



SCHOOL of  
GRADUATE STUDIES  
EAST TENNESSEE STATE UNIVERSITY

East Tennessee State University  
**Digital Commons @ East  
Tennessee State University**

---

Electronic Theses and Dissertations

Student Works

---

5-2005

# Using Domination to Analyze RNA Structures.

Travis Reves Coake  
*East Tennessee State University*

Follow this and additional works at: <https://dc.etsu.edu/etd>

 Part of the [Applied Mathematics Commons](#), and the [Genetics and Genomics Commons](#)

---

## Recommended Citation

Coake, Travis Reves, "Using Domination to Analyze RNA Structures." (2005). *Electronic Theses and Dissertations*. Paper 1022.  
<https://dc.etsu.edu/etd/1022>

This Thesis - Open Access is brought to you for free and open access by the Student Works at Digital Commons @ East Tennessee State University. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Digital Commons @ East Tennessee State University. For more information, please contact [digilib@etsu.edu](mailto:digilib@etsu.edu).

Using Domination to Analyze RNA Structures

---

A thesis  
presented to  
the faculty of the Department of Mathematics  
East Tennessee State University

In partial fulfillment of  
the requirements for the degree  
Master of Science in Mathematical Sciences

---

by  
Travis Reves Coake  
May 2005

---

Debra Knisley, Ph.D., Chair  
Teresa Haynes, Ph.D.  
Robert Gardner, Ph.D.  
Yue Zou, Ph.D.

Keywords: RNA, domination, total dominating set, locating dominating set,  
differentiating dominating set, paired-dominating set

## ABSTRACT

Using Domination to Analyze RNA Structures

by

Travis Reves Coake

Understanding RNA molecules is important to genomics research. Recently researchers at the Courant Institute of Mathematical Sciences used graph theory to model RNA molecules and provided a database of trees representing possible secondary RNA structures. In this thesis we use domination parameters to predict which trees are more likely to exist in nature as RNA structures. This approach appears to have promise in graph theory applications in genomics research.

## DEDICATION

I dedicate this thesis to all of those people who have supported me throughout my educational journey.

## ACKNOWLEDGEMENTS

It is a pleasure to thank the many people that made this thesis possible.

It is difficult not to overemphasize my gratitude to my advisors, Dr. Teresa Haynes and Dr. Debra Knisley. Without their great efforts, inspiration, and enthusiasm this thesis would not have been possible. I would also like to thank Kim Atkins, my fiancé for the past three years, for all the support and encouragement she has given me.

## CONTENTS

ABSTRACT . . . . .	2
DEDICATION . . . . .	3
ACKNOWLEDGEMENTS . . . . .	4
LIST OF FIGURES . . . . .	6
LIST OF TABLES . . . . .	7
1 INTRODUCTION . . . . .	8
1.1 Genomics Background . . . . .	8
1.2 Domination . . . . .	13
1.3 Locating, and Differentiating Dominating Sets . . . . .	19
1.4 Paired-Dominating Sets . . . . .	24
2 NEW RESULTS . . . . .	26
2.1 Trees of Order 2 through 6 . . . . .	27
2.2 Trees of Order 7 . . . . .	28
2.3 Trees of Order 8 . . . . .	32
2.4 Making a Prediction for Order 9 . . . . .	35
3 TAKING A CLOSER LOOK . . . . .	38
4 CONCLUSION . . . . .	40
BIBLIOGRAPHY . . . . .	41
APPENDICES . . . . .	43
Appendix A: List of trees for orders 2 through 9 . . . . .	43
VITA . . . . .	48

## LIST OF FIGURES

1	2D-RNA Example . . . . .	9
2	2D-RNA Illustration . . . . .	10
3	Corresponding 2D-RNA Tree to Figure 1 . . . . .	11
4	Domination of a Spider . . . . .	14
5	Domination Example . . . . .	14
6	Domination of a Tree . . . . .	15
7	Total Domination of a Spider . . . . .	16
8	Total Domination Example . . . . .	17
9	Total Domination of a Tree Example . . . . .	17
10	Locating Domination Example . . . . .	19
11	Complete Graph, $K_4$ . . . . .	20
12	Complete Graph, $K_4$ . . . . .	21
13	Tree of Order 9 . . . . .	22
14	Hypercube $Q_3$ . . . . .	24
15	Trees of Order 2 through 6 . . . . .	27
16	Trees of Order 7 . . . . .	28

## LIST OF TABLES

1	Table of Trees of Order 7 . . . . .	29
2	Results for Trees of Order 7 . . . . .	31
3	Table of Trees of Order 8 . . . . .	32
4	Results for Trees of Order 8 . . . . .	34
5	Table of Trees of Order 9 . . . . .	35
6	Results for Trees of Order 9 . . . . .	37
7	Best Candidates of Order 9 . . . . .	39



# 1 INTRODUCTION

## 1.1 Genomics Background

‘RNA molecules are integral components of the cellular machinery for protein synthesis and transport, transcriptional regulation, chromosome replication, RNA processing and modification, as well as other fundamental biological functions ’[3]. Due to current research, scientists are finding more and more functional RNAs; thus, the need for a method of cataloging the existing and novel RNAs as well as developing a better understanding of RNA is justified and is part of the motivation for this thesis. However, the number of known RNAs found in nature though increasing is still limited. Current research has not found a concrete method for determining which RNAs are likely to exist in nature. Nevertheless, enough data exists to provide researchers a means to develop a method of predicting RNA structures.

A secondary RNA structure typically denoted as 2D-RNA is a single-stranded RNA molecule made up of four nucleotide bases denoted by the letters *A,C,G*,and *U* that have been “folded” to form a more stable structure. For example, a single stranded RNA molecule will not be stable unless some of its exposed bases are protected from water or solvent; thus, the best way to protect these bases is to pair them with another base [3]. However, not all pairs within the molecule are suitable for such a pairing, so some choices of pairs are better than others. Therefore, a prediction method must find the most stable pairing of the bases to form the secondary RNA structure. Figure 1, illustrates a small stable 2D-RNA structure.

Acinetobacter\_sp\_AD1.trna45 Ala (GGC) 83.49 bits

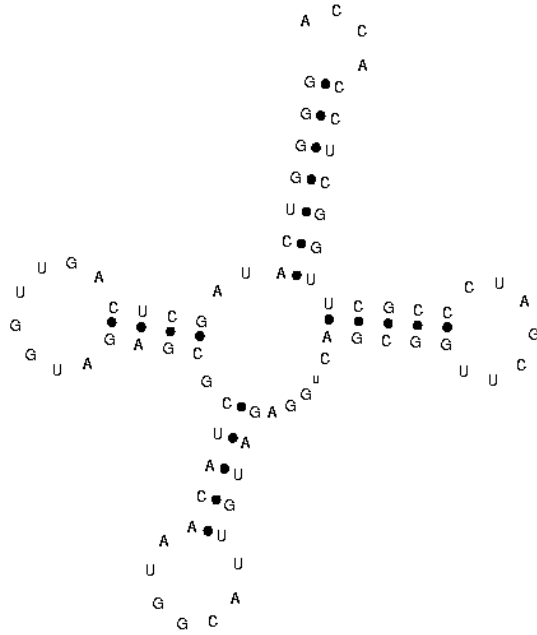


Figure 1: 2D-RNA Example

In Figure 1, we see not only complementary base pairs but also what seems to be loops as well. The stem is made up of two or more complementary base pairs, thus in Figure 1, we have four stems. However, in the structure we also find a number of unpaired bases between the stems which we call a loop. In this next example we see that not all stems are necessarily perfect like Figure 1 suggests, rather, stems can contain unpaired bases as well which we call bulges. Figure 2 illustrates part of a 2D-RNA structure where the stem contains a bulge.

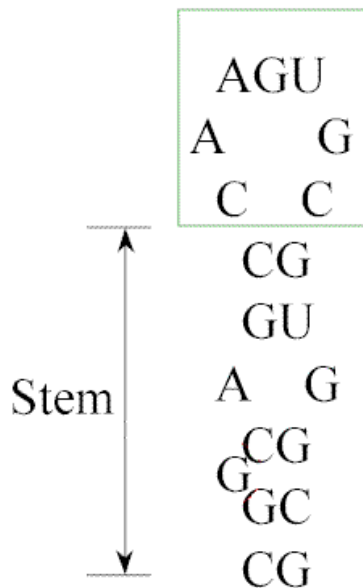


Figure 2: 2D-RNA Illustration

In Figure 2, we notice that the stem contains a base  $G$  that is not paired with any other base which creates a bulge. In addition, this illustration also defines where the stem stops and where the loop begins.

