THE PROPORTION OF TREES THAT ARE LINEAR

TANAY WAKHARE*, ERIC WITYK†, AND CHARLES R. JOHNSON§

Abstract. We study several enumeration problems connected to linear trees, a broad class which includes stars, paths, generalized stars, and caterpillars. We provide generating functions for counting the number of linear trees on \( n \) vertices, and prove that they form an asymptotically vanishing fraction of all trees on \( n \) vertices.

1. Introduction

A high degree vertex (HDV) in a simple undirected graph is one of degree at least 3. A tree is called linear if all of its HDV’s lie on a single induced path, and \( k \)-linear if there are \( k \) HDV’s. The linear trees include the familiar classes of paths, stars, generalized stars (g-stars, with exactly one HDV), double g-stars [3], and caterpillars [2], etc. They have become important, as all multiplicity lists of eigenvalues occurring among Hermitian matrices, whose graph is a given linear tree, may be constructed via a linear superposition principal (LSP) that respects the precise structure of the linear tree [3, 4]. For other, nonlinear trees, multiplicity lists require different methodology. For a tree to be nonlinear, there must be at least 4 HDV’s (and at least 10 vertices altogether). An example of a nonlinear tree and a linear tree, both on 13 vertices, is given in Figure 1.1.

Linear trees are a substantial generalization of caterpillars, and the problem of counting the number of non-isomorphic linear trees is significantly harder than for caterpillars. We define a bivariate generating function for the number of \( k \)-linear trees on \( n \) vertices, which enables the fast computation of these numbers. Additionally, we are able to obtain asymptotic upper bounds which show that the probability that a randomly chosen tree on \( n \) vertices will be linear approaches 0 as \( n \to \infty \). This shows that while the LSP is a useful characterization, it has limited applicability to studying the spectra of general trees. As \( n \) increases, the LSP characterizes the spectra of an asymptotically vanishing proportion of all trees. However, the proportion of linear trees vanishes slowly, so that the LSP is very important, especially for small numbers of vertices. This preliminary investigation suggests many open questions about linear trees – for instance, for fixed \( n \), what value of \( k \) maximizes the number of \( k \)-linear trees on \( n \)-vertices?

MSC(2010): Primary: 05C30; Secondary: 05C05.
Keywords: Linear trees; Caterpillars.
§ Corresponding author.
2. Generating Functions

There are strong links between nonisomorphic linear trees and partitions, which are famously difficult to enumerate. In constructing a generating function for $k$-linear trees on $n$ vertices, we will rely the generating function for integer partitions. Let

$$P(x) := \prod_{i=1}^{\infty} \frac{1}{1 - x^i} = \sum_{n=0}^{\infty} p(n) x^n$$

denote the generating function for $p(n)$, the number of unrestricted partitions of $n$. Let $r_{n,k}$ be the number of reflections of linear trees on $n$ vertices with $k$ HDVs (which counts linearly symmetric trees once and linearly asymmetric trees twice), and let $s_{n,k}$ denote the number of linearly symmetric trees on $n$ vertices with $k$ HDVs. We can conclude that the number of non-isomorphic $k$-linear trees on $n$ vertices is equal to

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

The following generating function allows us to compute recurrences for the coefficients which allow for fast computation of $a_{n,k}$.

**Theorem 1.** The generating function for $k$-linear trees on $n$ vertices is

$$2 \sum_{n=1}^{\infty} \sum_{k=0}^{\infty} a_{n,k} x^n y^k = \left( P(x) - \frac{1}{1-x} \right)^2 \frac{x^2 y^2}{(1 - x + xy - xy P(x))}$$

$$+ \frac{1}{1-x} \left( P(x^2) - \frac{1}{1-x^2} \right) \frac{x^2 y^2}{1 - \frac{(P(x^2)-1)x^2y^2}{1-x^2}}$$

$$+ \left( P(x^2) - \frac{1}{1-x^2} \right) \frac{P(x)-1}{1-x^2} \frac{x^3 y^3}{1 - \frac{(P(x^2)-1)x^2y^2}{1-x^2}}.$$

**Proof.** First, we enumerate the nonisomorphic generalized stars on $n$ vertices. Since two g-stars are non-isomorphic if and only if the lengths of their arms differ, we notice a one-to-one correspondence between nonisomorphic generalized stars and partitions. In particular, the number of nonisomorphic g-stars on $n$ vertices is $p(n-1)$ (the $-1$ accounting for the designated central vertex), with each partition corresponding to a distinct set of possible arm lengths. Linear trees are formed from generalized stars on $\geq 2$ vertices, with intermediate paths which can have non-trivial length. Therefore, we will use the generating function for the number of non-isomorphic generalized stars on $\geq 2$ vertices, which is $x(P(x) - 1)$.

