

PROTECTED VERTICES IN MOTZKIN TREES

ANTHONY VAN DUZER

ABSTRACT. In this paper we find recurrence relations for the asymptotic probability a vertex is k protected in all Motzkin trees. We use a similar technique to calculate the probabilities for balanced vertices of rank k . From this we calculate upper and lower bounds for the probability a vertex is balanced and upper and lower bounds for the expected rank of balanced vertices.

1. INTRODUCTION

There have been several papers written in the past looking at various different statistics concerning vertices in various different trees. Specifically there was an explosion of papers dealing with protected points in trees after Cheon and Shapiro introduced the term [1]. These papers however were limited in that they only looked at vertices which were protected meaning they broke vertices into two groups, a vertex was either protected or not protected. There were several such papers doing this such as [2, 3, 5, 6, 7]. Most of them went about this by trying to find an explicit function for the number of trees whose root was protected and then using that to calculate the desired probability. In this paper we instead intend to introduce more gradation in protection. Rather than simply looking at a binary choice of protected or not we instead look at distinguishing between how protected a vertex is. This change in focus obviously necessitates a slight change in technique. Rather than calculating the specific generating functions we calculate a recurrence relationship for the generating function we want and then use that to calculate the probabilities we are interested in or get upper and lower bounds for the probabilities we are interested in. This paper is concerned specifically with Motzkin trees and k -protectedness which we will define later, but in future papers I will apply the technique to other types of trees and perhaps other types of vertex statistics that might be interesting to look at.

To begin we will first go over many of the needed definitions.

1.1. Motzkin Tree. A Motzkin tree, also referred to as a 0-1-2 tree, or unary-binary tree, is a rooted plane tree where each vertex may have either 0, 1, or 2 children.

Theorem 1.1. *The generating function, $M(x) = \sum m_n x^n$, for the number of all Motzkin trees with n vertices is given by $M(x) = \frac{1 - x - \sqrt{1 - 2x - 3x^2}}{2x}$.*

Key words and phrases. Enumerative Combinatorics, Motzkin trees.

Proof. This comes from the relationship $M(x) = x + xM(x) + xM(x)^2$ which is based on the fact the root can be a leaf, the parent of a single child, or the parent of two children and each child of the root would be another Motzkin tree. Given that relationship you can use the quadratic formula to arrive at the desired generating function. \square

1.2. k -protected. A vertex is said to be of rank k if the shortest path, travelling strictly from parent to child, from the vertex to a leaf is of length k . A vertex is k protected if it is of rank j for some $j \geq k$. So a leaf would be 0 protected and rank 0, the parent of a leaf is rank 1 and both 0 and 1 protected, and the parent of children who are all 1 protected is 2 protected, but not necessarily rank 2. One of the advantages of working with k -protected over rank k is the recurrent nature of protection; a vertex is k protected if and only if all it's children are $k - 1$ protected the same statement does not hold true for rank k since a vertex could be rank 6 and have a child of rank 7. A vertex is balanced if the shortest path is the same length as the longest path so a balanced k -protected vertex would be a vertex where all paths to a leaf are of length at least k and all of these paths are the exact same length and a balanced vertex of rank k is a vertex where all paths to a leaf are of length exactly k .

1.3. Purpose of paper. In this paper we will find a recurrence relation for the probability a vertex is k -protected, we will find a recurrence relation for the probability a vertex is balanced and of rank k , and finally we will get upper and lower bounds for the probability a vertex is balanced and the expected rank of a balanced vertex.

2. LEAVES

The proportion of leaves in Motzkin trees had previously been calculated by Gi-Sang Cheon and Louis W. Shapiro in [1]; however, for the sake of completeness and to give a brief introduction to the technique used in this paper it is included.