In a paper entitled “Exploring the repertoire of RNA secondary motifs using graph theory; implications for RNA design” Tamar Schlick et al. defined a method of modeling 2D-RNA structures as a particular family of graphs called trees. Graphical trees have also been used in the past to represent chemical structures where the individual atoms were vertices and their bonds were the edges; however, the RNA model is different. Schlick et al. define how to represent a RNA structure as a tree such that vertices ( $\bullet$ ) represent the bulges and loops, and edges ( $-$ ) represent the stems.

If we consider the secondary RNA structure in Figure 1, we can see that the corresponding tree graph by Schlick's definition is illustrated in Figure 3.

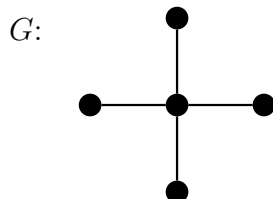


Figure 3: Corresponding 2D-RNA Tree to Figure 1

Two ways exist to count the total number of possible trees for any given number of vertices in a graph. One labels the vertices and the other does not. Our research pertains only to unlabelled vertices, and Harary and Prins obtained the counting polynomial  $t(x)$  whose coefficient  $N_v$  is the number of distinct graphs with  $N$  vertices [3].

$$\begin{aligned}
 t(x) &= \sum_{N=1}^{\infty} N_v x^v \\
 &= T(x) - \frac{1}{2} [T^2(x) - T(x^2)],
 \end{aligned} \tag{1}$$

where

$$T(x) = x \exp \left[ \sum_{r=1}^{\infty} \frac{1}{r} T(x^r) \right] \tag{2}$$

The counting polynomial up to the first 10 terms is

$$t(x) = x + x^2 + x^3 + 2x^4 + 3x^5 + 6x^6 + 11x^7 + 23x^8 + 47x^9 + 106x^{10}$$

We can see that given a particular order  $N$ , the number of total possible trees increases rather quickly when considering the total number of known 2D-RNA trees that exist in nature which is about 35. This suggests that there are a number of RNA trees that are still not found, yet could possibly exist in nature [10]. Thus, Schlick et al. developed a couple ways of predicting which trees are potential candidates to represent RNA structures that exist in nature. As a result they have categorized all possible topological trees into three classes: red, blue, and black. Red represents known RNA, blue represents theoretical RNA structures that based on their prediction method, they can expect to exist in nature. Finally, black represents hypothetical RNA structures they do not expect to see in nature. However, because we are dealing with nature, the possibility always exists.

As stated earlier, a better understanding of RNA is important in genomics research. As researchers today find more functional RNAs in nature, the need for a method of cataloging existing as well as novel RNA also becomes important. Therefore, with this application in mind, we are using graph theory as a method of predicting novel RNA in nature. In the next section, we will discuss the various domination parameters that we found useful in predicting 2D-RNA structures that could possibly exist in nature.

## 1.2 Domination

The concept of domination in graphs began in the 1850s with the game of chess. The goal of the problem was to use certain chess pieces to dominate the squares of a chessboard. In the game of chess, a queen can move horizontally, vertically, or diagonally. De Jaenish, in 1862, considered the problem of finding the minimum number of queens that can be placed on a chessboard such that every square is either occupied by a queen or can be occupied by a queen in a single move. It turns out that the minimum number of queens needed is five, and this became known as the Five Queens Problem [5, 8].

The connection between the Five Queens Problem and domination can be seen if we let each vertex of a graph represent a square of the 64 squares of a chessboard. Then, two vertices are adjacent in  $G$  if each corresponding square can be reached by a queen on the other square in a single move. This graph is referred to as the Queens graph. Hence, the minimum number of queens that can dominate the entire chessboard forms a dominating set in  $G$  [8]. We now consider the formal definitions and concepts of domination.

A set  $S$  of vertices of a graph  $G = (V, E)$  is a *dominating set* of  $G$  if every vertex in  $V - S$  is adjacent to some vertex of  $S$ . The *domination number*  $\gamma(G)$  is the minimum cardinality of a dominating set of  $G$ . In a tree, a vertex of degree one is referred to as a *leaf* and a vertex which is adjacent to a leaf is a *support vertex*. If a support vertex is adjacent to two or more leaves, it is called a *strong support vertex*. Consider the example of a *spider* shown in Figure 4.

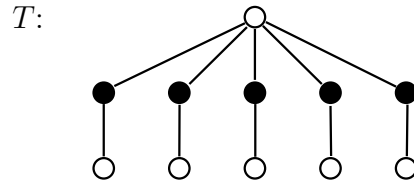


Figure 4: Domination of a Spider

Notice that if we let our set  $S$  be the five darkened vertices, then each of the remaining vertices is adjacent to a vertex in  $S$ . Hence  $\gamma(T) \leq 5$ . Cause at least one of each leaf and its support must be in  $S$ , we have  $\gamma(T) \geq 5$ . Therefore, it follows that  $\gamma(T) = 5$ .

Consider a different example, suppose we have the following graph shown in Figure 5.

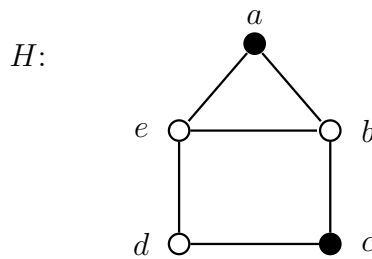


Figure 5: Domination Example

Then the set  $S = \{a, c\}$  forms a dominating set of  $H$ , and because no single vertex is adjacent to all other vertices, we have  $\gamma(H) = 2$ .

A vertex  $u$  is said to be *connected* to a vertex  $v$  in a graph  $G$  if there exists a

$u - v$  path in  $G$ . A graph  $G$  is connected if every pair of its vertices is connected. A *tree*  $T$  is a connected graph with no cycles. In many cases, we look at instances of domination of trees. For example, consider the tree shown in Figure 6.

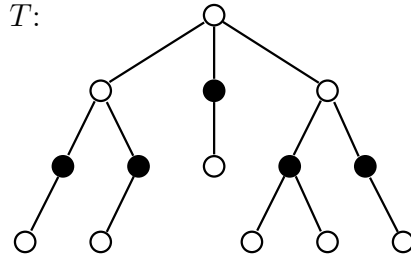


Figure 6: Domination of a Tree

Notice that, in this case,  $\gamma(T) = 5$  because the five darkened vertices dominate  $T$  and we must include at least the number of support vertices to dominate the leaves.

In order for a set of vertices  $S$  to be dominating every vertex not in the set must be adjacent to at least one vertex in the set. If we tighten the condition and require every vertex of a graph  $G$  to be adjacent to some vertex in  $S$ , then we have a total dominating set of  $G$ . Formally, a set  $S$  of vertices of a graph  $G = (V, E)$  is a *total dominating set* of  $G$  if every vertex in  $V$  is adjacent to some vertex in  $S$ . The minimum cardinality of a total dominating set of  $G$  is the *total domination number*  $\gamma_t(G)$ . Note that  $\gamma_t(G)$  is defined only for graphs with no isolated vertices. Cause every total dominating set is a dominating set, we have  $\gamma(G) \leq \gamma_t(G)$  for all graphs  $G$  with no isolated vertices.

Consider again the *spider* shown in Figure 4. The darkened vertices form a dom-



inating set  $S$  of  $T$ , but each of these five vertices need to be adjacent to a vertex in  $S$ . A total dominating set  $S'$  of  $T$  is illustrated in Figure 7.

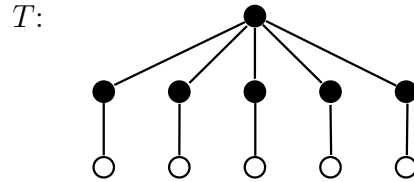


Figure 7: Total Domination of a Spider

Letting  $S'$  represent the darkened vertices, notice that every vertex of  $T$  is now adjacent to a vertex of  $S'$ . Hence,  $\gamma_t(T) \leq 6$ . Cause five vertices are necessary to dominate the leaves and none of the supports are adjacent to each other, it follows that a sixth vertex is needed and hence,  $\gamma_t(T) \geq 6$ . Therefore, we conclude that  $\gamma_t(G) = 6$ .

