12-2014

Linear and Nonlinear Trees: Multiplicity Lists of Symmetric Matrices

Eric Wityk
College of William and Mary

Linear and Nonlinear Trees: Multiplicity Lists of Symmetric Matrices

Eric Wityk

December 10, 2014

Acknowledgments

I would like to thank my thesis advisor Professor Charles Johnson for helping and supporting me throughout this past year, and Professor Ryan Vinroot for guiding me in my search for research topics. I would also like to thank Joe Corliss for providing me with his previous work on this topic, and Eric Walter for introducing me to and guiding my use of the SciClone. Finally, I would like to thank Professor Anke van Zuylen and Professor Sean Tarter for serving on my honors committee.

Abstract

Let A be a real symmetric matrix whose graph is a tree, T. If T is a linear tree (meaning all vertices with degree 3 or larger lie on the same induced path), then we can use a "Linear Superposition Principle" to determine all possible multiplicities of eigenvalues of A. If T is a nonlinear tree, we must use other ad hoc methods. I utilize these methods to compute all possible multiplicity lists of trees on 12 vertices, and augment an existing multiplicities database. This database allows us to examine of the effects that the structure of tree can have on a multiplicity list. Then, I investigate the enumeration of linear and nonlinear trees, and examine the ratio of nonlinear trees to total trees.

Contents

Chapter 1

Linear Superposition Principle

1.1 Introduction

Eigenvalues are undoubtedly among the most important properties of a matrix. Any n-by-n matrix has n eigenvalues counted by multiplicity. For general matrices, there is little that can be said about these multiplicities. However, we can more precisely describe the possibile multiplicities of eigenvalues of real symmetric matrices whose graph is a tree. We will focus on how the structure of such trees restrict the possiblities multiplicities of eigenvalues of these matrices.

Research into the relationship between trees and eigenvalue multiplicities motivated the creation of a "multiplicity database." This database catalogs all possible combinations of multiplicities of eigenvalues of these matrices for all trees on at most 11 vertices, as well as some useful information regarding the structure of these trees [1]. The aim of this database is to aid in formulating and testing conjectures on this topic. I extended the scope of the database to include all trees on 12 vertices, which more than doubled it in size. I appealed to an algorithm based on the *Linear Superposition Principle*[6] as my primary tool in determining possible multiplicity lists. If some combination of multiplicities is realizable, then the Linear Superposition Principle generates it. However, this tool can only be used on a particular type of tree known as *linear trees*.

The same process could be used to expand the multiplicity database further. However, the practical limitations of such an activity became of interest. The total number of trees on n vertices grows quickly as n grows. To illustrate, there are 551 trees on 12 vertices, and 1301 trees on 13 vertices. In addition, the algorithmic approach used on linear trees is more time consuming on larger trees. More alarmingly, there is no known algorithmic approach for determining possible combinations of multiplicities on trees that are not linear (*nonlinear trees*), so ad hoc manual techniques must be used. The research provided in this paper provides strong evidence the ratio of nonlinear trees to total trees converges asymptotically to 1 as the total number of vertices increases. As a result, replicating this process for larger trees would result in substantial manual work.

In Chapter 1, we discuss the process of expanding the multiplicity database, as well as the helpful tools that were utilized along the way. We then use this database to examine a set of conjectures [1]. In Chapter 2, we count linear trees on n vertices using generating functions. These generating functions yield closed formulas, and provide us with some information on the types of linear trees on n vertices. In Chapter 3, we discuss nonlinear trees and their structures. We will conclude that, with some restrictions, the number of nonlinear trees make up the vast majority of the total number of trees.

1.2 Definitions

Let $A = (a_{ij})$ by an $n \times n$ matrix. The graph of A is a graph G on n vertices, where there is an edge between vertices i and j if and only if $a_{ij} \neq 0$. We denote the vertex set of G as $V(G)$. We call the *multiplicity lists* of a graph G a collection of lists that describe all possibilities of combinations of multiplicities of eigenvalues that can be realized on a matrix whose graph is G. We will refer to the multiplicity list of all 1's (meaning all eigenvalues are distinct) as the *trivial multiplicity list*.

Trees have various properties that are of interest here. Let T be a tree, and let $V(T) = \{v_1, v_2, \ldots, v_n\}$ be its vertex set. We denote a path that begins on v_1 and ends on v_m as $v_1 - v_2 - \ldots - v_m$. We call the maximum length of the shortest path between any two vertices the *diameter* of the graph. All paths in a tree are unique, so the diameter of T is the length of the longest path. For our purposes we will count the length of the path on vertices, not edges. For example, if d is the diameter of T, then there is some path $v_{i_1} - v_{i_2} - \ldots - v_{i_d}$ which is the longest path in T. Also of critical importance is the *path cover number* of T. The path cover number of a tree is the number of the fewest disjoint paths such that every vertex in the tree is on one of these paths.

We say that a vertex is a *high degree vertex (HDV)* if and only if the degree of the vertex is 3 or greater. The only trees that do not contain high degree vertices are paths. On a path it is only possible to realize the trivial multiplicity list. So, the presence of high degree vertices is an immediate necessary condition for the realization of nontrivial multiplicity lists.

We define a *star* as a tree on $n \geq 4$ vertices, $n-1$ of which are adjacent to the same high degree vertex. This high degree vertex is referred to as the *central vertex.* The notion of a star can be expanded to include *generalized stars.* Consider $k \geq 3$ paths of length $l_1, l_2, \ldots, l_k$. A generalized star is a tree with exactly one high degree vertex which is adjacent to one endpoint of each of the k paths. We call each of these paths the *branches* of that generalized star. Therefore, the branch lengths of a generalized star sum to $n-1$. We denote any generalized star on n vertices as $T = [l_1, l_2, \ldots, l_k]$ where each l_i is the length of a particular branch, and $l_i \geq l_{i+1}$ for all $i \in \{1, 2, \ldots, k\}$.

Finally, we define a *linear tree* as a tree where all high degree vertices lie on the same path. We call this path the *induced path* of the linear tree. A *k-linear tree* is a linear tree with k high degree vertices. Let L be a linear tree. Let $v_1, v_2, \ldots, v_k$ be each high degree vertex in L such that there exists a path between any v_i to v_{i+1} for all $i \in \{1, 2, \ldots, k-1\}$, and no other high degree vertices lie on this path. Denote the length of this path without including v_i and v_{i+1} as s_i. Note that we allow s_i to be 0 in the case that v_i is adjacent to

v_{i+1}. Let T_i refer to the subtree that contains v_i and any pendant branches of v_i that do not include neither any high degree vertices, nor vertices lying on paths between two high degree vertices. Although T_i may be a path, in the context of a linear tree we refer to these subtrees as generalized stars. We denote such a linear tree L as $L = (T_1, s_1, T_2, s_2, \ldots, s_{k-1}, T_k)$.

Now, consider a linear tree $L = (T_1, s_1, T_2, s_2, \ldots, s_{k-1}, T_k)$. Because v_1 is adjacent to either v_2 or is the endpoint of a path between v_1 and v_2, its degree in L is higher than its degree in T_1 by 1. Therefore, the minimal degree of v_1 in T_1 that assures that it is a high degree vertex in L is 2. A similar argument can be made for v_k. If $k \geq 3$, let v_i be any high degree vertex other than v_1 or v_k. Then, v_i is either adjacent to v_{i-1} or it is adjacent to the endpoint of a path between v_i and v_{i-1}, and v_i is either adjacent to v_{i+1} or it is adjacent to the endpoint of a path between v_i and v_{i+1}. So, the minimal degree of v_i in T_i that assures that it is a high degree vertex in L is 1. We refer to T_1 and T_k as seither *exterior stars* or *peripheral stars*, and any other T_i where $i \neq 1, k$ as an *interior star*.

If a tree is not a linear tree, then we refer to it as a *nonlinear tree*. The smallest nonlinear

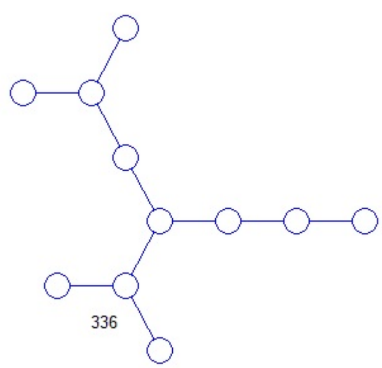

Figure 1.1: This linear tree contains three high degree vertices. There is a path of length 1 between two HDVs. The lengths of the branches of the peripheral stars are 1, and the length of the branch of the interior star is 3. So, we would denote this tree as either $([1,1], 0, [3], 1, [1,1])$ or $([1,1], 1, [3], 0, [1,1])$.

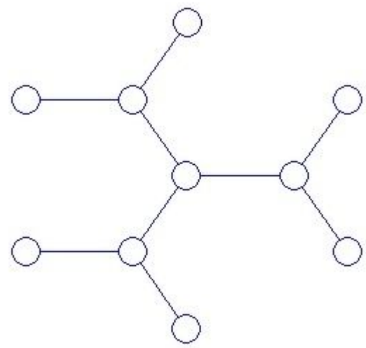

Figure 1.2: This nonlinear tree occurs on 10 vertices, which is the first time a nonlinear tree occurs as n increases.

tree is a 10 vertex tree and is shown in Figure 1.2. Unlike linear trees, there is a much richer variety of nonlinear trees, and as a result they are much more difficult to denote using a canonical notation. In addition, tools which are useful for linear trees, like the Linear Superposition Principle, do not easily generalize to nonlinear trees because of their complex structures. Despite these difficulties, we will make an effort to describe the various structures and commonalities that characterize nonlinear trees in Chapter 3.

1.3 Essential Background

We will discuss the tools and techniques that were used to determine the multiplicity lists of trees on 12 vertices. To expand the database, I primarily relied on the *method of assignments* and the Linear Superposition Principle. However, there is useful information that can be gathered from the path cover number, the diameter, and the degrees of vertices in a tree.

Let T be any tree on at least 2 vertices, and let $U(T)$ be the number of 1's in any possible multiplicity list of T. The smallest and largest eigenvalue of any Hermitian A whose graph is T must be unique. So, $U(T) \geq 2$ [3]. In addition, the path cover number of T is the largest multiplicity that is realized in the collection of multiplicity lists of T [4]. The diameter is a lower bound for the minimum number of distinct eigenvalues of a Hermitian matrix whose graph is a tree [5]. However, it is not guaranteed that this lower bound is always achieved.

Let A be any $n \times n$ matrix, and let $\alpha \subseteq \{1, 2, \ldots, n\}$. We use the notation $A(\alpha)$ to denote the square matrix created by the removal of the ith row and ith column of the matrix A for each $i \in \alpha$. The graphical representation of $A(\alpha)$ is simply the graph of A with the deletion of each vertex i and its incident edges. Conversely, $A[\alpha]$ is the matrix created by the removal of the jth row and jth column of A for all $j \in \{1, 2, \ldots, n\}/\alpha$. The interlacing inequalities classify how each multiplicity in A can change with the removal of a row and column.