Theorem 2.1. *The generating function for the number of leaves in all Motzkin trees with n vertices is given by $L(x) = \frac{x}{\sqrt{1 - 2x - 3x^2}}$.*

Proof. We observe the relationship $L(x) = x + xL(x) + 2xL(x)M(x)$. To prove this we isolate the root. If the root is a leaf it contributes x to the generating function. If the root is not a leaf it either has one child or two children. If it has a single child then the number of leaves the tree possess is simply the number of leaves of the subtree giving rise to the term $xL(x)$. If it has 2 children than we can again cut off the root creating two subtrees. On one subtree we count the total number of possible leaves which is given by $L(x)$; we multiply this by the total number of configurations. The total number of configurations with that many leaves is the total number of trees we can make for the right subtree which is simply the number of Motzkin trees hence the term $M(x)$. The power series $xL(x)M(x)$ is the generating function for the total number of leaves only counting the left subtree. Once we multiply this by 2 we get the total number of leaves

in all trees where the root has two children. Plugging in the known $M(x)$ and doing some algebraic manipulation we arrive at the given generating function. \square

Corollary 2.1.1. *The probability that a random vertex of a random Motzkin-tree is a leaf converges to $\frac{1}{3}$ as the number of vertices goes to infinity.*

Proof. We can extract the coefficient from the generating function and we get that $l(n)$, the total number of leaves in all Motzkin trees of size n , is asymptotically equal to

$$\frac{\left(\sqrt{\frac{3}{\pi}}\right) 3^n}{2\sqrt{n}}$$

whereas the number of vertices is asymptotically equal to

$$\frac{n3^{n+1}\sqrt{3}\left(1 + \frac{1}{16n}\right)}{(2n + 3)\sqrt{(n + 2)\pi}}.$$

Comparing the two and taking the limit as n goes to ∞ we get the ratio is $\frac{1}{3}$. \square

This is also the probability a vertex is balanced and of rank 0 since every leaf is balanced.

3. k -PROTECTED VERTICES

Lemma 3.1. *The generating function for the number of vertices that are k protected in all trees with n vertices is given by*

$$P_k(x) = \frac{R_k(x)}{\sqrt{1 - 2x - 3x^2}} = \frac{xR_{k-1}(x) + xR_{k-1}^2(x)}{\sqrt{1 - 2x - 3x^2}}$$

where $P_k(x)$ is the generating function for all k -protected vertices in all Motzkin tree with n -vertices and $R_k(x)$ is the number of trees where the root is k -protected.

Proof. This proof comes in two parts. The first part is that $P_k(x) = \frac{R_k(x)}{\sqrt{1 - 2x - 3x^2}}$ and the second is that $R_k(x) = xR_{k-1}(x) + xR_{k-1}^2(x)$.

First we will prove that $P_k(x) = \frac{R_k(x)}{\sqrt{1 - 2x - 3x^2}}$. This follows from the same recurrence relationship we used to find the generating function for the number of leaves. We have $P_k(x) = R_k(x) + xP_k(x) + 2xP_k(x)M(x)$ which gives us

$$P_k(x) = \frac{R_k(x)}{\sqrt{1 - 2x - 3x^2}} \text{ after a bit of manipulation.}$$

Next we need to prove the relationship $R_k(x) = xR_{k-1}(x) + xR_{k-1}^2(x)$. This comes from the fact the root of a tree can only be k protected is if it has a single child who is $k - 1$ protected or it has two children who are both $k - 1$ protected. \square

Lemma 3.2. (*Bender's Lemma*) [4]

Take generating functions $A(x) = \sum a_n x^n$ and $B(x) = \sum b_n x^n$ with radius of convergence $\alpha > \beta \geq 0$ where α goes with $A(x)$ and β goes with $B(x)$. If $\frac{b_{n-1}}{b_n}$ approaches a limit b as n approaches infinity and $A(b) \neq 0$ then $c_n \sim A(b)b_n$ where $\sum c_n x^n = A(x)B(x)$.

Theorem 3.3. Let p_k be the asymptotic proportion of all k -protected vertices in all Motzkin trees compared to all vertices in all Motzkin trees then $p_k = \frac{1}{3}p_{k-1} + \frac{1}{3}p_{k-1}^2$.