Consider again the graph shown in Figure 5. The set  $S = \{a, c\}$  forms a dominating set of  $H$  and we have  $\gamma(H) = 2$ . Cause these two vertices are not also adjacent to each other,  $S$  is not a total dominating set of  $H$ . However, the set  $S' = \{b, c\}$  forms a total dominating set of  $H$  shown in Figure 8 below.

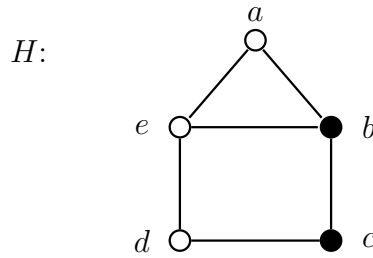


Figure 8: Total Domination Example

Therefore,  $\gamma_t(H) \leq 2$ . Cause  $2 = \gamma(H) \leq \gamma_t(H)$ , we have  $\gamma_t(H) = 2$ . Notice this is an example of a graph for which the domination and total domination numbers are equal.

Finally, consider again the tree shown in Figure 6. The support vertices of the tree are a dominating set of  $T$ . However, because the support vertices are not adjacent, this cannot also be a total dominating set of  $T$ . A total dominating set of  $T$  is shown below by the set of darkened vertices and it can be shown that  $\gamma_t(T) = 8$ .

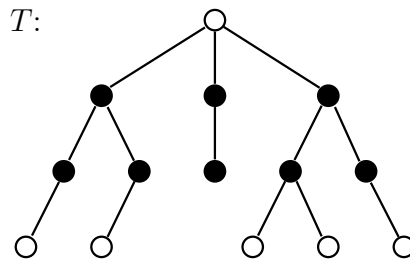


Figure 9: Total Domination of a Tree Example

In general, we follow the terminology of [5]. A more extensive study of domination in graphs can be found in [5, 6].

### 1.3 Locating, and Differentiating Dominating Sets

Consider the floor plan of a building as modelled by a graph where a vertex represents a room in the building and two vertices are adjacent if the corresponding rooms are adjacent. Suppose we wish to install expensive sensors in the building which will transmit a signal at the detection of an intruder (fire, burglar, etc.). Cause the sensors are expensive we wish to optimize their usage. This safeguards facility analysis of the corresponding graph motivated the concept of locating sets and further the idea of locating-dominating sets.

Let  $S = \{v_1, v_2, \dots, v_k\}$  be a set of vertices in a connected graph  $G = (V, E)$  and let  $v \in V$ . The  $k$ -vector (ordered  $k$ -tuple)  $c_s(v)$ , of  $v$  with respect to  $S$  is defined by

$$c_s(v) = (d(v, v_1), d(v, v_2), \dots, d(v, v_k))$$

where  $d(v, v_i)$  is the distance between  $v$  and  $v_i$  ( $1 \leq i \leq k$ ). The set  $S$  is called a *locating set* if the  $k$ -vectors  $c_s(v)$ , for all vertices  $v \in V$ , are distinct. This concept is studied in [12, 13].

For example, suppose we have the graph  $H$  given in Figure 10.

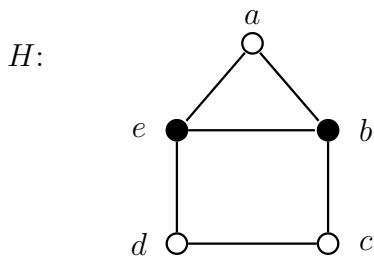


Figure 10: Locating Domination Example

In order for the set  $S = \{b, e\}$  to be a locating set of  $H$ , the 2-vectors  $c_s(v)$  must be distinct for all  $v \in V$ . Notice

$$c_s(a) = \{1, 1\}$$

$$c_s(b) = \{0, 1\}$$

$$c_s(c) = \{1, 2\}$$

$$c_s(d) = \{2, 1\}$$

$$c_s(e) = \{1, 0\}$$

Cause all of the 2-vectors are distinct, we conclude that  $S$  is a locating set of  $H$ .

Slater [13, 14] defined a *locating-dominating set* in a connected graph  $G$  to be a dominating set  $S$  of  $G$  such that for every two vertices  $u$  and  $v$  in  $V(G) - S$ ,  $N(u) \cap S \neq N(v) \cap S$ . The minimum cardinality of a locating-dominating set of  $G$  is the *location-domination number*  $\gamma_L(G)$ . Notice that the location-domination number is defined for every connected graph  $G$  since  $V$  is such a set. This concept is studied in [1, 2, 9, 13, 14, 15] and elsewhere. To illustrate, suppose we have  $G = K_4$ , a complete graph on four vertices, as shown in Figure 11 below.

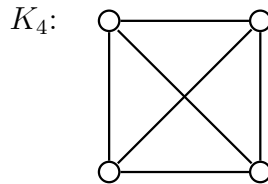


Figure 11: Complete Graph,  $K_4$

Notice that to dominate this graph, we need only one vertex, say vertex  $a$ , since all edges are present between any pair of vertices. Therefore  $\gamma(K_4) = 1$ . However, the dominating set  $S = \{a\}$  is not a locating-dominating set since the remaining vertices are all adjacent to  $a$ , and hence  $N(b) \cap S = \{a\}, N(c) \cap S = \{a\}, N(d) \cap S = \{a\}$ . In fact, any locating-dominating set of a complete graph must include all the vertices, except one. For instance  $S' = \{a, b, c\}$  is a locating-dominating set of  $K_4$  as illustrated in Figure 12.

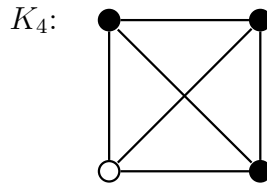


Figure 12: Complete Graph,  $K_4$

Hence  $\gamma_L(K_4) \leq 3$ . Notice that because  $K_4$  is a complete graph, any locating-dominating set must include all the vertices except one, and so  $\gamma_L(G) \geq 3$ . Therefore,  $\gamma_L(G) = 3$ . In general, for all complete graphs  $K_n$ ,  $\gamma_L(K_n) = n - 1$ .

In order for a set  $S$  to be a locating set, every vertex in  $V(G) - S$  must be distinguished in terms of its open neighborhood intersecting  $S$ . If we require all of the vertices of  $G$  to be distinguished, then we have a differentiating set of  $G$ . Gimbel et al. [4] defined a set  $S$  to be a *differentiating dominating set* if  $S$  is a dominating set and for every pair of vertices  $u$  and  $v$  in  $V$ ,  $N[u] \cap S \neq N[v] \cap S$ . The *differentiating domination number*  $\gamma_D(G)$  is the minimum cardinality of a differentiating dominating

set of  $G$ . Cause every differentiating dominating set is a dominating set, we have  $\gamma(G) \leq \gamma_D(G)$ . Consider the tree  $T$  of order 9 shown in Figure 13.

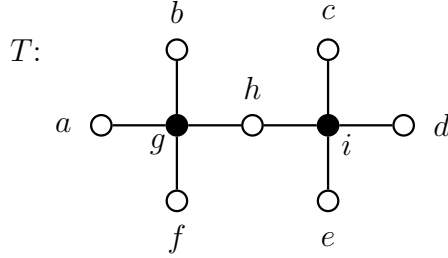
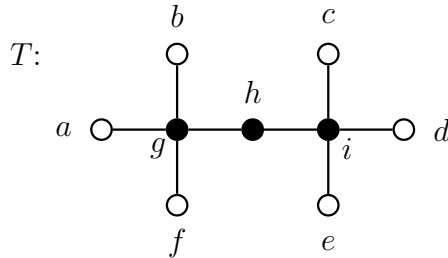


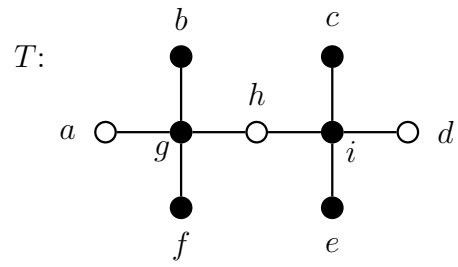
Figure 13: Tree of Order 9

Notice that the set  $S = \{g, i\}$  forms a dominating set of  $T$ , and it can be easily seen that  $\gamma(T) = 2$  and  $\gamma_i(T) = 3$ . However, notice that  $N[a] \cap S = \{g\}$  and  $N[b] \cap S = \{g\}$ , and so  $S$  is not differentiating.