Let an exterior star be a generalized star at the end of the linear tree. Such stars must have a central vertex of degree $\geq 2$, not counting the concatenating path. Therefore, there is a bijection between partitions of $n - 1$ with $\geq 2$ parts and non-isomorphic exterior stars on $n$ vertices. Since there is only a single partition of $n$ with one part, $n$ itself, the generating function for exterior stars is $x(P(x) - \frac{1}{1-x})$. Additionally, up to isomorphism, there is a unique path of length $i$, so that the generating function for the number of paths on $n$ vertices has the form $\frac{1}{1-x^i}$. 

---

**Table 1.** Appendix A The number of $k$-linear trees on $n$ vertices

| $n$ | $k = 1$ | $k = 2$ | $k = 3$ | $k = 4$ | $k = 5$ | $k = 6$ | $k = 7$ | $k = 8$ | $k = 9$ | $k = 10$ | $k = 11$ | Total |
|-----|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|------|
| 10  | 25      | 56      | 22      | 1       | 0       | 0       | 0       | 0       | 0       | 0       | 0       | 105  |
| 11  | 36      | 114     | 74      | 6       | 0       | 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     | 0       | 0       | 2,880,362|
| 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| 4,300,355| 1,434,484| 211,388 | 12,381  | 226     | 1       | 16,348,887|
| 25  | 1,562   | 91,754  | 1,497,739| 6,858,774 | 13,231,478| 11,567,238| 4,773,006| 915,546 | 75,951  | 2,288   | 17      | 38,955,354|
Therefore, the number of linear trees generated by concatenating an exterior star, \( k - 1 \) interior stars, and a trailing exterior star, by \( k - 1 \) paths of arbitrary length, is

\[
\sum_{n=1}^{\infty} \sum_{k=0}^{\infty} r_{n,k} x^n y^k = \sum_{k=2}^{\infty} \left( xP(x) - \frac{x}{1-x} \right)^2 (xP(x) - x)^{k-2} \frac{1}{(1-x)^{k-1}} y^k \frac{1}{(1-x)^k y^k} 
\]

\[
= \left( P(x) - \frac{1}{1-x} \right)^2 \sum_{k=2}^{\infty} (P(x) - 1)^{k-2} \frac{1}{(1-x)^{k-1}} x^k y^k 
\]

\[
= \left( P(x) - \frac{1}{1-x} \right)^2 \frac{x^2y^2}{(1-x + xy - xyP(x))}. 
\]

We now enumerate \( s_{n,k} \), the number of reflectionally symmetric \( k \)-linear trees on \( n \) vertices. These have a freely chosen central component, after which one half of the tree completely determines the other half. The component is a path when \( k \) is even, and a generalized star when \( k \) is odd.

If the central component is a path, it is free to have an arbitrary number of vertices, while every other component on \( n \) vertices determines \( 2n \) vertices due to reflectional symmetry. Therefore, we count the number of \( 2k \)-linear trees which can be generated by concatenating an exterior star, \( k - 1 \) interior stars, a freely chosen central path, and their reflections:

\[
\sum_{k=2}^{\infty} \left( x^2P(x^2) - \frac{x^2}{1-x^2} \right) (x^2P(x^2) - x^2)^{k-2} \frac{1}{(1-x^2)^{k-2}} \frac{1}{1-x} y^{2k-2} 
\]

\[
= \frac{1}{1-x} \left( P(x^2) - \frac{1}{1-x^2} \right)^2 \sum_{k=2}^{\infty} (P(x^2) - 1)^{k-2} \frac{1}{(1-x^2)^{k-2}} x^{2k-2} y^{2k-2} 
\]

\[
= \frac{1}{1-x} \left( P(x^2) - \frac{1}{1-x^2} \right)^2 \frac{x^2y^2}{(1 - (P(x^2) - 1)x^2y^2)} 
\]