Proof. Consider R_k . It is of the form $R_k = \frac{A_k + B_k(\sqrt{1-2x-3x^2})}{2x}$ where A_k and B_k are polynomials. This means that

$$R_{k+1} = x \frac{A_k^2}{(2x)^2} + x \frac{A_k}{2x} + x \frac{B_k^2(1-2x-3x^2)}{(2x)^2} + x \left(\frac{B_k}{2x} + \frac{2A_k B_k}{(2x)^2} \right) \sqrt{1-2x-3x^2}.$$

Now we go back to look at $P_k(x)$. We have

$$P_{k+1}(x) =$$

$$\frac{x \frac{A_k^2}{(2x)^2} + x \frac{A_k}{2x} + x \frac{B_k^2(1-2x-3x^2)}{(2x)^2} + x \left(\frac{B_k}{2x} + \frac{2A_k B_k}{(2x)^2} \right) \sqrt{1-2x-3x^2}}{\sqrt{1-2x-3x^2}}.$$

We can now simplify this and we get

$$P_{k+1}(x) =$$

$$\frac{x \frac{A_k^2}{(2x)^2}}{\sqrt{1-2x-3x^2}} + \frac{x \frac{A_k}{2x}}{\sqrt{1-2x-3x^2}} + \frac{\frac{B_k^2(1-2x-3x^2)}{(2x)^2}}{\sqrt{1-2x-3x^2}} + x \left(\frac{B_k}{2x} + \frac{2A_k B_k}{(2x)^2} \right).$$

Now we look at what each of these contribute to the asymptotic behavior of $P_{k+1}(x)$. We know that $x \left(\frac{B_k}{2x} + \frac{2A_k B_k}{(2x)^2} \right)$ is just a polynomial so this summand doesn't contribute anything to the asymptotic behavior. This is because they are polynomials so there exist some m such that for all $r > m$ the coefficient of x^r in $x \left(\frac{B_k}{2x} + \frac{2A_k B_k}{(2x)^2} \right)$ is equal to zero. Because the terms are eventually zero this part cannot contribute anything to the asymptotic behavior of the coefficients of the generating function we care about. That leaves us with

$$\frac{x \frac{A_k^2}{(2x)^2}}{\sqrt{1-2x-3x^2}} + \frac{x \frac{A_k}{2x}}{\sqrt{1-2x-3x^2}} + \frac{\frac{B_k^2(1-2x-3x^2)}{(2x)^2}}{\sqrt{1-2x-3x^2}}.$$

We would now like to apply Bender’s Lemma with the growth rate of $\frac{1}{3}$. We get a growth rate of $\frac{1}{3}$ because that is the growth rate of the function

$$\frac{1}{\sqrt{1 - 2x - 3x^2}}.$$

We need to check that plugging in the growth rate doesn’t give 0. This requirement does pose a slight problem as if we plug in $\frac{1}{3}$ into the term $\frac{B_k^2(1 - 2x - 3x^2)}{(2x)^2}$ we would get 0. We however can get around this relatively easily by instead breaking up

$$\frac{\frac{B_k^2(1 - 2x - 3x^2)}{(2x)^2}}{\sqrt{1 - 2x - 3x^2}}.$$

Specifically we will look at it as

$$\frac{\frac{B_k^2(1)}{(2x)^2}}{\sqrt{1 - 2x - 3x^2}} - \frac{\frac{B_k^2(2x)}{(2x)^2}}{\sqrt{1 - 2x - 3x^2}} - \frac{\frac{B_k^2(3x^2)}{(2x)^2}}{\sqrt{1 - 2x - 3x^2}}.$$

Now we can handle each of these seperately. Since we just have B_k multiplied by some non-zero polynomial it is sufficient for us to prove that $b_k(\frac{1}{3})$ is non-zero. Since B_k is just based on $B_{(k-1)}$ we can see that it is nonzero by induction. Once we see that we see that the contribution of

$$\frac{\frac{B_k^2(1 - 2x - 3x^2)}{(2x)^2}}{\sqrt{1 - 2x - 3x^2}}$$

to $l(n)$ is zero. So all that is left is

$$\frac{x \frac{A_k^2}{(2x)^2}}{\sqrt{1 - 2x - 3x^2}} + \frac{x \frac{A_k}{2x}}{\sqrt{1 - 2x - 3x^2}}.$$