Let  $S = \{b, c, e, f, g, i\}$ . Thus any pair of distinct vertices  $u$  and  $v$  in  $V$  can be differentiated and so  $\gamma_D(T) \leq 6$ . Cause it can be shown that no five vertices is a differentiating dominating set for  $T$ ,  $\gamma_D(T) \geq 6$ . Therefore,  $\gamma_D(T) = 6$ .

Now any pair of distinct vertices  $u$  and  $v$  in  $V$  can be differentiated and so



$\gamma_D(T) \leq 6$ . Cause no five vertices is a differentiating dominating set for  $T$ ,  $\gamma_D(T) \geq 6$ .

Therefore,  $\gamma_D(T) = 6$ .



## 1.4 Paired-Dominating Sets

Consider a police force where each police officer is adjacent to all the people in which he/she is responsible for protecting in a given area. Naturally cases will exist in which an individual officer can be overrun. However, if each officer was assigned a partner, the chances of overrunning either or both officers is less likely. This is an application of paired-domination. *Paired-domination* was introduced by Haynes and Slater [5] where a *dominating set*  $S$  with *matching*  $M$  is a dominating set  $S = \{v_1, v_2, \dots, v_{2t-1}, v_{2t}\}$  with independent edge set  $M = \{e_1, e_2, \dots, e_t\}$  where each edge  $e_i$  is incident to two vertices of  $S$ , that is,  $M$  is a perfect matching in  $\langle S \rangle$  [5]. The *paired-domination number*  $\gamma_{pr}(G)$  is the minimum cardinality of a paired dominating set  $S$  in  $G$ .

For example, for the graph  $Q_3$  in Figure 14, notice that the set  $S = \{a, g\}$  forms a dominating set of  $Q_3$ , and it can be shown that  $\gamma(Q_3) = 2$ .

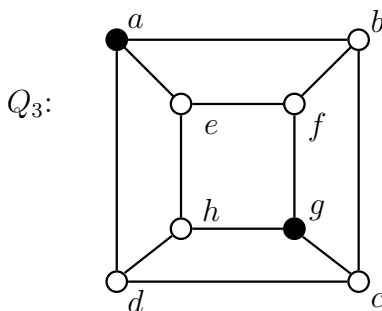
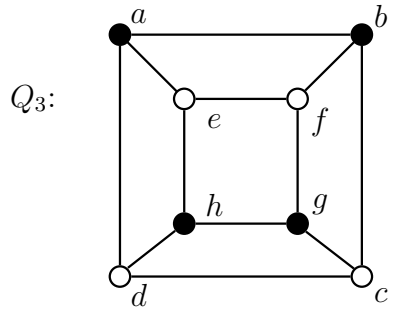


Figure 14: Hypercube  $Q_3$

Notice vertices  $a$  and  $g$  are not adjacent and, therefore, do not form a *paired-dominating* set. However, because  $S = \{a, g\}$  is a dominating set, suppose we add

any two vertices to  $S$  such that at least one vertex added to the dominating set is adjacent to another vertex already in  $S$ .



Cause vertices  $a$  and  $g$  already form a dominating set, the addition of vertices  $b$  and  $h$  now form a paired-dominating set because each element in  $S = \{a, b, g, h\}$  has a perfect matching. For instance,  $a$  is paired with  $b$  since  $a$  is adjacent to  $b$  in  $Q_3$ , and  $g$  is paired with  $h$  since  $g$  and  $h$  are adjacent in  $Q_3$ . Thus it can be shown that the *paired-domination* number  $\gamma_{pr}(Q_3) = 4$ .

## 2 NEW RESULTS

In this chapter, I will present the new results we have obtained for predicting 2D-RNA structures using only domination parameters.

Recall from [3] Schlick et al. categorized all possible topological trees of a given order into three classes: red, blue, and black where red represents known RNA structures. Blue represents theoretical RNA, structures we expect to exist in nature. Black represents hypothetical RNA, structures we expect not to exist in nature. However, because we are dealing with nature, it is always possible that a structure can exist despite what our results suggest.

Our prediction method of secondary RNAs is based upon domination parameters. We discovered that by taking the sum or quotient of a few select parameters resulted in a strong correlation between known RNA trees and hypothetical trees. Specifically, we found the following three parameters useful in the prediction of secondary RNA motifs:

1.  $\gamma_L + \gamma_D$
2.  $\gamma + \gamma_t + \gamma_{pr}$
3.  $\frac{\gamma}{\gamma_L + \gamma_D}$

## 2.1 Trees of Order 2 through 6

The following figure illustrates all topological trees of orders two through six which we obtained from [11]. The colors represent the classes predicted in [11].

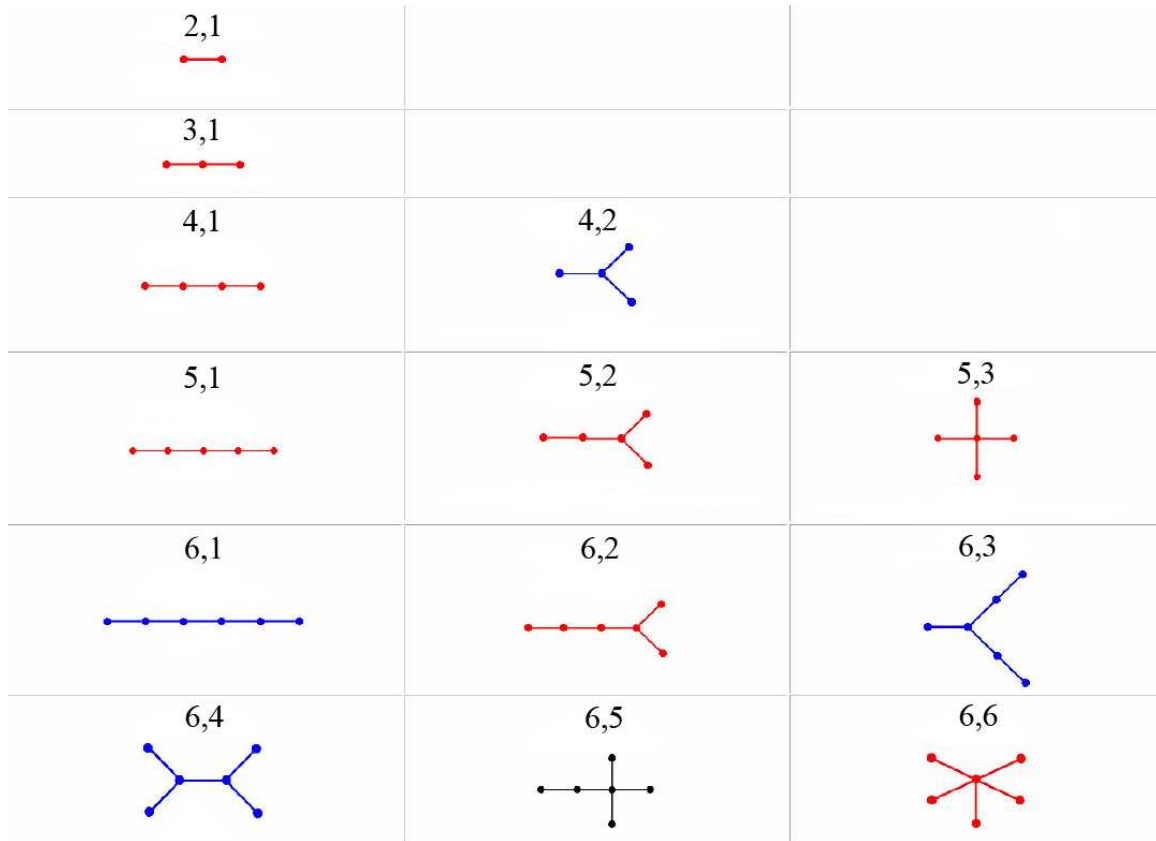


Figure 15: Trees of Order 2 through 6

As you can see in Figure 15, most of the RNA trees from order two through six have already been found in nature. Cause the order of these trees is small, we did not find the corresponding parameter values for these trees helpful in developing our

prediction method. However, once the order of the trees reached seven, we discovered a significant relationship between known RNA trees and our three domination parameters. In Figure 16 below, all topological trees of order seven are listed. As you can see, as the order of the trees gets larger the number of trees grows significantly. Therefore, for the remainder of this thesis we will reference the RNA trees by the number by which they are listed. A complete list of trees used in our research is attached in the appendix.