We can conduct a similar analysis for a \((2k + 1)\) -linear tree, where the central component is instead a generalized star. We obtain the generating function

\[
\sum_{k=2}^{\infty} \left( x^2P(x^2) - \frac{x^2}{1-x^2} \right) (x^2P(x^2) - x^2)^{k-2} \frac{1}{(1-x^2)^{k-1}} (xP(x) - x)y^{2k-1} 
\]

\[
= \left( P(x^2) - \frac{1}{1-x^2} \right) (P(x) - 1) \sum_{k=2}^{\infty} (P(x^2) - 1)^{k-2} \frac{1}{(1-x^2)^{k-1}} x^{2k-1} y^{2k-1} 
\]

\[
= \left( P(x^2) - \frac{1}{1-x^2} \right) \left( P(x) - \frac{1}{1-x^2} \right) \frac{x^3y^3}{(1 - (P(x^2) - 1)x^2y^2)}. 
\]

Noting that \( 2a_{n,k} = r_{n,k} + s_{n,k} \) and summing all three of these generating functions completes the proof. \( \square \)

From this generating function, we can extract the number of \( k \)-linear trees on \( n \) vertices for small values of \( n \). Table 7 displays this information for \( 10 \leq n \leq 25 \). Note that for fixed \( n \), the distribution of \( k \)-linear trees is not uniform or normal. Instead, the dominant contribution appears at around \( k \approx 0.3n \). It would be interesting to see what distribution the number of \( k \)-linear trees for fixed \( n \) assumes as we take \( n \rightarrow \infty \).
We wish to show that linear trees form an asymptotically small subset of all trees. Wityk \[6\] showed that the fraction of $k$-linear trees on $n$-vertices to the number of trees with $k$ high degrees vertices approaches 0 as the number of vertices tends to infinity. However, this was only for a fixed $k$, and only partial results were shown for the natural extension to account for all linear trees. Heuristically, we expect the proportion of trees that are linear to decrease as the number of vertices increases. Given a large tree, we can color all the high degree vertices. The probability that these HDVs all lie on a single induced path intuitively decreases as the number of vertices increases. The next theorem asymptotically proves this result, and Table 3 shows this phenomenon for small values of $n$.

**Theorem 2.** [6, Conjecture 3.4.7] The number of nonisomorphic linear trees on $n$ vertices grows at the rate $O(2^{7.46n})$. Hence, the probability that a randomly chosen tree on $n$ vertices will be linear approaches 0 as $n \to \infty$.

**Proof.** To begin, fix a finite $n$. We then wish to consider the number of linear trees on $n$ vertices, up to isomorphism. It is known that the number of nonisomorphic unlabelled trees on $n$ vertices, $T_n$, is asymptotically given by $T_n \sim Cn^{-\frac{5}{3}}\alpha^n$ where $C = 0.5349\ldots$ and $\alpha = 2.9557\ldots$. If we can show that the number of linear trees on $n$ vertices has an asymptotically smaller growth rate, we are done.

For fixed $n$, let the number of linear trees on $n$ vertices with an induced central path of length $k$ be denoted $T(k)$ (note that this is not equivalent to $k$-linearity). We will analyze an explicit expression for $T(k)$ with some overcounting. However, because the condition for a tree to be linear is so strong, this overcounting does not result in any difficulties. First, consider the following alternate construction of a linear trees from generalized stars:

- construct $k$ generalized stars $\{G_1, G_2, \ldots, G_k\}$ (where we now allow $|G_i| \geq 1$);
- connect the central vertices of $\{G_1, G_2, \ldots, G_k\}$ by single edges, in that order.

Note that every linear tree, up to isomorphism, is counted by this construction, since one of length longer than $k$.