Theorem 1.3.1 ([11]). *Let $A \in M_n$ be Hermitian, and let $i \in \{1, 2, \ldots, n\}$. Let $\lambda_1(A) \geq \lambda_2(A) \geq \ldots \geq \lambda_n(A)$ denote the eigenvalues of A. Then, the following inequality must hold: $\lambda_1(A) \geq \lambda_1(A(i)) \geq \lambda_2(A) \geq \lambda_2(A(i)) \geq \ldots \geq \lambda_{n-1}(A) \geq \lambda_{n-1}(A(i)) \geq \lambda_n(A)$.*

Let $m_A(\lambda)$ be the multiplicity of some eigenvalue λ in A. From the interlacing inequalities, we can conclude that $m_A(\lambda) - 1 \leq m_{A(i)}(\lambda) \leq m_A(\lambda) + 1$ for any λ. Furthermore, if the jth smallest eigenvalue in A increases in multiplicity in $A(i)$, then the $(j+1)$th and $(j-1)$th smallest eigenvalues cannot also increase in multiplicity in $A(i)$ [3]. We define an *upward multiplicity* of A as a multiplicity for which $m_A(\lambda) + 1 = m_{A(i)}(\lambda)$ for some i. An upward multiplicity m is denoted $\hat{m}$. The vertex whose removal results in an increase in the multiplicity of λ is called a *Parter vertex*.

Using the Parter-Weiner Theorem, we can understand which vertices are Parter vertices, and how their removal changes the multiplicities of the eigenvalues. The Parter-Weiner Theorem is stated as follows:

Theorem 1.3.2 ([3]). *Let A be a Hermitian matrix whose graph is a tree T. Let v be a vertex such that λ is an eigenvalue of both A and $A(v)$. Then:*

1. *There is a Parter vertex, v' for λ in T.*

2. *If* $m_A(\lambda) = 1$*, then there is a Parter vertex* v' *that has degree at least* 2*, and the two separated components of the graph of* $A(v)$*,* T_1 *and* T_2*, are such that* λ *has a multiplicity of* 1 *in both* $A[T_1]$ *and* $A[T_2]$*.*

3. *If* $m_A(\lambda) \geq 2$*, then there is a Parter vertex* v' *has degree at least* 3*, and at least three components of the graph resulting in its removal,* T_1, T_2 *and* T_3 *have* λ *appear as an eigenvalue of* $A[T_1]$*,* $A[T_2]$ *and* $A[T_3]$ *at least once.*

The Parter Theorem is the key theorem for using a powerful tool known as the method of assignments. The technique is employed as follows. We identify all high degree vertices in the tree. After removing some subset of these vertices, we are left with several disjoint components. We can then assign an eigenvalue to some or all of these components. If there are c components assigned to an eigenvalue after removal of p Parter vertices, then this eigenvalue will have a multiplicity of $c - p$. We can assign multiple distinct eigenvalues to the same subtrees as long as we do not assign more eigenvalues than there are vertices in that subtree. If some subtree S on m vertices is not overloaded, then the number of subtrees of S that are assigned an eigenvalue minus the number of Parter vertices used to make these assignments in S, cannot exceed m [10].

The method of assignments assists in determining the multiplicity lists, especially for nonlinear trees. However, it is not guaranteed that an assignment of eigenvalues can always be realized by a matrix. There is at least one known example of an assignment on a tree on 13 vertices where the ordering of the eigenvalues violates the interlacing inequalities [10]. So, the method of assignments must be used carefully, especially as the number of vertices in a tree increases.

So far, we have made no restriction on the order in which we list the multiplicities. Sometimes it can be useful to list multiplicities with respect to the order of their eigenvalues. We call such a multiplicity list an *ordered multiplicity list*. The ordered multiplicity list of a matrix Hermitian A whose distinct eigenvalues are $\lambda_1 < \lambda_2 < \ldots < \lambda_k$ is $(m_A(\lambda_1), m_A(\lambda_2), \ldots, m_A(\lambda_k))$. Because the largest and smallest eigenvalues are distinct,

$m_A(\lambda_1) = m_A(\lambda_k) = 1$. Often times we will also differentiate between upward and nonupward multiplicities when discussing ordered multiplicity lists.

Integer partitions are critically important for determining multiplicity lists and counting trees. Recall that a generalized star on n vertices has branch lengths $l_1, l_2, \ldots, l_m$, and $\sum_{i=1}^{m} l_i = n-1$. So, the branch lengths of a generalized star form a partition of the integer $n-1$. If $(l_1, l_2, \ldots l_m)$ is an integer partition of $n-1$, then the conjugate partition, denoted $(l_1, l_2, \ldots, l_m)^* = (l'_1, l'_2, \ldots, l'_{m'})$, where l'_i is the total number of $l_j \geq i$ for $j \in \{1, 2, \ldots, m\}$. The conjugate partition of an integer is another partition of that same integer.

Let $(a_1, a_2, \ldots, a_m)$ and $(b_1, b_2, \ldots, b_n)$ be two vectors of numbers. Without loss of generality, if $n \geq m$, then we define $a_i = 0$ for all $i \in \{m+1, m+2, \ldots, n\}$. Then, we say that $(a_1, a_2, \ldots, a_m)$ majorizes $(b_1, b_2, \ldots, b_n)$ if $\sum_{i=1}^{j} a_i > \sum_{i=1}^{j} b_i$ for all $j \in \{1, 2, \ldots, n-1\}$, and $\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} b_i$. We denote this as $(a_1, a_2, \ldots, a_m) \succ (b_1, b_2, \ldots, b_n)$. We also define weak majorization among these two vectors as $\sum_{i=1}^{j} a_i \geq \sum_{i=1}^{j} b_i$ for all $j \in \{1, 2, \ldots, n-1\}$, and $\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} b_i$. Weak majorization is denoted $(a_1, a_2, \ldots, a_m) \succeq (b_1, b_2, \ldots, b_n)$.

Theorem 1.3.3. *Let $T = [l_1, l_2, \ldots, l_m]$ be a generalized star on n vertices. Then, there exists a symmetric real matrix A that has eigenvalues with the ordered multiplicity list $(q_1, \ldots q_r)$ if and only if:*

1. $\sum_{i=1}^{r} q_i = n$
2. *if q_i is an upward multiplicity, then q_{i-1} and q_{i+1} are non-upward multiplicities.*
3. *Let $q_{i_1} \geq \ldots \geq q_{i_h}$ be upward multiplicities of A. Then, $(q_{i_1}+1, \ldots, q_{i+h}+1) \preceq (l_1, \ldots, l_m)^*$.*

[2]

This theorem gives us an easy way to compute the multiplicity lists of a generalized star. For example, consider the generalized star on 7 vertices $T = [3, 2, 1]$, depicted in Figure 1.3. The following vectors are majorized by $(3, 2, 1)$: $(3, 2, 1)$, $(3, 1, 1, 1)$, $(2, 2, 2)$,

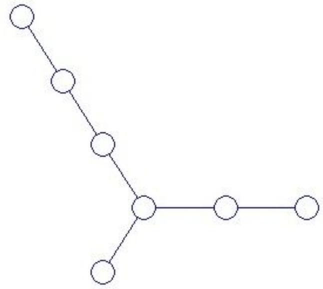

Figure 1.3: The generalized star $[3,2,1]$ has the unordered multiplicity lists $(2,1,1,1,1,1)$ and $(1,1,1,1,1,1,1)$. The complete set of ordered upward multiplicity lists is $(1,\hat{2},1,\hat{1},1,\hat{0},1)$, $(1,\hat{2},1,\hat{0},1,\hat{1},1)$, $(1,\hat{1},1,\hat{2},1,\hat{0},1)$, $(1,\hat{1},1,\hat{0},1,\hat{2},1)$, $(1,\hat{0},1,\hat{2},1,\hat{1},1)$, $(1,\hat{0},1,\hat{1},1,\hat{2},1)$, $(1,\hat{2},1,\hat{0},1,\hat{0},1,\hat{0},1)$, $(1,\hat{0},1,\hat{2},1,\hat{0},1,\hat{0},1)$, $(1,\hat{0},1,\hat{0},1,\hat{2},1,\hat{0},1)$, $(1,\hat{0},1,\hat{0},1,\hat{0},1,\hat{2},1)$, $(1,\hat{1},1,\hat{1},1,\hat{1},1)$, $(1,\hat{1},1,\hat{1},1,\hat{0},1,\hat{0},1)$, $(1,\hat{1},1,\hat{0},1,\hat{1},1,\hat{0},1)$, $(1,\hat{1},1,\hat{0},1,\hat{0},1,\hat{1},1)$, $(1,\hat{0},1,\hat{1},1,\hat{1},1,\hat{0},1)$, $(1,\hat{0},1,\hat{1},1,\hat{0},1,\hat{1},1)$, $(1,\hat{0},1,\hat{0},1,\hat{1},1,\hat{1},1)$, $(1,\hat{1},1,\hat{0},1,\hat{0},1,\hat{0},1,\hat{0},1)$, $(1,\hat{0},1,\hat{1},1,\hat{0},1,\hat{0},1,\hat{0},1)$, $(1,\hat{0},1,\hat{0},1,\hat{1},1,\hat{0},1,\hat{0},1)$, $(1,\hat{0},1,\hat{0},1,\hat{0},1,\hat{1},1,\hat{0},1)$, $(1,\hat{0},1,\hat{0},1,\hat{0},1,\hat{0},1,\hat{1},1)$, and $(1,\hat{0},1,\hat{0},1,\hat{0},1,\hat{0},1,\hat{0},1,\hat{0},1)$.

$(2,2,1,1)$, $(2,1,1,1,1)$, and $(1,1,1,1,1,1)$. Each number in these vectors indicate an upward multiplicity incremented by one. So, the possible upward multiplicities are $(\hat{2},\hat{1},\hat{0})$, $(\hat{2},\hat{0},\hat{0},\hat{0})$, $(\hat{1},\hat{1},\hat{1})$, $(\hat{1},\hat{1},\hat{0},\hat{0})$, $(\hat{1},\hat{0},\hat{0},\hat{0},\hat{0})$ and $(\hat{0},\hat{0},\hat{0},\hat{0},\hat{0},\hat{0})$. To generate a complete set of upward multiplicities, we must insert non-upward 1's between each upward multiplicity for every permutation of the lists of upward multiplicities. The complete set of ordered upward multiplicity lists is given in Figure 1.3.

The *Linear Superposition Principle* (often abbreviated LSP) is a tool which allows us to compute multiplicity lists of linear trees.. The Linear Superposition Principle is implemented as follows. Let $L = (T_1, s_1, T_2, s_2, \ldots, s_{k-1}, T_k)$ be a linear tree. Let $\hat{b_i}$ be any ordered upward multiplicity list for T_i, and let $\hat{c_j}$ be a list of s_j non-upward ones. Then, we can create a grid subject to the following conditions.

1. Augment each $\hat{b_i}$ and $\hat{c_j}$ into b_i^+ and c_j^+ by inserting non-upward 0's anywhere in these lists while keeping all lists the same length.

2. The rows of the grid in order will be $b_1^+, c_1^+, b_2^+, c_2^+, \ldots, c_{k-1}^+, b_k^+$.

3. The elements in any column of the grid are not all non-upward 0's.

4. Moving along any column of the grid, between any two non-upward 1's, there must be at least one upward multiplicity.

Then the column sums of this grid is a multiplicity list of L generated by the Linear Superposition Principle. Any multiplicity list that can be realized by a Hermitian matrix whose graph is L will be generated by the Linear Superposition Principle [6].

The Linear Superposition Principle is best understood through an example. Consider the tree $L = ([1,1,1], 0, [1,1], 2, [1,1])$. We will use the Linear Superposition Principle to generate a potential multiplicity list of L. First, we must extract an ordered multiplicity list of $[1,1,1], [1,1]$ and $[1,1]$. We will choose the ordered multiplicity lists $(1, \hat{2}, 1)$, $(1, \hat{0}, 1, \hat{0}, 1)$ and $(1, \hat{0}, 1, \hat{0}, 1)$. Then, we can set up a grid to determine a possible multiplicity list. Figure 1.4 depicts one possible way to arrange these multiplicities. In this case, we can extract $(1,1,1,1,1,3,2,1,1)$ as a possible ordered multiplicity list of L.