## 2.2 Trees of Order 7

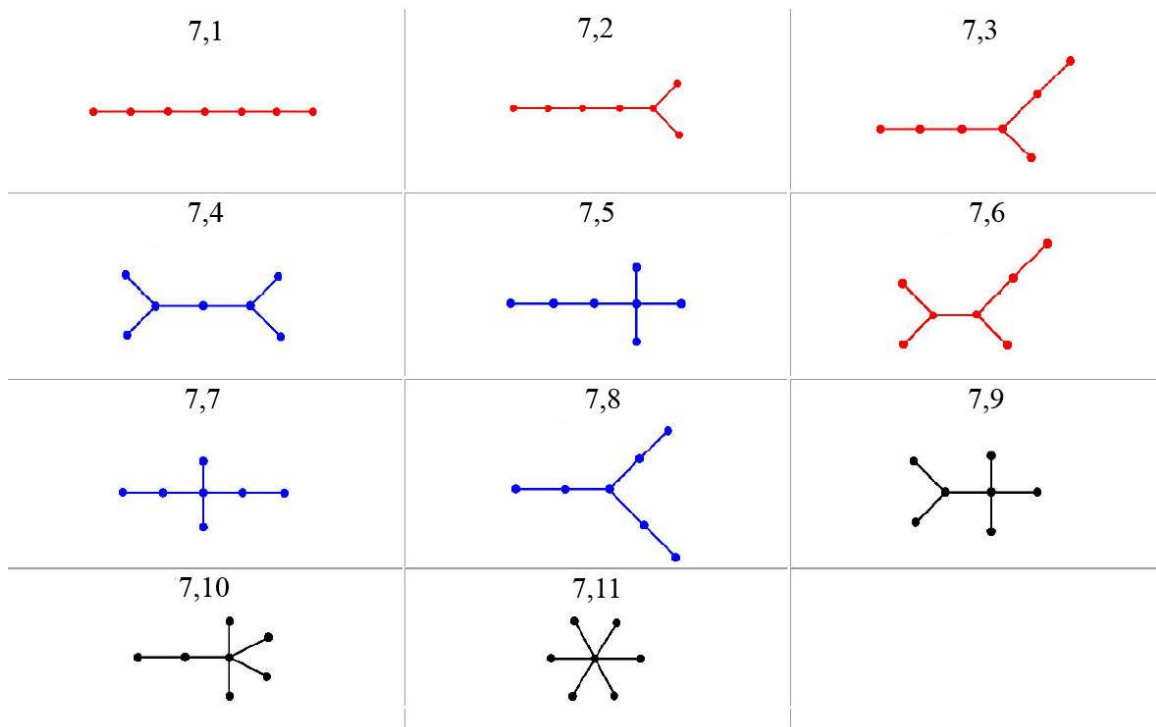


Figure 16: Trees of Order 7

Table 1: Table of Trees of Order 7

Tree	$\gamma_L + \gamma_D$	$\gamma + \gamma_t + \gamma_{pr}$	$\frac{\gamma}{\gamma_L + \gamma_D}$
1	7	11	.43
2	9	10	.22
3	7	11	.43
4	9	9	.22
5	9	9	.22
6	8	10	.38
7	8	10	.38
8	7	13	.43
9	10	6	.20
10	10	6	.20
11	12	5	.08

Consider Table 1 above where the values listed in the table correspond to the graphs in Figure 16. The trees that represent known RNA structures are highlighted in red. We notice that most of the values for the first parameter  $\gamma_L + \gamma_D$  that correspond to known RNAs are less than the other trees. In fact, all the red trees are less than the others with the exception of tree 2. Moreover, if we study the range of these values for the first parameter, we notice that these values are between seven to twelve inclusive; however, when we consider the range of values for the red trees we can see that these values span only from seven to nine which are at the lower end of the overall range.

In the second parameter  $\gamma + \gamma_t + \gamma_{pr}$  we see a similar pattern except the values for the red trees are typically larger than the rest. Furthermore, we note that the

values for the second parameter range from five to thirteen where the values for the red trees are at the higher end of the scale. Though a little more difficult to see, we also notice that the values for the the red trees for the third parameter  $\gamma/(\gamma_L + \gamma_D)$  are generally larger than the other values in the table, where the overall values vary from 0.08 to 0.43.

Because all of the values for the known RNA trees tend to be either higher or lower than the rest of the hypothetical trees, this suggests a bound may exist within the range for each parameter. Furthermore, this may also suggest that the farther away the value of the tree is from the bound, the more likely it is to exist. If we look at the values for the first parameter, it is clear that the median is nine, where the median is defined by ignoring repetition. Examining the values for the first parameter again, we find that the existing RNA trees have values that are less than the median. Studying the values of the second parameter, we find that the median is also nine and all the existing RNA trees have values that are greater than or equal to the median. When we calculate the median of the third parameter we find the median is 0.275, and all the red trees have values larger than the median except for tree 2 which has a value of 0.22. Because using the median is only a starting point for determining the bound then we simply change the value of the bound to 0.22 to accommodate tree 2. Because we are using the known RNA trees to develop a prediction method for novel RNA, then for any given order where there are known RNA trees we change the bound to accommodate them, if any, that contradict that bound.

Now, we have determined bounds for all three parameters for trees of order seven, we can now distinguish which RNA trees are likely to exist and are not likely to

exist. Table 2, illustrates these results where red circles represent RNA trees already found in nature, blue circles represent RNA trees that we expect to see in nature, and black circles represent hypothetical RNA trees we do not expect to see in nature. In addition, because such remarkable similarities in the end result with our prediction method and that used by Schlick et al. arised, we have included their results as well and have denoted them by squares.

Table 2: Results for Trees of Order 7

	<b>RAG</b>	$\gamma_L + \gamma_D$	$\gamma + \gamma_t + \gamma_{pr}$	$\frac{\gamma}{\gamma_L + \gamma_D}$
1	■	●	●	●
2	■	●	●	●
3	■	●	●	●
4	■	●	●	●
5	■	●	●	●
6	■	●	●	●
7	■	●	●	●
8	■	●	●	●
9	■	●	●	●
10	■	●	●	●
11	■	●	●	●

We determined a tree to be blue for the first parameter if its value is less than or equal to the median, for the second and the third parameter if its value is greater than or equal to the median, and black otherwise. As you can see, each parameter



makes the same prediction for the trees of order seven as that of Schlick et al. Thus, if we were to make a prediction on order seven, the final result would be the same as that of Schlicks et al.; however, though remarkably similar, this is not so for orders greater than seven at this time.

### 2.3 Trees of Order 8

Table 3: Table of Trees of Order 8

Tree	$\gamma_L + \gamma_D$	$\gamma + \gamma_t + \gamma_{pr}$	$\frac{\gamma}{\gamma_L + \gamma_D}$
1	9	11	.33
2	9	11	.33
3	9	11	.33
4	10	10	.20
5	8	14	.38
6	10	10	.20
7	9	11	.33
8	9	11	.33
9	11	9	.18
10	9	14	.33
11	8	12	.50
12	9	11	.33
13	10	10	.30
14	11	9	.18
15	10	10	.30
16	9	13	.33
17	10	10	.30
18	12	6	.13
19	12	6	.13
20	8	14	.50
21	10	10	.30
22	12	6	.13
23	14	5	.07

Table 3 corresponds to the graphs of trees of order eight. Again the trees that

represent known RNA structures are highlighted in red. We notice the same pattern in parameter values for trees of order eight that we found in trees of order seven. For example, we notice that the range of values for the first parameter  $\gamma_L + \gamma_D$  is between eight to fourteen inclusive, and that the values of the known RNA structures span only eight to ten. Note that these values are at the lower end of the range compared to the other trees.

In the second parameter  $\gamma + \gamma_t + \gamma_{pr}$  most of the values for the red trees are typically larger than the rest. The values for the second parameter vary from five to fourteen where the values for the red trees are at the higher end of the scale. The values for the the red trees for the third parameter  $\gamma/(\gamma_L + \gamma_D)$  are usually larger than the other values in the table, where the values extend from 0.07 to 0.5.