We again note that we have a correspondence between partitions of $n - 1$ and non-isomorphic generalized stars on $n$ vertices. Hence, letting $n_i$ denote the size of the $i$-th g-star in our construction, we wish to estimate the multiple sum

\[
T(k) \leq \sum_{n_1 + n_2 + \cdots + n_k = n} p(n_1 - 1)p(n_2 - 1)\cdots p(n_k - 1)
\]

(3.1)

\[
= \sum_{n_1 + n_2 + \cdots + n_k = n-k} p(n_1)p(n_2)\cdots p(n_k).
\]

(3.2)
We now appeal to the well-known asymptotic expansion for the partition function [11, p. 97]

\[ p(n) \sim \frac{1}{4\sqrt{3n}} \exp \left( \pi \sqrt{\frac{2n}{3}} \right), \tag{3.3} \]

the expansion which led to the introduction of the circle method. Substituting into (3.2) gives

\[ T(k) \leq \sum_{n_1 + n_2 + \cdots + n_k = n - k} p(n_1)p(n_2)\cdots p(n_k) \sim \sum_{n_i \geq 0} \frac{1}{(4\sqrt{3})^k} \prod_{i=1}^{k} n_i \exp \left( \pi \sqrt{\frac{2}{3} \sum_{i=1}^{k} \sqrt{n_i}} \right). \]

Applying the method of Lagrange multipliers to the \( k \) variable function

\[ \frac{1}{\prod_{i=1}^{k} n_i} \exp \left( \pi \sqrt{\frac{2}{3} \sum_{i=1}^{k} \sqrt{n_i}} \right) \]

subject to the constraints \( n_i \geq 0 \) and \( \sum_{i=1}^{k} n_i = n - k \), we have the estimate

\[ T(k) = O \left( \frac{1}{(4\sqrt{3})^k} \prod_{i=1}^{k} n_i \exp \left( \pi \sqrt{\frac{2}{3} \sum_{i=1}^{k} \sqrt{n_i}} \right) \right) \]

subject to the constraints \( n_i \geq 0 \) and \( \sum_{i=1}^{k} n_i = n - k \) shows that the function achieves its maximum value at

\[ n_i = \left( \frac{n}{k} - 1 \right) \geq 0 \text{ (since } k < n\text{)}, \]

and that this maximum value is

\[ \frac{1}{\left( \frac{n}{k} - 1 \right)^k} \exp \left( \pi \sqrt{\frac{2}{3} (n-k)k} \right). \]

Hence, noting that the Diophantine equation \( n_1 + n_2 + \cdots + n_k = n - k \) has \((n-1)\choose (k-1)\) solutions in non-negative integers by a stars and bars argument, we have the estimate

\[ T(k) = O \left( \frac{1}{(4\sqrt{3})^k} \prod_{i=1}^{k} n_i \exp \left( \pi \sqrt{\frac{2}{3} (n-k)k} \right) \right). \]

\[ = O \left( \frac{1}{(4\sqrt{3})^k} \frac{n}{k} \exp \left( \pi \sqrt{\frac{2}{3} (n-k)k} \right) \frac{n}{k} \right). \]

Note that, if \( k \) is a fixed constant, \((n)\choose (k)\) grows polynomially in \( n \), and the result is proven for any fixed finite \( k \).

Therefore, now let \( k = Cn \), \( 0 < C \leq 0.5389 \). Then we have to appeal to Stirling’s asymptotic expansion \( n! \sim \left( \frac{n}{e} \right)^n \sqrt{2\pi n} \) to write

\[ \left( \frac{n}{Cn} \right) \sim \frac{1}{\sqrt{2\pi n C(1-C)}} \frac{n}{(1-C)} \left( \frac{1-C}{C} \right)^{Cn}. \]

Our asymptotic formula for \( T(k) \) then reduces to

\[ T(k) = O \left( \frac{1}{(4\sqrt{3})^k} \frac{n}{k} \exp \left( \pi \sqrt{\frac{2C(1-C)}{3}} \right) \frac{n}{Cn} \right) \]

\[ = O \left( \frac{1}{(4\sqrt{3})^k} \frac{n}{k} \exp \left( \pi \sqrt{\frac{2C(1-C)}{3}} \right) \frac{n}{Cn} \right) \]

\[ = O \left( \frac{1}{(4\sqrt{3})^k} \frac{n}{k} \exp \left( \pi \sqrt{\frac{2C(1-C)}{3}} \right) \frac{n}{Cn} \right). \]

Numerical examination reveals that

\[ \frac{1}{1-C} \left( \frac{1}{4\sqrt{3}} \right)^C \exp \left( \pi \sqrt{\frac{2C(1-C)}{3}} \right) < 2.744 \]

whenever \( C \leq 0.5389 \). Hence, in this regime \( T(k) = O(2.745^n) \).

