z = \frac{q}{q+r+1}$$

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

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: Let $P = \{\{a^4\}, \{\hat{a}\}, \{\hat{a}^3\}\}$. Then, according to Definition 1.3 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

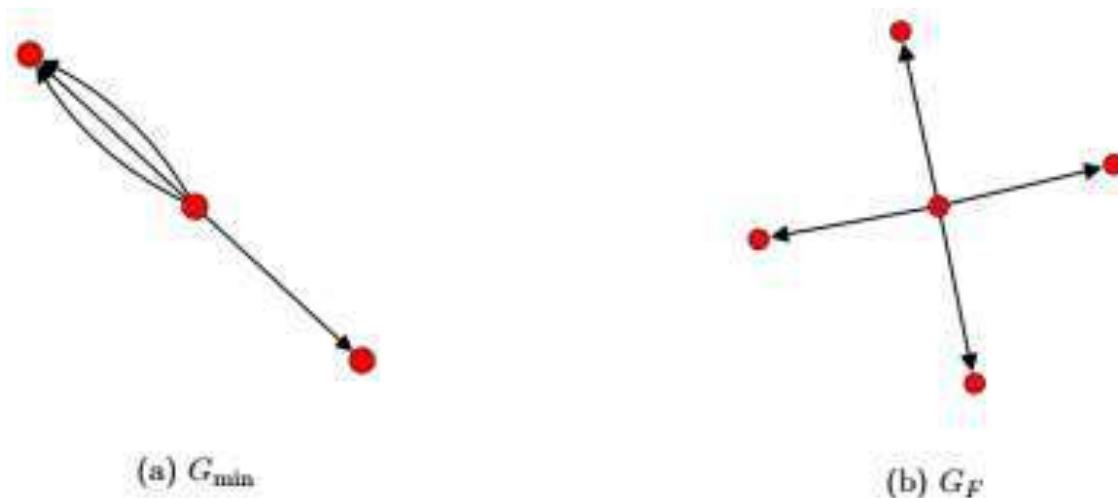


Fig 3: Two Graphs for $P = \{a^4, \hat{a}, \hat{a}^3\}$

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: [EMPB⁺14] The order of G_{\min} , denoted mp , is $mp = \min\{\text{lcm}\{b_j \mid r_j \neq 0 \text{ and } r_j = \frac{a_j}{b_j}\}$

where, $\langle r_1, \dots, r_p \rangle \in \mathcal{S}(P)\}$

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

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

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

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 hated 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. \square

Example 1: 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: 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,

$$m + 1 > q + r + 1$$

Orders of Graphs Realized by P

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 The pot P realizes a disconnected graph for an order S if and only if, for at least one tile distribution (R_{11}, \dots, R_{j1})

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

associated to S , we can write

where each j -tuple (R_{1i}, \dots, 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}, \dots, R_{j1}) be the tile distribution used to

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

realize G . Then

where $H_i \cap H_j = \emptyset$ 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}, \dots, R_{ji}) , that constructs H_i . Further, since each $H_i \subset G$, it follows that $R_{\ell i} \leq R_{1i}$ for all ℓ .

Thus we translate Equation 4.9 into the language of tile distributions to arrive at

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

where each j -tuple (R_{1i}, \dots, R_{ji}) is a tile distribution of P that realizes some graph H_i . Conversely, if we suppose an

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

order S has at least one tile distribution

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 $S = s_1 + s_2 + \dots + s_\ell$

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: 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.

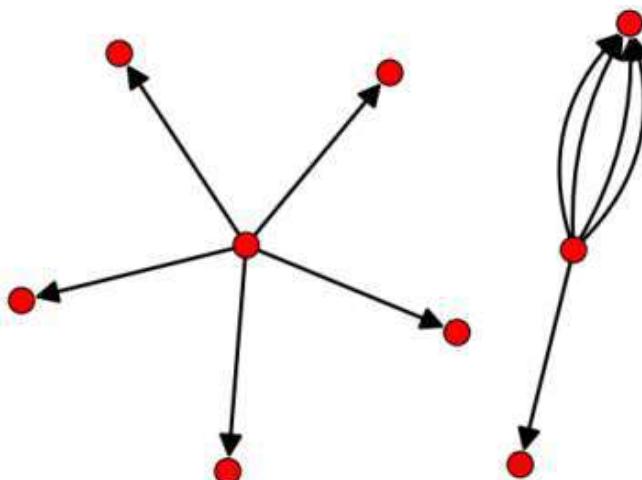


Fig 4: Graph of order 9 for $\{\{a^5\}, \{\hat{a}\}, \{\hat{a}^4\}\}$

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.