We looked for bounds in trees of order eight following the same procedure as described for trees of order seven. If we examine the values for the first parameter for trees of order eight, we see that the median is ten. We also find that the values for the first parameter for existing RNA trees are less than or equal to the median. The median for the second parameter is also ten, and all the existing RNA trees have values that are greater than or equal to the median. We calculated that the median of the third parameter is 0.28, and all the red trees values are greater. Thus, these are the same results we found for order seven.

Because we have calculated bounds for all three parameters for trees of order eight, we can determine which RNA trees are likely to exist and are not likely to exist. Again, we have compared our results from our prediction method to Schlick et al, indicating their results with squares.

Table 4: Results for Trees of Order 8

RAG	$\gamma_L + \gamma_D$	$\gamma + \gamma_t + \gamma_{pr}$	$\frac{\gamma}{\gamma_L + \gamma_D}$	
1	■	●	●	●
2	■	●	●	●
3	■	●	●	●
4	■	●	●	●
5	■	●	●	●
6	■	●	●	●
7	■	●	●	●
8	■	●	●	●
9	■	●	●	●
10	■	●	●	●
11	■	●	●	●
12	■	●	●	●
13	■	●	●	●
14	■	●	●	●
15	■	●	●	●
16	■	●	●	●
17	■	●	●	●
18	■	●	●	●
19	■	●	●	●
20	■	●	●	●
21	■	●	●	●
22	■	●	●	●
23	■	●	●	●

Consider the results we obtained with both orders seven and eight and notice how remarkably similar our results are to those produced by Schlick's method. Based on these results, we make predictions for the existence of trees for order nine.

## 2.4 Making a Prediction for Order 9

Table 5: Table of Trees of Order 9

Tree	$\gamma_L + \gamma_D$	$\gamma + \gamma_t + \gamma_{pr}$	$\frac{\gamma}{\gamma_L + \gamma_D}$	Tree	$\gamma_L + \gamma_D$	$\gamma + \gamma_t + \gamma_{pr}$	$\frac{\gamma}{\gamma_L + \gamma_D}$
1	9	14	.33	25	13	9	.15
2	11	11	.27	26	11	14	.27
3	11	11	.27	27	9	15	.44
4	9	12	.44	28	12	10	.25
5	11	11	.27	29	9	15	.44
6	9	14	.33	30	10	12	.40
7	10	11	.30	31	11	11	.27
8	10	11	.30	32	13	9	.15
9	12	10	.17	33	11	13	.27
10	10	14	.30	34	11	13	.27
11	9	15	.44	35	12	10	.25
12	9	12	.44	36	12	10	.25
13	10	11	.30	37	12	10	.25
14	9	15	.44	38	10	14	.40
15	12	10	.17	39	12	10	.25
16	11	11	.27	40	14	6	.14
17	10	14	.30	41	14	6	.14
18	10	15	.30	42	9	17	.44
19	10	14	.30	43	10	14	.40
20	11	11	.27	44	11	11	.27
21	10	11	.30	45	12	10	.25
22	11	11	.27	46	14	6	.14
23	12	9	.17	47	16	5	.06
24	10	12	.40				

Table 5, listed above shows parameter values for trees of order nine. We already

know of four RNA structures that exist in this order, which are highlighted in red. We test our prediction method developed using trees of order seven and order eight to accurately predict the four known RNA structures as well as the possibility of other trees in order nine.

Because we have assumed that we do not know which trees are red, our final result will contain only blue and black circles. Unfortunately, Schlick et al. did not publish their findings for order nine; thus, we are unable to compare our results with theirs.

The range of values for the first parameter  $\gamma_L + \gamma_D$  are between 9 to 16 inclusive with a median value of 12. Therefore, all trees with values less than or equal to 12 will be denoted by a blue circle. In the second parameter  $\gamma + \gamma_t + \gamma_{pr}$ , values range from 5 to 17 with a median value of 11; thus, trees with values greater than or equal to 11 will be denoted by a blue circle. The third parameter  $\gamma/(\gamma_L + \gamma_D)$  contains trees with values from 0.06 to 0.44 with a median value of 0.27 where all trees with values greater than or equal to 0.27 will be denoted as a blue circle as well. Three predictions for each tree of order 9, one from each parameter, are illustrated in Table 6. Our final result is the intersection of these three results.

Studying the results on the next page, we see that all four known RNA trees numbered 6, 11, 13, and 27 are blue according to our predictions. Thus, we believe using our method we can accurately predict the existence of other trees of order nine.

Table 6: Results for Trees of Order 9

1	●	●	●	■	25	●	●	●	■
2	●	●	●	■	26	●	●	●	■
3	●	●	●	■	▶ 27	●	●	●	■
4	●	●	●	■	28	●	●	●	■
5	●	●	●	■	29	●	●	●	■
▶ 6	●	●	●	■	30	●	●	●	■
7	●	●	●	■	31	●	●	●	■
8	●	●	●	■	32	●	●	●	■
9	●	●	●	■	33	●	●	●	■
10	●	●	●	■	34	●	●	●	■
▶ 11	●	●	●	■	35	●	●	●	■
12	●	●	●	■	36	●	●	●	■
▶ 13	●	●	●	■	37	●	●	●	■
14	●	●	●	■	38	●	●	●	■
15	●	●	●	■	39	●	●	●	■
16	●	●	●	■	40	●	●	●	■
17	●	●	●	■	41	●	●	●	■
18	●	●	●	■	42	●	●	●	■
19	●	●	●	■	43	●	●	●	■
20	●	●	●	■	44	●	●	●	■
21	●	●	●	■	45	●	●	●	■
22	●	●	●	■	46	●	●	●	■
23	●	●	●	■	47	●	●	●	■
24	●	●	●	■					

### 3 TAKING A CLOSER LOOK

Furthermore, if we take another look at the list of values for trees of order seven that we have determined to be good candidates for novel RNA, we notice that some trees are better candidates than others. For instance, all values for known trees seem not only to be greater than or less than the median for a given parameter but also to be significantly greater than or less than in several cases. This is particularly true for known trees of order eight. For example, looking back at Table 3, we can see that all trees that represent known RNA have values that are significantly greater than or less than the median for a given parameter have already been found. For instance, notice trees 5, 11, and particularly 20. This suggests that not only can we predict which RNA trees we can expect to see in nature, but also we can determine which trees are better candidates than others.

The following determines which trees are best, good, and fair candidates. As stated before, the more extreme the values within the bounds for each parameter a tree has the better its possibility of existing. Therefore, a tree is the best candidate if all three of its parameters are extremes within the bounds. A tree is a good candidate if two of its parameters are extremes, or are near the extreme values, and trees with values that are close or equal to the bounds are considered fair candidates.

For example, tree 42 of order nine, highlighted in dark-blue, is the best candidate within this entire order because all three of its parameter values are extremes within the given bounds. Tree 42 has a value of 9 for the first parameter, the smallest value in this bound. For the second and third parameters tree 42 has values of 17 and .44, respectively, the largest values in each of these parameters' bounds.

Table 7: Best Candidates of Order 9

Tree	$\gamma_L + \gamma_D$	$\gamma + \gamma_t + \gamma_{pr}$	$\frac{\gamma}{\gamma_L + \gamma_D}$	Tree	$\gamma_L + \gamma_D$	$\gamma + \gamma_t + \gamma_{pr}$	$\frac{\gamma}{\gamma_L + \gamma_D}$
1	9	14	.33	25	13	9	.15
2	11	11	.27	26	11	14	.27
3	11	11	.27	27	9	15	.44
4	9	12	.44	28	12	10	.25
5	11	11	.27	29	9	15	.44
6	9	14	.33	30	10	12	.40
7	10	11	.30	31	11	11	.27
8	10	11	.30	32	13	9	.15
9	12	10	.17	33	11	13	.27
10	10	14	.30	34	11	13	.27
11	9	15	.44	35	12	10	.25
12	9	12	.44	36	12	10	.25
13	10	11	.30	37	12	10	.25
14	9	15	.44	38	10	14	.40
15	12	10	.17	39	12	10	.25
16	11	11	.27	40	14	6	.14
17	10	14	.30	41	14	6	.14
18	10	15	.30	42	9	17	.44
19	10	14	.30	43	10	14	.40
20	11	11	.27	44	11	11	.27
21	10	11	.30	45	12	10	.25
22	11	11	.27	46	14	6	.14
23	12	9	.17	47	16	5	.06
24	10	12	.40				

Now consider trees 14 and 29, they too are good candidates; however, by this method they are not as likely to exist as tree 42. These trees, highlighted in light-blue, only have extreme values for the first and third parameters. Moreover, tree 1, for example, is only a fair candidate since the first parameter is the only extreme value. Recall that using our prediction method, the red trees 6, 11, 13, and 27 are included in the set of expected trees.