We know that plugging in $\frac{1}{3}$ into A_k does not return zero by induction so we can use Bender’s lemma. This gives us the desired result that $p_{k+1} = \frac{1}{3}p_k^2 + \frac{1}{3}p_k$. □

Corollary 3.3.1. *The growth rate of the asymptotic probabilities that a vertex is k protected is $\frac{1}{3}$.*

Proof. From the previous theorem we know that $p_k = \frac{1}{3}p_{k-1} + \frac{1}{3}p_{k-1}^2$. From this we can deduce that the lower growth rate is $\frac{1}{3}$ since the p_k are probabilities and hence non-negative. For the upper growth rate we observe that the p_k are strictly decreasing and non-negative. They are strictly decreasing because for every vertex that is k -protected there is at least one vertex that is $(k - 1)$ -protected, its child. This means that for

any ϵ there exists an m such that for any $l > m$, $p_l \leq \epsilon$. Thus for $k \geq m$ we have $p_{k+1} \leq \frac{1}{3}p_k + \frac{1}{3}\epsilon p_k$ that gives an upper growth rate of $\frac{1}{3} + \frac{\epsilon}{3}$ and that goes towards $\frac{1}{3}$ as ϵ goes to zero. \square

Protection level	Probability k -protected \approx
1	.66666667
2	.37037037
3	.16918153
4	.06593464
5	.02342734
6	.007992060

4. BALANCED VERTEXES

We can use the same technique as in the preceding to calculate the proportion of balanced vertices of rank k .

Lemma 4.1. *The generating function for the number of trees whose root is balanced and of rank k , call it $B_k(x)$, is a polynomial.*

Proof. If the root is balanced and rank k then all paths to a leaf must be of length k . Thus the largest tree we can have is a full binary tree with $k + 1$ levels. This means we can have at most $2^{k+1} - 1$ vertices in our tree. Thus the largest term that can appear in $B_k(x)$ is an $x^{2^{k+1}-1}$. \square

Theorem 4.2. *The generating function for the number of vertices who are k protected and balanced is given by $B_k^*(x) = \frac{B_k(x)}{\sqrt{1-2x-3x^2}}$.*

Proof. The proof is exactly the same as the proof for the number of vertices that are k protected. \square

The fact that we are only dealing with polynomials makes balanced trees and balanced vertices much easier to deal with and allows us to calculate rank k rather than k protected.

Theorem 4.3. *The asymptotic proportion of balanced vertices of rank k is given by $b_k = \frac{1}{3}b_{k-1} + \frac{1}{3}b_{k-1}^2$.*

Proof. The fact that $B_k(x)$ is a polynomial means we can simply plug in the growth rate of $\frac{x}{\sqrt{1-2x-3x^2}}$ directly into the polynomial and as we saw before that growth rate is $\frac{1}{3}$. So we plug $\frac{1}{3}$ into B_k and $B_k = xB_{k-1}(x) + xB_{k-1}^2(x)$ which gives the desired recurrence relationship. \square

It is natural now to consider the proportion of vertices that are balanced.

Lemma 4.4. *Let $B^*(x)$ equal the generating function for the number of vertices who are balanced. Then*

$$B^*(x) = \sum_{k \geq 0} B_k^*(x) = \frac{\sum_{k \geq 0} B_k(x)}{\sqrt{1 - 2x - 3x^2}}.$$

Proof. This follows from the fact that a vertex is balanced if and only if it is balanced of rank k for some k . We know that $\sum_{k \geq 0} B_k^*(x)$ is well defined as a formal power series since the coefficient of x^n in $B_k^*(x)$ is 0 for $n < k + 1$. We also know that $\sum_{k \geq 0} B_k(x)$ is well-defined. □

The form of the power series shows that we would like to use Bender’s lemma to get a value for the probability a vertex is balanced so we need to show that $\sum_{k \geq 0} B_k(x)$ is non-zero and has radius of convergence strictly greater than $\frac{1}{3}$.