It is unknown whether or not the Linear Superposition Principle is a sufficient condition for realizable multiplicity lists for all linear trees. There are some special cases where sufficiency has been proven. A linear tree with 2 high degree vertices (sometimes called a double

	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9
b_1^+	0	0	0	1	0	$\hat{2}$	1	0	0
c_1^+	0	0	0	0	0	0	0	0	0
b_2^+	1	0	0	$\hat{0}$	1	0	$\hat{0}$	0	1
c_2^+	0	0	1	0	0	0	1	0	0
b_3^+	0	1	$\hat{0}$	0	0	1	$\hat{0}$	1	0
	1	1	1	1	1	3	2	1	1

Figure 1.4: This grid demonstrates the Linear Superposition Principle. Note that every other row is an augmented ordered upward multiplicity list of a generalized star. Furthermore, the only column where two non-upward multiplicities occur is the column for λ_7. However, they are separated by an upward $\hat{0}$ in the row for b_2^+, so this does not violate the 4th condition of the Linear Superposition Principle. The bottom row is the multiplicity list for L generated by LSP.

generalized star) has a realizable multiplicity list if an only if it is generated by LSP. LSP is also sufficient for linear trees where the branch lengths of each generalized star is 1. These trees are called linear trees of depth 1. (Alternatively, we could define this as a tree where all vertices are at most one edge away from a single induced path that contains all high degree vertices.)[6]

1.4 Multiplicity Database and Applications

The Linear Position Principle is a powerful tool, especially aided by automation. Previous research led to the development of a Matlab code which computes the multiplicity lists generated by LSP. Using this, a database was created detailing a complete set of multiplicity lists for all trees on 11 vertices and fewer [1].

I obtained a copy of the database and the code with the goal of expanding this database to include trees on 12 vertices. Utilizing the SciClone supercomputer and the LSP code, I computed the multiplicity lists for all linear trees [12]. Using the method of assignments, I determined the multiplicity lists for all 19 nonlinear trees, and double checked the multiplicity lists for linear trees. As a final test, I ensured that this data did not contradict any other proven results.

In the multiplicity database, each multiplicity list is abbreviated in the following way. A multiplicity list in the database is denoted $m_1 m_2 \ldots m_r$ where each $m_1, m_2, \ldots, m_r > 1$. For any multiplicity list for a tree on n vertices, there are $n - \sum_{i=1}^{r} m_i$ ones that are implicitly included in the unordered list.

This database is a powerful tool for multiplicity research. The database information on the path cover number, diameter, $U(T)$ and degree counts for all trees, which can be used to quickly test new conjectures and open questions. I introduced a new field, which flags nonlinear trees, to aid in research on the differences between linear and nonlinear trees. Once the database was expanded to include trees on 12 vertices, I used it to examine various conjectures [1].

Conjecture 1.4.1. *For any linear tree T, the length of the shortest multiplicity list of T equals the diameter of T.*

Using a technique known as branch duplication, it has been shown that this conjecture is true whenever the diameter is at most 6. However, when the diameter is 7 or larger, this conjecture is not true for all nonlinear trees. Reassuringly, I confirmed this conjecture for all 116 trees of diameter at most 6 on 12 vertices. I also confirmed this conjecture for the remaining 385 trees with a diameter of 7 or greater. Previous work had validated this conjecture on all trees with 11 vertices or fewer,[1] so this result shows this conjecture holds on all trees with 12 vertices or fewer.

Conjecture 1.4.2. *For any linear tree T, there exists a minimal length multiplicity list of T that attains $U(T)$.*

This conjecture was validated for all trees on fewer than 12 vertices [1]. However, on 12 vertices, there are some counterexamples. The generalized star $[4, 1, 1, 1, 1, 1, 1, 1]$ has a diameter of 6, so there are at least 6 distinct eigenvalues for such a matrix. Furthermore, this minimum number of distinct elements is attained by the unordered multiplicity list $(7, 1, 1, 1, 1, 1)$. However, $U(T)$ for this tree is 4. This is attained by the multiplicity list $(4, 2, 2, 1, 1, 1, 1)$. However, there are 7 distinct eigenvalues in this multiplicity list. There are no multiplicity lists which contain both four 1's and 6 distinct elemenets.

In addition, the linear tree $([3, 1, 1, 1], 0, [1, 1, 1, 1])$ is another counterexample. The diameter of this tree is 6, which is attained by the unordered multiplicity list $(4, 4, 1, 1, 1, 1)$, but $U(T) = 2$ is attained by the multiplicity list $(2, 2, 2, 2, 2, 1, 1)$. There is no multiplicity list with both two 1's and 6 distinct elements.

Finally, the linear tree $([3, 2, 1], 0, [1, 1, 1, 1])$ is a third counterexamples on 12 vertices. The diameter of this tree is 6, which is attained by the unordered multiplicity list $(4, 3, 2, 1, 1, 1)$, but $U(T) = 2$ is achieved by the list $(2, 2, 2, 2, 2, 1, 1)$. There is no multiplicity list that both has two 1's and 6 distinct elements. Thus, we have proven that this conjecture is false.

Conjecture 1.4.3. *Given k high degree vertices of degree $d_1, \ldots, d_k$, there is a multiplicity list whose only entries greater than 1 are $d_1 - 1, \ldots, d_k - 1$.*

This conjecture has been proven for all linear trees, and all trees on fewer than 12 vertices [1]. However, this remains an open question for nonlinear trees. I validated this conjecture on the 19 nonlinear trees on 12 vertices. So, this conjecture holds for all trees on 12 vertices or fewer.

Conjecture 1.4.4. *Let $U(T)$ be the smallest number of 1s can occur in a multiplicity list of T. Let $D_i(T)$ refer to the number of vertices of degree i in T. Then, $U(T) \leq 2 + D_2(T)$.*

When $D_2(T) = 0$, this formula suggests $U(T) \leq 2$. Since we know $U(T)$ is bounded below by 2, this would mean that $U(T) = 2$. It has been shown that Conjecture 1.4.3 guarantees that this inequality will hold. Since Conjecture 1.4.3 held on all trees on 12 vertices or fewer, this conjecture does as well.

Conjecture 1.4.5. *Let T be a tree, and let $(m_1, m_2, \ldots, m_r)$ be a realizable unordered multiplicity list of T. Then, for any $j \in \{1, 2, \ldots, r\}$ such that $m_j \geq 2$, the unordered multiplicity list $(m_1, m_2, \ldots, m_{j-1}, m_j - 1, m_{j+1}, \ldots, m_r, 1)$ is realizable in T.*

This conjecture has been proven for linear trees in which the Linear Superposition Principle is a sufficient condition for a realizable multiplicity list [1]. Since the Linear Superposition

Figure 1.5: These trees are all counterexamples to conjecture 1.4.2. The multiplicity lists given are in the same notation that they appear in the multiplicity database.

Tree	Multiplicity Lists
$[4, 1, 1, 1, 1, 1, 1, 1]$	7; 6; 52; 5; 43; 422; 42; 4; 332; 33; 322; 32; 3; 222; 22; 2
$([3, 1, 1, 1], 0, [1, 1, 1, 1])$	6; 52; 5; 44; 432; 43; 422; 42; 4; 332; 33; 3222; 322; 32; 3; 22222; 2222; 222; 22; 2
$([3, 2, 1], 0, [1, 1, 1, 1])$	52; 5; 432; 43; 422; 42; 4; 332; 33; 3222; 322; 32; 3; 22222; 2222; 222; 22; 2

Principle is a sufficient condition in the case where $n = 12$, this conjecture is true for all linear trees on 12 vertices. In addition, I confirmed this conjecture for the nonlinear trees on 12 vertices. Since the database was previously used to confirm the conjecture for all trees on fewer than 12 vertices[1], we can conclude that this conjecture is valid for all trees on at most 12 vertices.

Chapter 2

Counting Linear Trees

To expand the multiplicity database, it was necessary to generate all linear trees on 12 vertices. This raised the natural question of how many linear trees there are on n vertices. In this section, I attempt to answer this question using generating functions. The result is a formula for determining the total number of k-linear trees on n vertices.

2.1 Linear Symmetry

Suppose we have the linear tree $(T_1, s_1, T_2, s_2, \ldots s_{k-1}, T_k)$. We call this tree *linearly symmetric* if and only if

- For each $i \in \{1, 2, \ldots, k\}$, $T_i \cong T_{k-i+1}$
- For each $j \in \{1, 2, \ldots, k-1\}$, $s_j = s_{k-j}$.

If k is even, then $j = \frac{k}{2}$ implies $k - j = j$. Similarly, when k is odd, then $i = \frac{k+1}{2}$ implies $k - i + 1 = i$. In these cases, we call T_i and s_j the *central components* of the tree. These components have the property that they can be different from every other component in a linearly symmetric tree, because they have no symmetric counterpart. Finally, if a tree is not linearly symmetric, then it is *linearly asymmetric*.

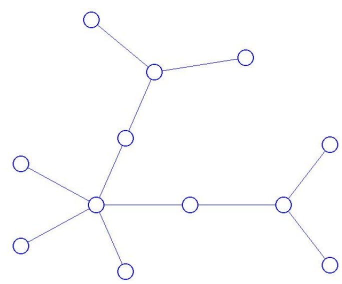

Figure 2.1: *The tree* $([1,1],1,[1,1,1],1,[1,1])$ *is linearly symmetric. The component* $[1,1,1]$ *is the central component.*

Notice that a linearly asymmetric tree can be denoted in two different ways using the canonical notation: $(T_1, s_1, T_2, s_2, \ldots, s_{k-1}, T_k)$ or $(T_k, s_{k-1}, \ldots, s_2, T_2, s_1, T_1)$. We refer to these two notations as *reflections* of the tree. We also define the reflection of T_i, as T_{k-i+1}, and the reflection of s_j as s_{k-j}. Observe that if $(T_1, s_1, T_2, s_2, \ldots, s_{k-1}, T_k)$ is linearly symmetric then both reflections of this tree are the same. So, linearly asymmetric trees have two reflections, whereas linearly symmetric trees have one reflection.

Let $r_{n,k}$ be the number of reflections of linear trees on n vertices with k HDVs. Let $s_{n,k}$ be the number of linearly symmetric trees on n vertices with k HDVs. Note that $r_{n,k}$ counts linearly symmetric trees once, but counts both reflections of linearly asymmetric trees. We can conclude that the number of k-linear trees on n vertices can be defined as

$$a_{n,k} = \frac{1}{2}\bigg(r_{n,k} + s_{n,k}\bigg).$$

2.2 Generating Functions

Generating functions are a crucial tool that we will use to count linear trees. We will derive a bivariate generating function for the total number of linear trees on n vertices where k of

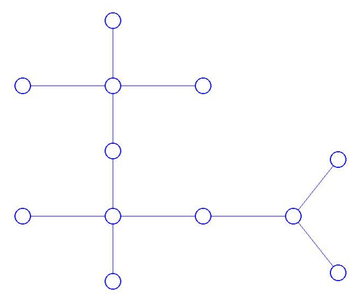

Figure 2.2: *The tree* $([1,1,1],1,[1,1],1,[1,1])$ *is linearly asymmetric. In the star labeled case, this tree would be distinct from* $([1,1],1,[1,1],1,[1,1,1])$.

those vertices are HDVs. From this generating function, we will extract closed formulas. We define the generating function of interest as $A(x,y) = \sum_{n=1}^{\infty}\sum_{k=0}^{\infty} a_{nk}x^n y^k$ where a_{nk} is the number of k-linear trees on n vertices. We begin by determining generating functions for each T_i and each s_j in a linear tree. The generating functions for each individual component can then be multiplied together and after some adjustments for linear symmetry, the end result will be a generating function for linear trees.