## 4 CONCLUSION

Although relatively few RNA trees exist in nature, the fact that we were able to accurately predict four of the existing RNA trees of order nine using only domination parameters suggests that this approach has promise in graph theory applications in genomics research. In addition, since our results for orders seven and eight were remarkably close to the results from the RAG database, then this suggest that other parameters may also be useful in predicting novel RNA structures. Furthermore, since there exists polynomial time algorithms for finding the domination number of a tree, then this new application of graph theory, and domination in particular, is an exciting avenue for future research.









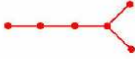
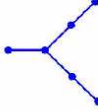
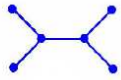
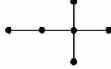

## BIBLIOGRAPHY



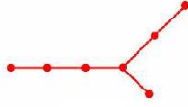
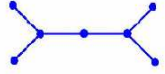
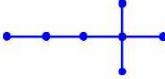

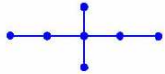
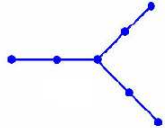
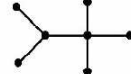
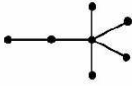

- [1] C.J. Colbourn, P.J. Slater, and L.K. Stewart, Locating-dominating sets in series-parallel networks. *Congr. Numer.* **56** (1987), 135-162.
- [2] A. Finbow and B.L. Hartnell, On locating dominating sets and well-covered graphs. *Congr. Numer.* **65** (1988), 191-200.
- [3] H.H. Gan, S. Pasquali, and T. Schlick, Exploring the repertoire of RNA secondary motifs using graph theory; implications for RNA design. *Nucleic Acids Res.* **31** (2003), 2926-2943.
- [4] J. Gimbel, B. van Gorden, M. Nicolescu, C. Umstead, and N. Vaianna, Location with dominating sets. *Congress Numer.* **151** (2001), 129-144.
- [5] T.W. Haynes, S.T. Hedetniemi, and P.J. Slater, *Fundamentals of Domination in Graphs*, Marcel Dekker, New York, 1998.
- [6] T.W. Haynes, S.T. Hedetniemi, and P.J. Slater (eds), *Domination in Graphs: Advanced Topics*, Marcel Dekker, New York, 1998.
- [7] M.A. Henning and O.R. Oellermann, Metric-Locating-Dominating Sets in Graphs. To appear in *Ars Combin.*
- [8] L. Lesniak and G. Chartrand, *Graphs and Digraphs*, 3rd ed, Chapman and Hall, New York, 1996.
- [9] D.F. Rall and P.J. Slater, On location-domination numbers for certain classes of graphs. *Congr. Numer.* **45** (1984), 97-106.

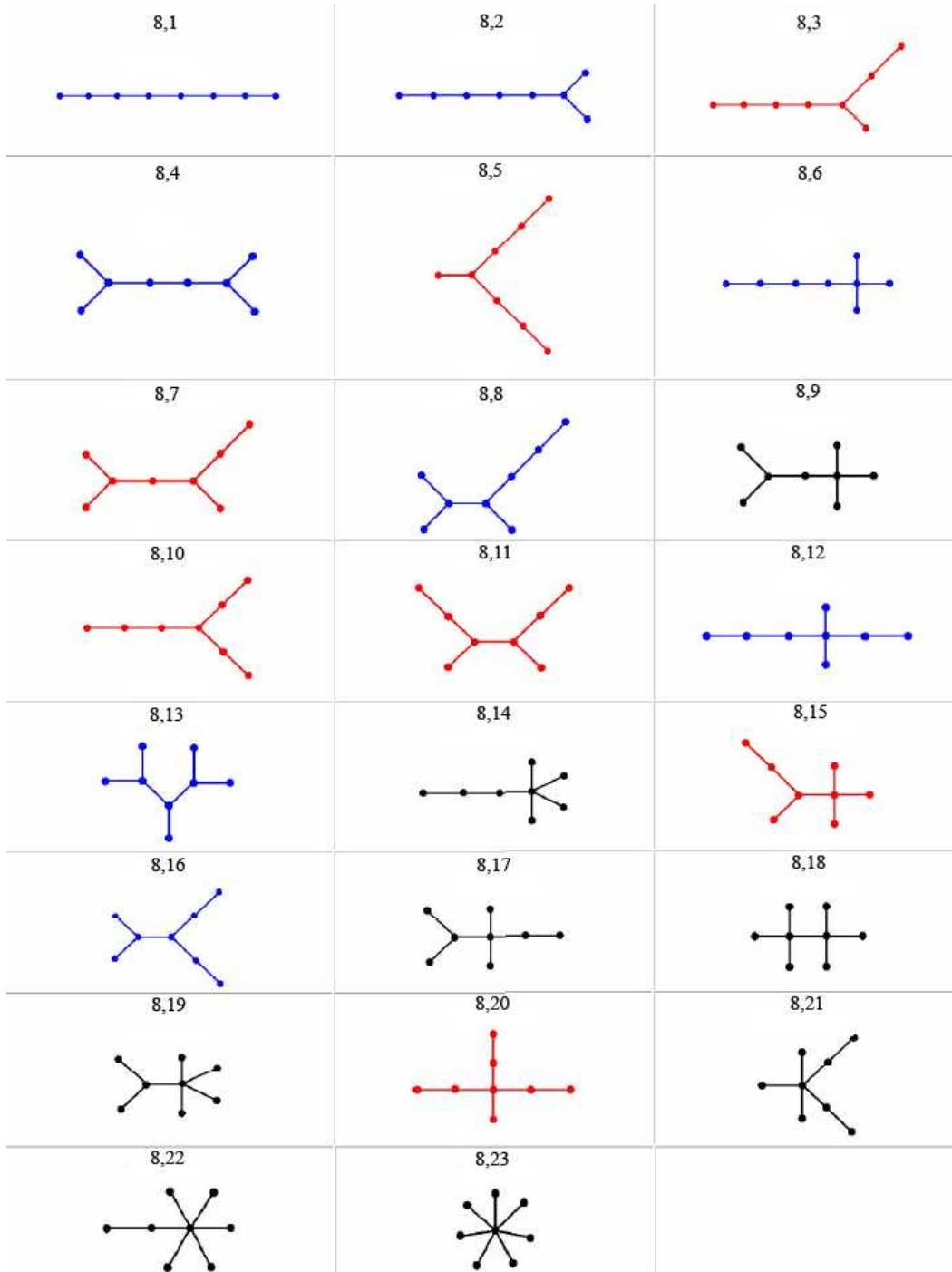
- [10] T. Schlick, RAG: RNA-As-Graphs. *BMC Bioinformatics*. **5** (2004), 1471-2105.
- [11] T. Schlick, RAG: RNA-As-Graphs database—concepts, analysis, and features. *BMC Bioinformatics*. **20** (2004), 1285-1291.
- [12] P.J. Slater, Leaves of trees. *Ars Combin.* **2** (1976), 549-559.
- [13] P.J. Slater, Dominating and location in acyclic graphs. *Networks* **17** (1987), 55-64.
- [14] P.J. Slater, Dominating and reference sets in graphs. *J. Math. Phys. Sci.* **22** (1988), 445-455.
- [15] P.J. Slater, Locating dominating sets and locating-dominating sets. In Y. Alavi and A. Schwenk, editors, *Graph Theory, Combinatorics, and Applications, Proc. Seventh Quad. Internat. Conf on the Theory and Applications of Graphs*, pages 1073-1079. John Wiley & Sons, Inc., 1995.