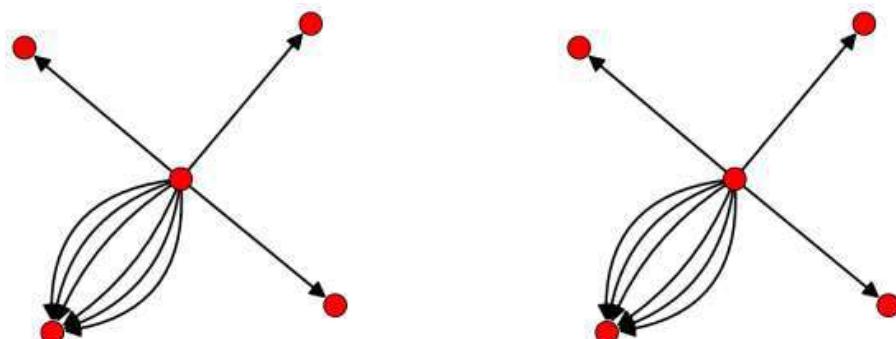


Fig 5: Disconnected graph of order 10 for $\{\{a^9\}, \{\hat{a}\}, \{\hat{a}^6\}\}$

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

$$S = s_1 + s_2 + \dots + s_\ell$$

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

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 $\eta = [12.57] = 13$. 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.

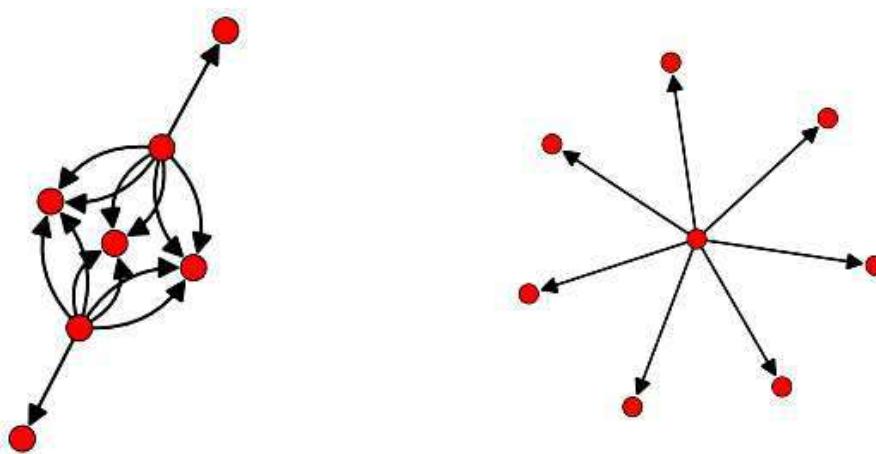


Fig 6: Disconnected graph of order 15 for $\{\{a^7\}, \{\hat{a}\}, \{\hat{a}^4\}\}$

What is perhaps more interesting is that we can obtain a disconnected graph on 12 vertices. This is because $m_p = 5 < \zeta$.

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 $G \in O(P)$ 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.

References

1. Aldinger J, Klapper I, Tabor M. Formulae for the calculation and estimation of writhe. *J Knot Theory Ramifications* 1995;4:343-372.
2. Antman SS. Nonlinear problems of elasticity, of Applied Mathematical Sciences, Springer, New York, second ed., 2005, 107.
3. Antman SS, Kenney CS. Large buckled states of nonlinearly elastic rods under torsion, thrust, and gravity, *Arch. Rational Mech. Anal* 1981;76:289-338.
4. Antman SS, Liu T-P. Travelling waves in hyperelastic rods, *Quart. Appl. Math.* 1979;36:377-399.
5. Arsuaga J, Va'zquez M, Trigueros S, Sumners D, Roca J. *Knottin*.