There is only one possible path on i number of vertices. So, we can define

$$s(x) = \frac{1}{1-x} = \sum_{i=0}^{\infty} x^i$$

as the generating function for the total number of paths with i vertices.

Consider any interior generalized star on i vertices. The degree of the central vertex must be at least 1 in order for the degree of that vertex to be at least 3 in a linear tree. So, there there are $m \geq 1$ branches of length $l_1, \ldots, l_m$ where $\sum_{j=1}^{m} l_j = i-1$ and each $l_j \geq 1$. Notice that since $m \geq 1$, $i-1 \geq 1$. Therefore, a minimum of 2 vertices are needed to form an interior star. Recall that the lengths of the pendant branches of a generalized star on i vertices form an integer partition of $i-1$. The number of integer partitions of $i-1$ is the coefficient of x^{i-1} in the generating function $\prod_{i=1}^{\infty} \frac{1}{1-x^i} = 1 + x + 2x^2 + 3x^3 + 5x^4 + \ldots$ [9]. The sum of the branch lengths of an interior star is at least 1, so $-1 + \prod_{i=1}^{\infty} \frac{1}{1-x^i}$ is the generating function for branch lengths of an interior star. Finally, to account for the central vertex, we multiply this function by x to obtain the generating function of an interior star as

$$T_{int}(x) = x\left(-1 + \prod_{i=1}^{\infty} \frac{1}{1-x^i} \right).$$

Exterior stars are similar to internal stars, except that the degree of the central vertex of exterior stars must be at least 2. Viewing branch lengths as parts of an integer partition, the total number of different possibilities of branch lengths of an exterior star on i vertices is the number of integer partitions of $i-1$ with at least 2 parts. For any integer $i-1$ there is only one partition containing fewer than two parts, $i-1$ itself. So, the coefficient of each x^{i-1} in the generating function for partition numbers must be decreased by 1. Therefore the generating function for partition numbers with 2 parts or more is $-\frac{1}{1-x} + \prod_{i=1}^{\infty} \frac{1}{1-x^i}$. Multiplying this function by x to account for the central vertex results in the generating function for exterior stars:

$$T_{ext}(x) = x\left(-\frac{1}{1-x} + \prod_{i=1}^{\infty} \frac{1}{1-x^i} \right).$$

Define $R(x,y)$ to be the generating function for $r_{n,k}$, the total number of reflections of linear trees. We will restrict ourselves to the case where $k \geq 2$. So,

$$R(x,y) = \sum_{n=1}^{\infty} \sum_{k=2}^{\infty} r_{n,k} x^n y^k.$$

Any reflection of a linear tree can be denoted $(T_1, s_1, T_2, s_2, \ldots, s_{k-1}, T_k)$. Both T_1 and T_k are exterior stars, and there are $k-2$ remaining T_i (for $i \in \{2, \ldots, k-1\}$) which are interior stars. In addition, there are $k-1$ different s_j (for $j \in \{1, 2, \ldots, k-1\}$) which refer to paths between each generalized star. Thus, we can define

$$R(x,y) = \sum_{j=2}^{\infty} \Big(T_{ext}(x)\Big)^2 \Big(T_{int}(x)\Big)^{j-2} \Big(s(x)\Big)^{j-1} y^j.$$

One can carefully derive a generating function for linearly symmetric trees using the same techniques. Consider some linearly symmetric tree $(T_1, s_1, \ldots, s_{k-1}, T_k)$ on n vertices. Each T_i (s_j) must be the same as T_{k-i+1} (s_{k-j}) for all $i \in \{1, 2, \ldots, k\}$ $(j \in \{1, 2, \ldots, k-1\})$. So, if we count the number of linearly symmetric trees while fixing T_i (s_j), we implicitly fix T_{k-1} (s_{k-1}) as well. This means that if one non-central component has r vertices, then its reflection also has r vertices, and the rest of the tree will contain $n-2r$ vertices. The central component is free to be different from every other component in the tree because it has no reflection. So, if the central component is an interior star (path), the generating function for this central component is $T_{int}(x)$ $(s(x))$. For non-central components, we denote their generating functions as $s^*(x), T^*_{int}(x)$ and $T^*_{ext}(x)$.

We begin by considering the generating function for a non-central path. As before, there is only one possible path on a given number of vertices. Any choice for such a path also determines its symmetric counterpart, so each vertex must be counted twice. Therefore, the only powers of x that can have a nonzero coefficient are even powers. Thus, the generating function for non-central paths in a linearly symmetric tree is

$$s^*(x) = s(x^2) = \frac{1}{1-x^2}.$$

We can apply similar arguments to the interior and exterior stars. This means that the generating functions for non-central interior and exterior stars respective are

$$T^*_{int}(x) = T_{int}(x^2) = x^2\bigg(-1 + \prod_{i=1}^{\infty} \frac{1}{1-x^{2i}}\bigg)$$

and

$$T^*_{ext}(x) = T_{ext}(x^2) = x^2\left(-\frac{1}{1-x^2} + \prod_{i=1}^{\infty} \frac{1}{1-x^{2i}} \right).$$

This information is enough for us to determine $S(x,y) = \sum_{n=0}^{\infty}\sum_{k=2}^{\infty} s_{n,k}x^n y^k$. We note that in this generating function, the central component will be a path when k is even, and a generalized star when k is odd.

$$S(x,y) = \sum_{j=2}^{\infty} T^*_{ext}(x)\Bigg(\Big(T^*_{int}(x)\Big)^{j-2}\Big(s^*(x)\Big)^{j-2} s(x)$$

$$+\Big(T^*_{int}(x)\Big)^{j-2}\Big(s^*(x)\Big)^{j-1} T_{int}(x)y\Bigg)y^{2j-2}$$

We conclude by defining the generating function for the total number of unlabeled linear trees as

$$A(x,y) = \frac{1}{2}\Big(R(x,y) + S(x,y)\Big).$$

There is one final observation worth noting. The first nonzero coefficient for $T_{ext}(x)$ is the coefficient for x^3, and the first nonzero coefficient for $T_{int}(x)$ is the coefficient for x^2. In the symmetric case, the first nonzero coefficient for $T^*_{ext}(x)$ is the coefficient for x^6, and the first nonzero coefficient for $T^*_{int}(x)$ is the coefficient for x^4. Both $s(x)$ and $s^*(x)$ have a constant term as their first nonzero coefficient. So, given some y^k, the smallest power of x with a nonzero coefficient is $x^{2k+2}y^k$. Thus, the smallest k-linear tree has $2k+2$ vertices.

2.3 Closed Formulas

It is possible to obtain closed formulas for k-linear trees on n vertices. If $k = 0$, then there are no high degree vertices and the tree is a path. If $k = 1$, then the tree is a generalized star. In such a tree, there are at least three branches whose lengths sum to $n-1$. This is the same as the number of integer partitions of $n-1$ with at least 3 parts. We define $p(n)$ to be the number of integer partitions of n. The total number of partitions of $n-1$ with at least

3 parts is the number of integer partitions of $n-1$, minus all integer partitions with two or one parts. For any $n-1$, there are $\lfloor \frac{n-1}{2} \rfloor$ integer partitions of $n-1$ with two parts, and one integer partition with one part. So, the total number of generalized stars on n vertices is $p(n-1) - \big(1 + \lfloor \frac{n-1}{2} \rfloor\big)$.

For larger values of k, we can obtain a closed formula from each coefficient in the generating function. Consider the term $T_{ext}(x)^2 s(x) y^2$ in $R(x,y)$. The nonzero coefficients of each x^n in $T_{ext}(x) = x\big(\frac{-1}{1-x} + \prod_{i=1}^{\infty} \frac{1}{1-x^i}\big) = t_n^{ext} x^n$ are the number of integer partitions of $n-1$ decreased by 1. These coefficients are nonzero whenever $n \geq 3$. So, $t_n^{ext} = p(n-1) - 1$. The coefficients of all powers of x in $s(x)$ are simply 1. When the power of y is 2, the smallest power of x with a nonzero coefficient in $R(x,y)$ is x^6. So, $T_{ext}(x)^2 s(x) y^2 = \sum_{n=6}^{\infty} r_{n,2} x^n y^2$. Thus, we can define

$$r_{n,2} = \sum_{i=3}^{n-3} \sum_{j=3}^{n-i} \Big(p(i-1) - 1\Big)\Big(p(j-1) - 1\Big).$$

Note that in $R(x,y)$, when $k \geq 3$, the term where the power of y is $k+1$ is

$$T_{ext}^2(x) T_{int}^{k-1}(x) s^k(x) y^{k+1} = \Big(T_{ext}^2(x) T_{int}^{k-2}(x) s^{k-1}(x) y^k\Big)\Big(T_{int}(x) s(x) y\Big),$$

where $T_{ext}^2(x) T_{int}^{k-2}(x) s^{k-1}(x) y^k$ is the generating function for k-linear trees on n vertices. So, we can define $r_{n,k}$ recursively, when $k \geq 3$ as follows

$$r_{n,k} = \sum_{i=2k}^{n-2} \sum_{j=1}^{n-i-1} r_{i,k-1} p(j).$$

Now that we have extracted the coefficients for $R(x,y)$, we must extract the coefficients for $S(x,y)$. The term in$S(x,y)$ where the power of y is 2 is $T_{ext}^*(x) s(x) y^2$. The coefficient of this term is the sum of the coefficients of $T_{ext}^*(x)$, and the coefficient of x^i in $T_{ext}^*(x)$ is $(p(i/2-1) - 1)$ if i is even, and 0 otherwise. So, we can write this sum as

$$s_{n,2} = \sum_{i=3}^{\lfloor \frac{n}{2} \rfloor} \Big(p(i-1) - 1\Big).$$

The remaining coefficients for $s_{n,k}$ can be extracted in a similar way. A more detailed discussion of this process can be found in the appendix. The following recursive formula allows us to count $2l$ interior stars on $2m$ vertices. These are the coefficients of $T_{int}^*(x)^l s^*(x)$ in the generating function $S(x,y)$.

$$s_{rec}(m,l) = \begin{cases} \sum_{i=0}^{m} \sum_{j=0}^{m-i} p(1+i) s_{rec}(m-i-j, l-1) & \text{if } l \geq 1 \\ 1 & \text{if } l = 0 \text{ and } m = 0 \\ 0 & \text{otherwise} \end{cases}$$

This recursive formula allows us to carefully extract the coefficients of $(T_{int}^*(x))^{l-2}(s^*(x))^{l-2}$ as it appears in $S(x,y)$. However, there are other factors in the generating function whose coefficients we must still compute. If $k \geq 3$ is odd, then we must count the two non-central exterior stars, two non-central paths, and the central interior star. These missing factors in the generating function are $T_{ext}^*(x)s(x)^*T_{int}(x)$. If $k \geq 4$ is even, then the missing factors are $T_{ext}^*(x)s(x)$. We provide the closed formulas for coefficients of $S(x,y)$ for powers of y that are odd and larger than 3. We omit the exact details of the derivation, which is similar to the derivation of the closed formulas we have already derived in this section.