**Case 2:** \( 0.5389n \leq k \leq 1 \).

In this regime, we have many extremely small components. For instance, we cannot even have a single component of size \( \frac{n}{k} \), since we have \( k \geq \frac{n}{C} \) star components. While the asymptotic expansion (3.3) is good for large values of \( n \), it is a poor bound for small \( n \). Therefore, we instead appeal to the easy bound \( p(n) \leq 2^{n-1} \), which counts the number of compositions of \( n \).
We have the simple estimate
\[ T(k) \leq \sum_{n_1+n_2+\cdots+n_k=n} p(n_1-1)p(n_2-1)\cdots p(n_k-1) \]
\[ \leq \sum_{n_1+n_2+\cdots+n_k=n} 2^{n_1-1}2^{n_2-1}\cdots 2^{n_k-1} \]
\[ = 2^{n-k} \sum_{n_1+n_2+\cdots+n_k=n} 1 \]
\[ = 2^{n-k} \binom{n-1}{k-1} \]
\[ \leq 2^{n-k} \binom{n}{k}. \]

Again writing \( k = Cn \), \( .5389 \leq C \leq 1 \), and using Stirling’s expansion to asymptotically expand the binomial coefficient gives
\[ T(k) = O \left( \left( \frac{2^{1-C}}{1-C} \left( \frac{1-C}{C} \right)^C \right)^n \right). \]

For \( .5389 \leq C \leq 1 \), numerical examination reveals
\[ \frac{2^{1-C}}{1-C} \left( \frac{1-C}{C} \right)^C < 2.745, \]
so that \( T(k) = O(2.745^n) \). Combining both cases reveals that the total number of linear trees on \( n \) vertices is
\[ \sum_{k=1}^{n} T(k) = O(n2.745^n) = O(2.746^n), \]
which is dominated by \( T_n \sim Cn^{-\frac{5}{2}}2.9557^n \), completing the proof. \( \square \)

Note that we have still overcounted some classes of linear trees, and that the true asymptotic growth rate of the number of linear trees on \( n \) vertices may be even smaller. We count reflectionally symmetric linear trees twice. We also overcount linear trees which correspond to compositions of \( n \) with leading or trailing ones, as illustrated in Figure 3.1 since we can regard some vertices both as degenerate star centers or as parts of an arm.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{overcounting}
\caption{Blue vertices lead to overcounting, since they can be regarded either as degenerate star centers or as part of an arm}
\end{figure}

\section*{Acknowledgements}

Part of this work was carried out by the second author as part of his honors thesis at the College of William and Mary. T.W. would also like to thank Roberto Costas-Santos for being an excellent advisor during the College of William and Mary Matrix REU, where part of this work was completed.
References

[1] G. Andrews, The theory of partitions. Addison-Wesley Publishing Co., Reading, Mass, 1976
[2] F. Harary and A. J. Schwenk, The number of caterpillars, Discrete Math., 6, 359–365, 1973
[3] C. R. Johnson and C. M. Saiago, Eigenvalues, multiplicities and graphs. Cambridge University Press, Cambridge, 2018
[4] C. R. Johnson, A. A. Li, and A. J. Walker, Ordered multiplicity lists for eigenvalues of symmetric matrices whose graph is a linear tree, Discrete Math., 333, 39–55, 2014
[5] R. Otter, The number of trees, Ann. of Math. (2), 44, 583–599, 1948
[6] E. Wityk, Linear and Nonlinear Trees: Multiplicity Lists of Symmetric Matrices, College of William and Mary, Undergraduate Honors Theses, Paper 113, 1–55, 2014

* University of Maryland, College Park, MD 20742, USA
  E-mail address: twakhare@gmail.com

† Georgia Institute of Technology, Atlanta, GA 30332, USA
  E-mail address: eric.wityk@gmail.edu

§ College of William and Mary, Williamsburg, VA 23185, USA
  E-mail address: crjohn@wm.edu