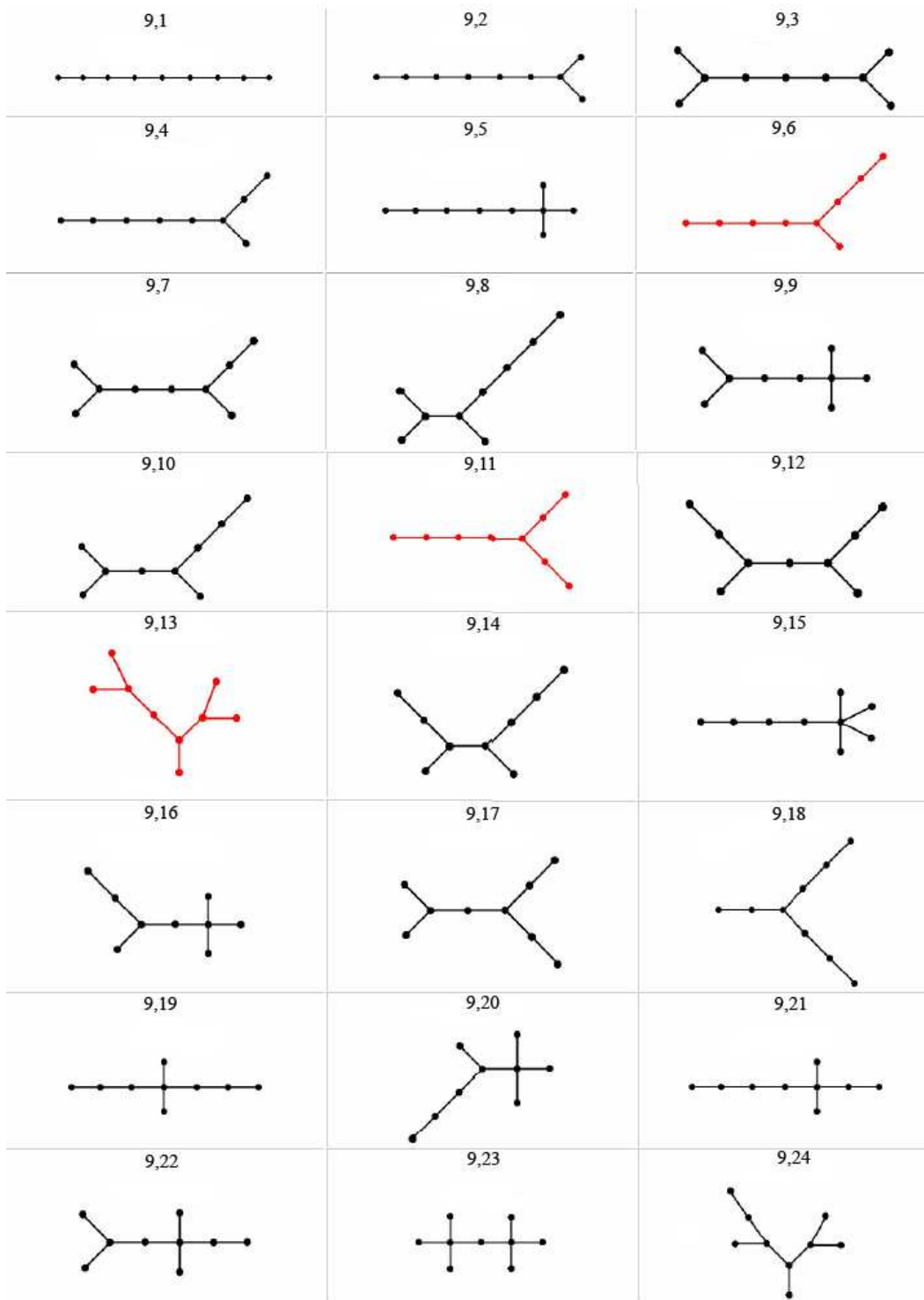
# APPENDICES

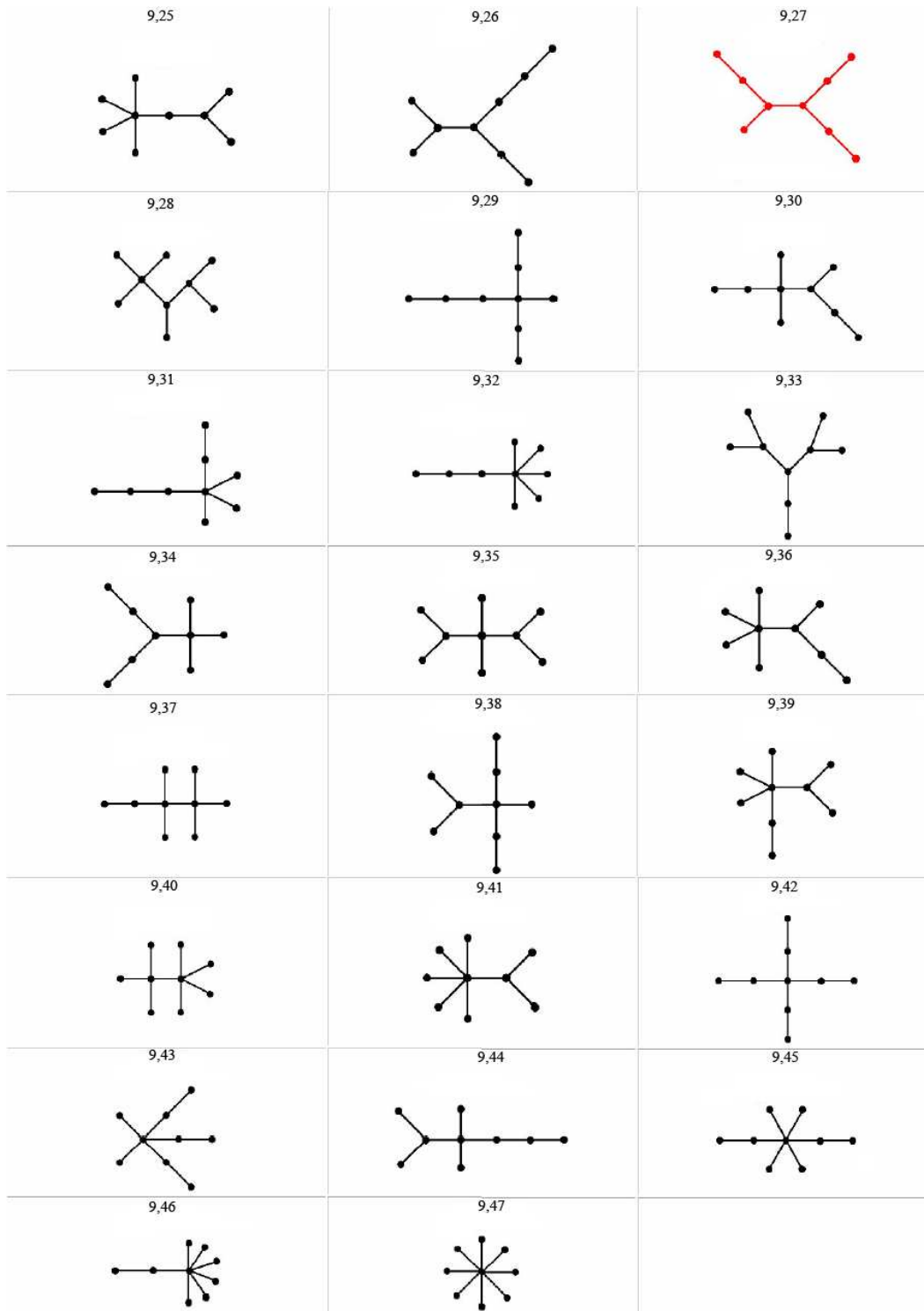
List of trees for orders 2 through 9

2,1 		
3,1 		
4,1 	4,2 	
5,1 	5,2 	5,3 
6,1 	6,2 	6,3 
6,4 	6,5 	6,6 

<p>7,1</p> 	<p>7,2</p> 	<p>7,3</p> 
<p>7,4</p> 	<p>7,5</p> 	<p>7,6</p> 
<p>7,7</p> 	<p>7,8</p> 	<p>7,9</p> 
<p>7,10</p> 	<p>7,11</p> 	









## VITA

TRAVIS R. COAKE

### **Education**

- Bachelors of Science Degree in Mathematics, Radford University,  
December 2002.
- Masters of Science Degree in Mathematics, East Tennessee State University,  
May 2005.

### **Professional Experience**

- Graduate Assistant/Teaching Assistant  
East Tennessee State University, Johnson City, TN, 2003-2004.