$$s_{odd}(n,k) =$$
$$\sum_{i=0}^{\lfloor \frac{n-2-2k}{2} \rfloor} \sum_{j=0}^{\lfloor \frac{n-2-2k-2i}{2} \rfloor} \sum_{t=0}^{\lfloor \frac{n-2-2k-2i-2j}{2} \rfloor} \Big(p(2+i) - 1\Big) p(n-1-2k-2i-2j-2t) s_{rec}(j, \frac{k-3}{2})$$

The following is the closed formula for coefficients in $S(x,y)$ where the power of y is even and larger than 4:

$$s_{even}(n, k \geq 4) = \sum_{i=0}^{\lfloor \frac{n-2-2k}{2} \rfloor} \sum_{j=0}^{\lfloor \frac{n-2-2k-2i}{2} \rfloor} \Big(p(2+i) - 1\Big) s_{rec}(j, \frac{k-2}{2})$$

This collection of closed formulas gives us a complete description of the coefficients of $S(x, y)$ as follows:

$$s_{n,k} = \begin{cases} s_{n,2} & \text{if } k = 2 \\ s_{odd}(n, k \geq 3) & \text{if } k \geq 3 \text{ is odd} \\ s_{even}(n, k \geq 4) & \text{if } k \geq 4 \text{ is even} \end{cases}$$

To conclude, we recall that $A(x, y) = \sum_{n=0}^{\infty} \sum_{k=2}^{\infty} a_{n,k} x^n y^k = \frac{1}{2}\big(R(x, y) + S(x, y)\big)$. So, the coefficients of $A(x, y)$, are $a_{n,k} = \frac{1}{2}\big(r_{n,k} + s_{n,k}\big)$.

These formulas were invaluable for determining that all linear trees had been generated when expanding the multiplicity database. Althought it is somewhat complicated to derive these formulas, it is relatively easy to compute them. Using these closed formulas in Matlab, have computed the total number of k-linear trees on n vertices for all $k \leq 11$ and $n \leq 25$. The resulting computations can be found in the appendix.

2.4 Growth Rates of k-Linear Trees

The tables in the appendix raise an interesting question. Given some n and k, we may wish to determine when $a_{n,k} > a_{n-i,k-r}$ for some $i < n$ and some $r < k$. Using the closed formulas for $a_{n,k}$, we can prove the following theorem:

Theorem 2.4.1. *For any $k > 2$ and any $n \geq 2k + 4$, $a_{n,k} > a_{n-3,k-1}$.*

Proof. Let $k > 2$ and $n \geq 2k + 4$. Then,

$$r_{n,k} = \sum_{i=2k}^{n-2} \sum_{j=1}^{n-i-1} r_{i,k-1} p(j)$$

$$= r_{n-2,k-1} + 3r_{n-3,k-1} + \sum_{i=2k}^{n-4} \sum_{j=1}^{n-i-1} r_{i,k-1} p(j) > 2r_{n-3,k-1}.$$

Recall that $r_{n,k}$ counts symmetric trees once, and asymmetric trees twice. So, $r_{n,k} > s_{n,k}$. We can show

$$a_{n-3,k-1} = \frac{1}{2}(r_{n-3,k-1} + s_{n-3,k-1}) < \frac{1}{2}(r_{n-3,k-1} + r_{n-3,k-1}) = r_{n-3,k-1},$$

and

$$a_{n,k} = \frac{1}{2}(r_{n,k} + s_{n,k}) > \frac{1}{2} r_{n,k} > \frac{1}{2}(2r_{n-3,k-1}) = r_{n-3,k-1}.$$

So we conclude that $a_{n,k} > r_{n-3,k-1} > a_{n-3,k-1}$.

$\square$

The following conjecture is a natural extension of this theorem:

Conjecture 2.4.2. *Given some k, there exists some N such that for all $n > N$, the number of k-linear trees surpasses the number of $k-1$-linear trees. That is, $a_{n,k} > a_{n,k-1}$.*

This conjecture is of particular interest because of its implications for the ratio of nonlinear trees to total trees. Some examination of the tables in the appendix indicates that this seems to be the case for all $k \leq 5$. Examination of the values of $a_{2k+3,k}$ for each k reveals that linear trees with more high degree vertices seem to have a higher initial growth rate with n. However, this conjecture has not been rigorously proven.

Chapter 3

Properties of Nonlinear Trees

3.1 Homeomorphic Skeletons

Consider any edge, $\{u, v\}$, in a graph. When we apply *edge subdivision* to that edge, we introduce a new vertex x whose only incident edges are $\{u, x\}$ and $\{x, v\}$, and delete the edge $\{u, v\}$. In addition, we can apply *reverse edge subdivision* (which is sometimes referred to as *smoothing out*) to two edges on a graph. Suppose there is a graph with three vertices, u, v and x, where $\deg x = 2$. Suppose that $\{u, x\}$, and $\{x, v\}$ are edges in the graph, but $\{u, v\}$ is not an edge in the graph. Then, applying reverse edge subdivision removes x and its incident edges, and constructs the new edge $\{u, v\}$. Edge subdivision and reverse edge

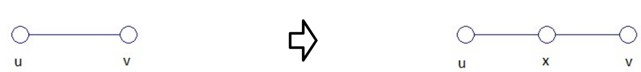

Figure 3.1: Edge Subdivision

subdivision are illustrated in Figure 3.1 and Figure 3.2 respectively.

Two graphs G and K are *homeomorphic* when some G' is isomorphic to K', where G' and K' are both graphs created by applying edge subdivision or reverse edge subdivision to G and K any number of times. The *homeomorphic skeleton* of a tree, T, is the tree with the fewest vertices that is homeomorphic to T. We denote the homeomorphic skeleton of a tree $Sk(T)$.

Lemma 3.1.1. *Let T be any tree. Reverse edge subdivision preserves the degree of all other vertices in the resulting tree, except for the vertex that is removed in the process.*

Proof. Let T be a tree. Now, consider some vertex v, and any reverse edge subdivision that does not remove v. Consider any edge $\{v, x\}$ in T. If x does not have degree 2, then there is no reverse edge subdivision that will affect this edge, since x has degree greater than 2. If $\deg(x) = 2$, let $\{v, x\}$ and $\{x, u\}$ be the edges incident to x. If the reverse edge subdivision were applied to these edges, both of these edges and x will be removed. This would decrease the degree of v by 1. But, the edge $\{v, u\}$ would be created. Note that by the definition of reverse edge subdivision, this edge could not have been an edge in T. So, $\{v, u\}$ will be a new edge, and the degree of v will be unchanged.

□

Since a single reverse edge subdivision does not affect the degree of any remaining vertices,

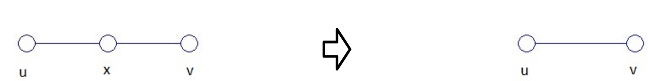

Figure 3.2: Reverse Edge Subdivision

multiple edge subdivisions will not affect the degree of any of the remaining vertices. Since reverse edge subdivision decreases the total number of vertices by 1 without changing the degrees of the resulting vertices, the homeomorphic skeleton can be found using the following algorithm. Let T be any tree. If there is no vertex of degree 2, then we cannot use reverse edge subdivision to reduce the number of vertices in T, so it is its own homeomorphic skeleton. If not, then find a vertex of degree 2. Apply reverse edge subdivision to this vertex and its neighbors. Repeat the process with the resulting tree. This leads us to the following observation:

Observation 3.1.2. *Let T be a tree, and $Sk(T)$ be its homeomorphic skeleton. Then, there are no vertices of degree 2 in $Sk(T)$.*

If this observation did not hold, then we could apply reverse edge subdivision on the degree 2 vertex of a tree which would result in a homeomorphic tree on fewer vertices. Let T be any tree, and let $Sk(T)$ be its homeomorphic skeleton. From the previous lemma, we know that all vertices have the same degree as their counterparts in T, and there are no vertices of degree 2. So, every vertex in $Sk(T)$ whose degree is larger than 1 is a high degree vertex. Let $v_1, v_2, \ldots v_r$ be all vertices of degree 1 in $Sk(T)$. We define the *linear structure* of T as the subgraph of $Sk(T)$ includes all vertices and edges in $Sk(T)$ except for $v_1, v_2, \ldots v_r$ and their incident edges. When a linear structure is a tree on k vertices, we sometimes refer to it as a *k-structure*. The counterparts to the vertices in the linear structure of T are the high degree vertices in T..

Observation 3.1.3. *The k-structure of a linear tree T is a path on k vertices.*

This is fairly easy to show. For any linear tree, the branches of each generalized star and the paths between high degree vertices can be smoothed out. So, the homeomorphic skeleton of a linear tree is a tree of depth 1 with edges between the high degree vertices. The high degree vertices all lie on the same induced path, and there are k such vertices. So, the removal of the degree 1 vertices from the homeomorphic skeleton of a linear tree results in a path of length k.

We can also make a useful observation from linear structures. On any tree with a given k-structure, there are $k-1$ paths between high degree vertices which do not contain other high degree vertices (some paths lengths may be zero). Furthermore, the number of generalized stars on n vertices whose central vertex is a vertex of degree 1 in the k-structure can be counted the same way as the number of exterior stars in the linear case, and the number of generalized stars on n vertices whose center is a vertex of degree 2 in the k-structure can be counted the same way as the number of interior stars in the linear case.

We define a *central star* as a generalized star whose central vertex has high degree in the linear structure, and its pendant branches (not including other high degree vertices or paths between them). When considering a generating function for a central star, we note that since the central vertex has degree of at least 3 without the addition of any pendant branches, no additional pendant branches are necessary. So, we can define its generating fuction as

$$T_{cent}(x) = x \prod_{i=1}^{\infty} \frac{1}{1-x^i}.$$

3.2 Homeomorphic Subtrees

Nonlinear trees are not as conveniently structured as linear trees. However, using homemorphic skeletons and linear structures, we can determine some commonalities among all nonlinear trees.

Lemma 3.2.1. *If a tree is nonlinear, then it must contain at least* 4 *HDVs.*

Proof. Consider the linear structure of any tree. If the linear structure is a path, then that tree must be a linear tree. So if T is a nonlinear tree, its linear structure cannot be a path. The first linear structure that is not a path is a star on 4 vertices. Thus, a nonlinear tree must contain at least 4 high degree vertices.

□

Lemma 3.2.2. *Let T be a nonlinear tree, and consider any three high degree vertices, v_1, v_2 and v_3. Then, either these vertices lie on the same path, or there is a fourth high degree vertex that lies on the path between any v_i and v_j for any $i, j \in \{1, 2, 3\}$, $i \neq j$.*

Proof. Take any three high degree vertices in T, and label them v_1, v_2 and v_3. Suppose they do not lie on the same path. There is a path between v_1 and v_2, a path from v_2 to v_3 and, a walk from v_1 to v_2 to v_3. This walk cannot be a path. So some, but not all vertices on the path from v_1 to v_2 must be used on the path from v_2 to v_3. Let $v_1 - v_{i_1} - v_{i_2} - \ldots - v_{i_{k-1}} - v_{i_k} - v_2$ be the path from v_i to v_2, and let $v_2 - v_{j_1} - v_{j_2} - \ldots - v_{j_{l-1}} - v_{j_l} - v_3$ be the path from v_2 to v_3. Now, let v^* be the vertex with the smallest $a \in \{1, 2, \ldots, k\}$ such that $v_{i_a} = v_{j_b}$ for some $b \in \{1, 2, \ldots, l-1, l\}$. So, the following must be true. If $a = 1$ then v^* is adjacent to v_1. Otherwise it is adjacent to $v_{i_{a-1}}$. If $a = k$, then v^* is adjacent to v_2. Otherwise it is adjacent to $v_{i_{a+1}}$. If $b = l$, then v^* is adjacent to v_3. Otherwise it is adjacent to $v_{j_{b+1}}$. So, the degree of v^* is at least 3, and it is a high degree vertex. Furthermore, v^* lies on the paths between any pair of v_1, v_2 or v_3.

□

Lemma 3.2.3. *Let T be a tree. If there is a nonlinear subtree of T, then T is nonlinear.*

Proof. Let T be a tree and let there be some nonlinear subtree of T. Then, there are at least 3 high degree vertices in this subtree, v_1, v_2 and v_3, that do not lie on the same path. Suppose that in T there is a path that does contain all 3 HDVs. Without loss of generality, say this path is $v_1 - n_1 - \ldots - n_i - v_2 - n_{i+1} - \ldots - n_j - v_3$. Then, because this path cannot be realized in the subtree, at least one n_l must not be in the subtree for some $l \in \{1, \ldots, j\}$. However, there is exactly one path between any two vertices in a tree. So, if n_l is not in the subtree, then there is no path from v_1 to v_3. Thus, the subtree is disconnected, which is a contradiction. Therefore, there must be no path containing these three vertices in the T, and the T is nonlinear.

□

Theorem 3.2.4. *A tree is nonlinear if and only if it contains a subtree that is homeomorphic to the nonlinear* 10 *vertex tree.*

Proof. Suppose a tree contains a subtree that is homeomorphic to the nonlinear 10 vertex tree. Let v_1 be the vertex homeomorphic to the central vertex of the nonlinear 10 vertex tree, and let v_2, v_3 and v_4 be the vertices homeomorphic to the other HDVs in the 10 vertex tree. Then, there is a path from v_2 to v_1, from v_3 to v_1, and from v_4 to v_1. Furthermore, the only vertex in common with these paths is v_1. So, there is a path from v_2 to v_1 to v_3, from v_2 to v_1 to v_4, and from v_3 to v_1 to v_4. So, there is no path on which all of v_1, v_2, v_3 and v_4 lie. Thus, the tree is nonlinear.

Now, suppose a tree is nonlinear. Then, we can identify at least 3 HDVs which do not fall on a single path. Call these vertices v_1, v_2, and v_3. Then, we know that there exists an additional high degree vertex, v_4, which lies on the path between v_1 and v_2, on the path between v_1 and v_3, and on the path between v_2 and v_3. Let T' be a subtree that contains the path from v_1 to v_2 (which includes v_4), the path from v_4 to v_3, and two branches of length 1 on each of v_1, v_2 and v_3. Note that for each of v_1, v_2 and v_3, we can include two branches of length 1 and the path, since they have degree at least 3. Furthermore, v_4 has degree 3 in this tree, since the paths from v_4 to each of v_1, v_2 and v_3 are distinct. This subtree, T', is homeomorphic to the 10 vertex nonlinear tree. To show this, we simply smooth out the paths from v_1 to v_4, from v_2 to v_4, and from v_3 to v_4 so that a single edge incident to each of v_1, v_2 and v_3 that is also incident to v_4. Then, we are left with a graph where there are three degree 3 vertices which are adjacent to v_4. This is the 10 vertex nonlinear tree.

□

3.3 Automorphism Groups

Let G be a graph. We say that ϕ is an *automorphism* if ϕ is an isomorphism that maps the graph onto itself. The set of automorphisms of any graph form a group with the trivial automorphism, defined as $\phi(v) = v$ for all $v \in V(G)$, as the identity. We make the following general observation about isomorphisms.

Observation 3.3.1. *If T and T' are isomorphic, then the vertices of a generalized star in T must each map to the vertices of an isomorphic generalized star T'.*

This observation follows from the fact that an isomorphism defined from T to T' will also define an isomorphism between any subtree of T and the image of the vertices of that subtree. It also follows that the image of vertices that form a path in T must form a path in T'. The existence of an isomorphism between T and T' also has important consequences for that tree's linear structure.

Theorem 3.3.2. *Let T and T' be two trees. Let G and G' be their respective linear structures. Then,*

1. *if T and T' are isomorphic then G is isomorphic to G'.*
2. *suppose T is isomorphic to T'. Let $\phi : V(T) \to V(T')$ be an isomorphism between T and T'. Then the same mapping of $\phi : V(G) \to V(G')$ is an isomorphism.*

Proof. Let T and T' be two trees, and let G and G' be their respective linear structures. Suppose T and T' are isomorphic. Then, they share the same homeomorphic skeleton. Since they share the same homeomorphic skeleton, they will also share the same linear structure. So, G and G' are isomorphic.

Now, let $\phi : V(T) \to V(T')$ be an isomorphism between T and T'. Consider the same mapping ϕ, but from $V(T) \supseteq V(G) \to V(G')$. Let $v, u \in V(G)$. We must show that $\{v, u\}$ is an edge in G if and only if $\{\phi(v), \phi(u)\}$ is an edge in G'. We know that $\{v, u\}$ is an edge in G if and only if it is an edge in T, or there is a path $v - v_1 - v_2 - \ldots - v_m - u$ where each v_i has degree 2 for all $i \in \{1, 2, \ldots, m\}$. If $\{u, v\}$ is an edge in T, then $\{\phi(u), \phi(v)\}$ is an edge in T'. Suppose $\{u, v\}$ is not an edge in T. Since u and v are in G, they must be high degree vertices in T. We also know that $v - v_1 - v2 - \ldots - v_m - u$ is a path in T if and only if $\phi(v) - \phi(v_1) - \phi(v_2) - \ldots - \phi(v_m) - \phi(u)$ is a path in T', since ϕ is an isomorphism. Furthermore, each $\phi(v_i)$ has degree 2 for all $i \in \{1, 2, \ldots, m\}$ since their pre-images have degree 2 in T. So, after smoothing out this path, $\{\phi(v), \phi(u)\}$ is an edge in $Sk(T')$. Since the degree of v and u is at least 3 in T, the degree of $\phi(v)$ and $\phi(u)$ is at least 3 in T' and in $Sk(T')$. So, they remain in G'. We therefore conclude that $\{\phi(v), \phi(u)\}$ is an edge in G' if and only if $\{u, v\}$ is an edge in G, and ϕ is also an isomorphism on G and G'.

□

The automorphism group of a tree's linear structure critically determines if one can count trees by counting the possibilities for generalized stars, similar to counting $r_{n,k}$. If the automorphism group of a tree's linear structure is nontrivial, then this approach will over-count trees. For example, consider linear trees. The linear structure of a linear tree is a path, and the automorphism group of any path $v_1 - v_2 - \ldots - v_n$ contains two elements: the trivial automorphism, and an automorphism that maps v_i to v_{n-i+1} for all $i \in \{1, 2, \ldots, n\}$. There is a connection between this nontrivial automorphism, and the double-counting of linearly asymmetric trees that we noted in $r_{n,k}$

Theorem 3.3.3. *Consider the set of trees whose linear structure is G. If the automorphism group of G is trivial, then we can count all such trees on n vertices by enumerating all possibilities for each individual generalized star and each individual path between generalized stars.*

Proof. Suppose we wish to count all trees on n vertices whose linear structure is G, and suppose G has a trivial automorphism group. Let v be a vertex in G, and consider the generalized star whose center is v. Suppose we wish to count all possible trees where this generalized star is fixed as T_1. We may be concerned that we may count this set of trees again when the generalized star whose center of v is not T_1. Let T be a particular tree that is counted when the generalized star centered on v is T_1, and T' be a tree that double counts this tree when the generalized star centered on v is not T_1. If we have double counted, then there exists an isomorphism from T to T'. Since G has a trivial automorphism group, any isomorphism from T to T' must map v to itself. But, we know that if that is the case, then the vertices of T_1 must map to the vertices of the star centered on v in T', a contradiction. So, we do not over-count trees using this method.

□

We will illustrate with an example. Let $B(x)$ be the generating function for a nonlinear tree whose linear structure is the generalized star with three branches with distinct lengths. The automorphism group of such a generalized star is trivial. We also note that the smallest such generalized star is $[3, 2, 1]$, which is a generalized star on 7 vertices. So, if this linear structure contains $k \geq 7$ vertices, we can define

$$B_k(x) = \sum_{n=0}^{\infty} b_{n,k}x^n = T_{ext}(x)^3 T_{cent}(x) T_{int}(x)^{k-4} s(x)^{k-1}.$$

3.4 Ratio of Nonlinear Trees to Total Trees

Let $d_{n,k}$ be the number of nonlinear trees on n vertices, k of which are high degree. Recall that $a_{n,k}$ is the total number of k-linear trees. We wish to show that $\frac{d_{n,k}}{d_{n,k}+a_{n,k}}$ when $n \geq 2k+2$ goes to 1 as k increases.

Lemma 3.4.1. *Let $a_{n,k}$ be the total number of k-linear trees on n vertices, and $r_{n,k}$ be the total number of reflections of k-linear trees on n vertices. Then, for all n and k, $a_{n,k} \leq r_{n,k}$.*

Proof. Recall that the number of star labeled trees will count linearly symmetric trees once, and linearly asymmetric trees twice. The number of linear trees will count linearly symmetric and linearly asymmetric trees exactly once. So, there are at least as many reflections as there are linear trees.

□

We now have an overestimate of $a_{n,k}$. Next, we will obtain an underestimate of $d_{n,k}$. Let $p_{3,d}(k)$ be the number of integer partitions of k into three distinct parts. Recall that this is number of generalized stars with three branches with distinct lengths on $k+1$ vertices.

Lemma 3.4.2. *Let $p_{3,d}(k)$ be the number of integer partitions of k into three distinct parts. Then, as k increases, $p_{3,d}(k)$ is non-decreasing and unbounded.*

Proof. Define $P_k = \{(a,b,c) | a+b+c = k, a > b > c \geq 1\}$ be the set of integer partitions of k into three distinct parts. Consider the mapping $f : P_k \to X \subseteq P_{k+1}$ defined by $f((a,b,c)) = (a+1,b,c)$. We must show that this is an injective mapping. Suppose $f((a,b,c)) = f((a',b',c'))$. Then, $(a+1,b,c) = (a'+1,b',c')$. So, $a = a'$, $b = b'$, $c = c'$, and $(a,b,c) = (a',b',c')$. We therefore conclude that this function is injective, and $p_{3,d}(k) = |P_k| \leq |P_{k+1}| = p_{3,d}(k+1)$. So $p_{3,d}(k)$ is non-decreasing as k increases. Now, suppose k is even. Then, $(\frac{k}{2}, \frac{k}{2}-1, 1) \in P_k$ cannot have a pre-image in P_{k-1}, since $f^{-1}((\frac{k}{2}, \frac{k}{2}-1, 1)) = (\frac{k}{2}-1, \frac{k}{2}-1, 1)$ is not a partition with distinct parts. So if k is even then $p_{3,d}(k) > p_{3,d}(k-1)$. Since $p_{3,d}$ is non-decreasing and is not constant for all $k > N$ for some positive integer N, $p_{3,d}$ is unbounded as k increases.

□

If we are given a linear structure that is a generalized star on k vertices with three branches with distinct length, then the generating function $B_k(x)$ can be used to count nonlinear trees with that linear structure. Since there are $p_{3,d}(k-1)$ such linear structures for any k, the generating function for all nonlinear trees whose linear structure is a generalized star with three branches with disinct lengths is $p_{3,d}(k-1)B_k(x)$.

Lemma 3.4.3. $p_{3,d}(k-1)b_{n,k} \leq d_{n,k}$

Proof. Consider nonlinear trees on n vertices, k of which are high degree. Then, there are $p_{3,d}(k-1)$ linear structures that are a generalized star with three branches with distinct lengths, and $b_{n,k}$ trees on such a linear structure. So, there are $p_{3,d}(k-1)b_{n,k}$ total nonlinear trees whose linear structure is a generalized star with three branches with distinct lengths. We conclude that $p_{3,d}(k-1)b_{n,k} \leq d_{n,k}$.

□

In order to estimate the ratio of nonlinear trees to total trees, we must be able to directly compare the number of linear trees to the number of nonlinear trees. Using generating functions is a simple way to make such a comparison..

Lemma 3.4.4. *For $k \geq 9$, $r_{n.k} \leq 2b_{n,k}$.*

Proof. Let $R_k(x)$ be the portion of the generating function $R(x,y)$ where the power of y is k. We will show this by proving the coefficients of $2B_k(x) - R_k(x)$ are nonnegative for terms where the power of y is at least 9. We can simplify the generating functions as follows:

$$2T_{int}(x)^{k-4}T_{ext}(x)^3T_{cent}(x)s(x)^{k-1} - T_{int}(x)^{k-2}T_{ext}(x)^2s(x)^{k-1}$$

$$= T_{int}(x)^{k-4}T_{ext}(x)^2s(x)^{k-1}\left[2T_{ext}(x)T_{cent}(x) - T_{int}(x)^2\right].$$

Note that each of $T_{int}(x)$, $T_{ext}(x)$, and $s(x)$ has nonnegative coefficients. So, their product must have nonnegative coefficients. We must show that $2T_{ext}(x)T_{cent}(x) - T_{int}(x)^2$ has nonnegative coefficients. Observe the following:

$$2T_{ext}(x)T_{cent}(x) - T_{int}(x)^2$$

$$= 2x\biggl(-s(x) + \prod_{i=1}^{\infty}\frac{1}{1-x^i}\biggr)x\prod_{i=1}^{\infty}\frac{1}{1-x^i} - \biggl(x\biggl(-1 + \prod_{i=1}^{\infty}\frac{1}{1-x^i}\biggr)\biggr)^2$$

$$= x^2\Bigg(2\Bigg(-s(x)+\prod_{i=1}^{\infty}\frac{1}{1-x^i}\Bigg)\prod_{i=1}^{\infty}\frac{1}{1-x^i}-\Bigg(-1+\prod_{i=1}^{\infty}\frac{1}{1-x^i}\Bigg)^2\Bigg).$$

The x^2 term also has nonnegative coefficients, so this leaves only needing to check the coefficients of

$$2\Bigg(-s(x)+\prod_{i=1}^{\infty}\frac{1}{1-x^i}\Bigg)\prod_{i=1}^{\infty}\frac{1}{1-x^i}-\Bigg(-1+\prod_{i=1}^{\infty}\frac{1}{1-x^i}\Bigg)^2$$

We note the first nonzero coefficient in this resulting formula is the coefficient for for x^2. If $n = 2$ then this is $2((p(2)-1)p(2-2)-p(1)p(1) = 2-1 = 1$, and if $n = 3$ then this is $2((p(3)-1)p(3-2)+p(3-1)p(3-3))-(p(1)p(2)+p(2)p(1)) = 2(2+2)-(2+2) = 4$.

Next, we can extract the coefficient x^n for any $n \geq 4$ as

$$2\sum_{i=2}^{n}(p(i)-1)p(n-i)-\sum_{j=1}^{n-1}p(j)p(n-j)$$

$$= 2\sum_{i=2}^{n}\big(p(i)p(n-i)-p(n-i)\big)-\sum_{j=1}^{n-1}p(j)p(n-j).$$

But,

$$2\sum_{i=2}^{n}p(i)p(n-i)-p(n-i)$$

$$= p(n)p(0)-p(0)+2\Bigg(\sum_{i=2}^{n-1}p(i)p(n-i)-p(n-i)\Bigg)$$

$$= p(n)-1+2\Bigg(\sum_{i=2}^{n-1}p(i)p(n-i)-p(n-i)\Bigg)$$

and

$$\sum_{j=1}^{n-1} p(j)p(n-j)$$

$$= p(1)p(n-1) + \sum_{j=2}^{n-1} p(j)p(n-j)$$

$$= p(n-1) + \sum_{j=2}^{n-1} p(j)p(n-j).$$

So,

$$2\sum_{i=2}^{n} \big(p(i)p(n-1) - p(n-1)\big) - \sum_{j=1}^{n-1} p(j)p(n-j)$$

$$= p(n) - 1 + 2\Bigg(\sum_{i=2}^{n-1} \big(p(i)p(n-i) - p(n-i)\big)\Bigg) - (p(n-1) + \sum_{j=2}^{n-1} p(j)p(n-j))$$

$$= p(n) - 1 - p(n-1) + \Bigg(\sum_{i=2}^{n-1} \big(2p(i)p(n-i) - 2p(n-i) - p(i)p(n-i)\big)\Bigg)$$

$$= p(n) - 1 - p(n-1) + \Bigg(\sum_{i=2}^{n-1} p(i)p(n-i) - 2p(n-i)\Bigg)$$

$$= p(n) - 1 - p(n-1) + \sum_{i=2}^{n-1} (p(i) - 2)p(n-i).$$

We note that $p(i) - 2$ is positive when $i \geq 3$, and 0 when $i = 2$. So, we can rewrite this formula as

$$p(n) - 1 - p(n-1) + \sum_{i=3}^{n-1} (p(i) - 2)p(n-i).$$

Because $n \geq 2$, the $p(n) - 1 - p(n-1) \geq p(2) - 1 - p(2-1) = 2 - 1 - 1 = 0$. So, the total sum is nonnegative. Since each coefficient of $2B_k(x) - R_k(x)$ is nonnegative, the difference $2b_{n,k} - r_{n,k}$ is nonnegative and $r_{n,k} \leq 2b_{n,k}$.

□

It is interesting to note that the following proof relies on the fact that nonlinear trees have many more linear structures than linear trees. Thus, the rich structure for nonlinear trees is the key reason that, as the number of high degree vertices increases, there are more nonlinear trees than linear trees.

Lemma 3.4.5. *When $n \geq 2k+2$ and $k \geq 9$, then* $\frac{p_{3,d}(k-1)b_{n,k}}{p_{3,d}(k-1)b_{n,k}+r_{n,k}} \geq \frac{p_{3,d}(k-1)}{2+p_{3,d}(k-1)}$.

Proof. When $k \geq 9$, then $p_{3,d}(k-1) \geq 2$, since $(5,2,1)$ and $(4,3,1)$ are two partitions of 8 into three distinct parts, and $p_{3,d}(k-1)$ is non-decreasing. Furthermore, when $n \geq 2k+2$, $r_{n,k} = 1$. So the denominator of $\frac{p_{3,d}(k-1)b_{n,k}}{p_{3,d}(k-1)b_{n,k}+r_{n,k}}$ will be nonzero. Since $2b_{n,k} \geq r_{n,k}$, $\frac{p_{3,d}(k-1)b_{n,k}}{p_{3,d}(k-1)b_{n,k}+r_{n,k}} \geq \frac{p_{3,d}(k-1)b_{n,k}}{p_{3,d}(k-1)b_{n,k}+2b_{n,k}} = \frac{p_{3,d}(k-1)}{2+p_{3,d}(k-1)}$, which gives us our desired result.

□

Theorem 3.4.6. *When $n \geq 2k+2$,* $\lim_{k\to\infty} \frac{d_{n,k}}{d_{n,k}+a_{n,k}} \to 1$

Proof. When $n \geq 2k+2$, then $a_{n,k} \geq 1$. So, $d_{n,k} + a_{n,k} \geq 1$, so the denominator of this ratio is not zero. Now, $d_{n,k} \geq p_{3,d}(k-1)b_{n,k}$, and $a_{n,k} \leq r_{n,k} \leq 2b_{n,k}$. So, $\frac{d_{n,k}}{d_{n,k}+a_{n,k}} \geq \frac{p_{3,d}(k-1)b_{n,k}}{p_{3,d}(k-1)b_{n,k}+2b_{n,k}} = \frac{p_{3,d}(k-1)}{p_{3,d}(k-1)+2}$. Since $p_{3,d}(k-1)$ is unbounded as k increases, $\lim_{k\to\infty} \frac{p_{3,d}(k-1)}{p_{3,d}(k-1)+2} = 1$. So, $\lim_{k\to\infty} \frac{d_{n,k}}{d_{n,k}+a_{n,k}} = 1$.

□

This theorem is dependent on the number of linear and nonlinear trees given a fixed value for k. A natural extension of this theorem is to consider how the ratio of nonlinear trees to total trees changes with no restrictions on k. This is statement is formalized by the following conjecture:

Conjecture 3.4.7. *The total fraction of nonlinear trees goes to* 1 *as n increases.*

It can be shown that this ratio is at least $\frac{1}{2}$ if for any k, there exists some N such that for all $n > N$, $a_{n,k} > a_{n,k-1}$ (Conjecture 2.4.2). Let $m > 8$ be some positve integer such that $p_{3,d}(m-1) > 18$, and let N_i be the integer for which $a_{n,i} > a_{n,i-1}$ for all $n > N_i$ for all $i \in \{1, 2, \ldots, m\}$. Let $N = \max_{i \in \{1,2,\ldots,m\}}\{N_i\}$. Then, for all $n > N$, the following inequality holds: $\sum_{i=0}^{8} a_{n,i} < 1 + 8a_{n,m}$. So, $\sum_{i=0}^{8} a_{n,i} \leq 8a_{n,m} < 16b_{n,m}$.

We notice that

$$\frac{\sum_{k=0}^{\infty} a_{n,k} + \sum_{k=4}^{\infty} d_{n,k}}{\sum_{k=0}^{\infty} a_{n,k}} > \frac{\sum_{k=0}^{\infty} a_{n,k} + \sum_{k=4}^{\infty} d_{n,k}}{16b_{n,m} + \sum_{k=9}^{\infty} a_{n,k}}$$

$$> \frac{16b_{n,m} + \sum_{k=9}^{\infty} 2b_{n,k} + \sum_{k=9}^{\infty} p_{3,d}(k-1)b_{n,k}}{16b_{n,m} + \sum_{k=9}^{\infty} 2b_{n,k}}$$

$$= 1 + \frac{\sum_{k=9}^{\infty} p_{3,d}(k-1)b_{n,k}}{16b_{n,m} + \sum_{k=9}^{\infty} 2b_{n,k}}$$

Observe that

$$\sum_{k=9}^{\infty} p_{3,d}(k-1)b_{n,k} = \left(\sum_{k=9}^{m-1}(p_{3,d}-2)(k-1)b_{n,k} + (p_{3,d}(m-1)-18)b_{n,m} + \sum_{k=m+1}^{\infty}(p_{3,d}(k-1)-2)b_{n,k}\right)$$

$$+\left(\sum_{k=9}^{m-1} 2b_{n,k} + 18b_{n,m} + \sum_{k=m+1}^{\infty} 2b_{n,k}\right),$$

where

$$\sum_{k=9}^{m-1}(p_{3,d}-2)(k-1)b_{n,k} + (p_{3,d}(m-1)-18)b_{n,m} + \sum_{k=m+1}^{\infty}(p_{3,d}(k-1)-2)b_{n,k}$$

is positive. So,

$$\frac{\sum_{k=9}^{\infty} p_{3,d}(k-1)b_{n,k}}{16b_{n,m} + \sum_{k=9}^{\infty} 2b_{n,k}} > 1,$$

and

$$\frac{\sum_{k=0}^{\infty} a_{n,k} + \sum_{k=4}^{\infty} d_{n,k}}{\sum_{k=0}^{\infty} a_{n,k}} > 2.$$

If we take the reciprocal of both sides, this inequality becomes

$$\frac{\sum_{k=0}^{\infty} a_{n,k}}{\sum_{k=0}^{\infty} a_{n,k} + \sum_{k=4}^{\infty} d_{n,k}} < \frac{1}{2}.$$

Thus we can conclude,

$$\lim_{n\to\infty} \frac{\sum_{k=4}^{\infty} d_{n,k}}{\sum_{k=4}^{\infty} d_{n,k} + \sum_{k=0}^{\infty} a_{n,k}} > \frac{1}{2}.$$

The tables in the appendix provide strong evidence that this ratio is in fact larger than $\frac{1}{2}$. About 63% of trees on 25 vertices are nonlinear, and this ratio grows rapidly as the the number of vertices grows. However, the exact ratio of convergence remains unknown.

Chapter 4

Appendix

4.1 Tables

We use this chapter to give a detailed breakdown of computations of the number of k-linear trees, and the fraction of nonlinear to linear trees. The total number of unlabeled trees and number of integer partitions were obtained from the Online Encyclopedia of Integer Sequences [7] [8].

Num Vertices	Paths	1 HDV	2 HDV	3 HDV	Total
1	1	0	0	0	1
2	1	0	0	0	1
3	1	0	0	0	1
4	1	1	0	0	2
5	1	2	0	0	3
6	1	4	1	0	6
7	1	7	3	0	11
8	1	11	10	1	23
9	1	17	24	5	47

n	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10	k=11	Total Linear
10	25	56	22	1	0	0	0	0	0	0	0	105
11	36	114	74	6	0	0	0	0	0	0		231
12	50	224	219	37	1	0	0	0	0	0	0	532
13	70	441	576	158	8	0	0	0	0	0	0	1,254
14	94	733	1,394	591	58	1	0	0	0	0	0	2,872
15	127	1,252	3,150	1,896	304	9	0	0	0	0	0	6,739
16	168	2,091	6,733	5,537	1,342	82	1	0	0	0	0	15,955
17	222	3,393	13,744	14,812	5,085	508	11	0	0	0	0	37,776
18	288	5,408	26,969	37,133	17,232	2,635	112	1	0	0	0	89,779
19	375	8,440	51,185	87,841	53,200	11,523	804	12	0	0	0	213,381
20	480	12,982	94,323	198,267	152,316	44,704	4,730	145	1	0	0	507,949
21	616	19,650	169,453	429,199	409,105	156,513	23,451	1,182	14	0	0	1,209,184
22	781	29,388	297,533	896,731	1,040,846	504,869	102,186	7,862	184	1	0	2,880,382
23	990	43,394	512,006	1,814,978	2,526,691	1,517,918	400,074	43,602	1,682	15	0	6,861,351
24	1,243	63,430	865,050	3,572,810	5,887,488	4,300,385	1,434,484	211,388	12,381	226	1	16,348,887
25	1,562	91,754	1,437,739	6,858,774	13,231,478	11,567,238	4,773,006	915,546	75,951	2,288	17	38,955,354

Num Vertices	Nonlinear Trees	% Nonlinear	Linear Trees	Total
10	1	0.9%	105	106
11	4	1.7%	231	235
12	19	3.4%	532	551
13	47	3.6%	1,254	1,301
14	287	9.1%	2,872	3,159
15	1,002	13%	6,739	7,741
16	3,365	17%	15,955	19,320
17	10,853	22%	37,776	48,629
18	34,088	28%	89,779	123,867
19	104,574	33%	213,381	317,955
20	315,116	38%	507,949	823,065
21	935,321	44%	1,209,184	2,144,505
22	2,743,374	49%	2,880,382	5,623,756
23	7,966,723	54%	6,861,351	14,828,074
24	22,951,010	58%	16,348,887	39,299,897
25	65,681,536	63%	38,955,354	104,636,890

4.2 Discussion of the Closed Formulas of $s_{n,k}$

We will discuss in depth the derivation of the closed formula for $s_{n,k}$.

Any term where the power of y in $S(x,y)$ is greater than 2 will include factors for the generating functions of some non-central interior stars and non-central paths. In the $S(x,y)$, these terms appear as $(T^*_{int}(x))^{j-2}(s^*(x))^{j-2}$ when the power of y is even, and $(T^*_{int}(x))^{j-2}(s^*(x))^{j-1}$ when the power of y is odd for $j \geq 2$. When k is even, a k-linear tree will contain $k-2$ non-central interior stars and paths, and when k is odd, a k-linear tree will contain $k-3$ non-central interior stars and $k-1$ paths. Note that the number of non-central interior stars is always even, and there are always at least as many non-central paths as there are non-central interior stars. We will create a recursive function that counts $2l$ interior stars and $2l$ paths on $2m$ vertices in the linearly symmetric case. Due to the symmetry, this is same as counting l interior stars and l paths on m vertices. Using the closed formulas for counting interior stars and paths, we can create the following recursive formula:

$$s_{rec}(m,l) = \begin{cases} \sum_{i=0}^{m}\sum_{j=0}^{m-i} p(1+i)s_{rec}(m-i-j,l-1) & \text{if } l \geq 1 \\ 1 & \text{if } l=0 \text{ and } m=0 \\ 0 & \text{otherwise} \end{cases}$$

This recursive formula allows us to carefully extract the coefficients of $(T^*_{int}(x))^{l-2}(s^*(x))^{l-2}$ as it appears in $S(x,y)$. However, there are other factors in the generating function whose coefficients we must still compute. If $k \geq 3$ is odd, then we must count the two non-central exterior stars, two non-central paths, and the central interior star. These uncounted factors in the generating function are $T^*_{ext}(x)s(x)^*T_{int}(x)$. If $k \geq 4$ is even, then the uncounted factors are $T^*_{ext}(x)s(x)$. We provide the closed formulas for coefficients of $S(x,y)$ for powers of y that are odd and larger than 3.

Recall that the coefficients of $T_{ext}(x)s(x)T_{int}(x)$ are

$$\sum_{i=3}^{n-2}\sum_{t=0}^{n-2-i}(p(i-1)-1)p(n-i-t) = \sum_{i=0}^{n-5}\sum_{t=0}^{n-5-i}(p(i+2)-1)p(n-i-t).$$

This is similar to the coefficient of $T^*_{ext}(x)s^*(x)T_{int}(x)$ except that $T^*_{ext}(x) = T_{ext}(x^2)$ and $s^*(x) = s(x^2)$. So, there can only be $\lfloor\frac{n-5}{2}\rfloor$ coefficients in $T^*_{ext}(x)$ and $\lfloor\frac{n-5-i}{2}\rfloor$ coefficients of $s^*(x)$ that will impact the nth coefficient of this product. In particular, the coefficients of $T^*_{ext}(x)s^*(x)T_{int}(x)$ are $\sum_{i=0}^{\lfloor\frac{n-5}{2}\rfloor}\sum_{t=0}^{\lfloor\frac{n-5-i}{2}\rfloor}(p(i+2)-1)p(n-2i-2t)$. Given this, we can compute the closed formula for linearly symmetric trees on an odd number of high degree vertices as:

$$s_{odd}(n,k) =$$

$$\sum_{i=0}^{\lfloor\frac{n-2-2k}{2}\rfloor}\sum_{j=0}^{\lfloor\frac{n-2-2k-2i}{2}\rfloor}\sum_{t=0}^{\lfloor\frac{n-2-2k-2i-2j}{2}\rfloor}\Big(p(2+i)-1\Big)p(n-1-2k-2i-2j-2t)s_{rec}(j,\frac{k-3}{2})$$

When there are an even number of high degree vertices, we must account for the factor $T^*_{ext}(x)s(x)$. So, we can compute the closed formula for linearly symmetric trees with an even number of high degree vertices as:

$$s_{even}(n,k\geq 4) = \sum_{i=0}^{\lfloor\frac{n-2-2k}{2}\rfloor}\sum_{j=0}^{\lfloor\frac{n-2-2k-2i}{2}\rfloor}\Big(p(2+i)-1\Big)s_{rec}(j,\frac{k-2}{2})$$

This collection of closed formulas gives us a complete description of the coefficients of $S(x,y)$ as follows:

$$s_{n,k} = \begin{cases} s_{n,2} & \text{if } k=2 \\ s_{odd}(n,k\geq 3) & \text{if } k\geq 3 \text{ is odd} \\ s_{even}(n,k\geq 4) & \text{if } k\geq 4 \text{ is even} \end{cases}$$

Bibliography

[1] S. P. Buckley, J. G. Corliss, C. R. Johnson, C. A. Lombardia, C. Saiago. Questions, Conjectures, and Data about Multiplicity Lists for Trees.

[2] C. R. Johnson, A. Leal-Duarte, C. M. Saiago. Inverse eigenvalue problems and lists of multiplicities of eigenvalues for matrices whose graph is a tree: the case of generalized stars and double generalized stars. *Linear Algebra and its Applications* 373:311-330 (2003).

[3] C. R. Johnson, A. Leal-Duarte, C. M. Saiago. The Parter-Weiner Theorem: Revinement and Generalization. *SIAM Journal on Matrix Analysis and Applications* 25(2):562-361 (2003).

[4] C. R. Johnson, A. Leal-Duarte. The Maximum Multiplicity of an Eigenvalue in a Matrix Whose Graph is a Tree. *Linear and Multilinear Algebra* 46:139-144 (1999).

[5] C. R. Johnson, A. Leal-Duarte. On the Minimum Number of Distinct Eigenvalues for a Symmetric Matrix Whose Graph is a Given Tree. *Mathematical Inequalities and Applications* 5(2):175-180 (2002).

[6] C. R. Johnson, A. A. Li, A. J. Walker. Ordered Multiplicity Lists for Eigenvalues of Symmetric Matrices Whose Graph is a Linear Tree. *Discrete Mathematics* 333:39-55 (2014).

[7] OEIS Foundation Inc. (2011), The On-Line Encyclopedia of Integer Sequences, http://oeis.org/A000055.

[8] OEIS Foundation Inc. (2011), The On-Line Encyclopedia of Integer Sequences, http://oeis.org/A000041.

[9] A. Tucker. *Applied Combinatorics* 6th ed. Hoboken: John Wiley & Sons, Inc. 2012.

[10] C. R. Johnson, C. Jordan-Squire, and D. A. Sher. Eigenvalue assignments and the two largest multiplicities in a Hermitian matrix whose graph is a tree. Discrete Applied Mathematics 158:681-691 (2010).

[11] R. A. Horn, and C. R Johnson. *Matrix Analysis* 2nd ed. New York: Cambridge University Press. 2013.

[12] This work was performed [in part] using computational facilities at the College of William and Mary which were provided with the assistance of the National Science Foundation, the Virginia Port Authority, Sun Microsystems, and Virginia's Commonwealth Technology Research Fund.