Lemma 4.5. *Let $b(n)$ be the number of Motzkin trees on n vertices where the root is balanced. Then $b(n) \leq \frac{2.9^n}{n^2}$.*

Proof. This can be proved by induction and the fact that $b(n) < \sum_{k=\log_2(n)}^{n-\log_2(n)} b(k)b(n-k-1) + b(n-1)$. To see why

$$b(n) < \sum_{k=\log_2(n)}^{n-\log_2(n)} b(k)b(n-k-1) + b(n-1)$$

consider what it means for the root to be balanced. If a root is balanced then either it has a single child which is balanced. That gives us the term $b(n-1)$ or both of its children are balanced and of the same rank. The fact that they have to be balanced of the same rank is what gives us our upper and lower bounds for the sum. A tree that is balanced of rank m can have at most 2^m vertices, this would be a full binary tree and must have at least $m + 1$ vertices, this would be the stick tree. Because of that we know that a tree of size n cannot have a balanced root if one of its children’s subtrees is of size less than $\log_2(n)$. We can use a strict inequality here because there will be trees of size m with a balanced root and trees of size $n - m$ with a balanced root such that the rank of the roots are different. Once we know that fact we can apply induction to get the required inequality. □

From this we know that the exponential growth rate is 2.9 and hence it has a radius of convergence greater than 3. This fact tells us that the probability a vertex is balanced converges as n goes to infinity.

Corollary 4.5.1. *The probability that a vertex is balanced is between 0.568362259762727779 and 0.5683622597627278*

Proof. The probability a vertex is balanced is simply the sum of the probabilities that it is k balanced taken over all k . This is because if a vertex is balanced it must be balanced of some rank k . So we have $P(\text{balanced}) = \sum_{k=0}^{\infty} b_k$. Now we need to estimate this sum.

We have that $\left(\frac{1}{3}\right)^k (b_m) \leq b_{k+m} \leq \left(\frac{1}{3} + b_m\right)^k (b_m)$. We can prove this inductively since we know that $b_{k+1} = \frac{1}{3}b_k + \frac{1}{3}b_k^2$ which means $\frac{1}{3}b_k \leq b_{k+1} = \left(\frac{1}{3} + b_k\right) b_k$. We can then continue this getting that

$$\left(\frac{1}{3}\right)^2 b_k \leq b_{k+2} = \left(\frac{1}{3} + b_{k+1}\right) \left(\frac{1}{3} + b_k\right) b_k \leq \left(\frac{1}{3} + b_k\right)^2 b_k.$$

This is because b_k is a decreasing sequence. We can continue this to get the originally stated inequality. This gives us

$$\sum_{k=0}^{19} b_k + \sum_{k=20}^{\infty} \left(\frac{1}{3}\right)^{20-k} b_{20} \leq P(\text{balance}) \leq \sum_{k=0}^{19} b_k + \sum_{k=20}^{\infty} \left(\frac{1}{3} + b_{20}\right)^{20-k} b_{20}.$$

The infinite sum on the far left and far right side of this inequality both are simply geometric series so are easily summable. This gives us the stated lower and upper bounds. \square

4.1. Expected Value. We can use a similar technique to calculate the expected rank of balanced vertices.

Theorem 4.6. *The expected value of the rank of balanced vertices exists as n goes to infinity exists.*

Proof. Let $EB(x) = \sum_{k=0}^{\infty} kB_k^*(x)$. This is well defined as a formal power series for the same reason that $B_k^*(x)$ was well defined. To apply Bender's lemma to this we need to find the exponential growth rate of the numerator. Let $eb(n)$ be the coefficient of x^n in the numerator of $EB(x)$. We have that $eb(n) \leq nb^*(n)$ since x^n will not appear in any $B_k^*(x)$ for $k \geq n$ and $b^*(m) = \sum_{k=0}^{\infty} b_k^*(m)$. From this we know that $eb(n) \leq \frac{2.9^n}{n}$, so the numerator of the generating function has exponential growth rate of at most 2.9. Because the exponential growth rate is less than 3 we know the function has radius of convergence greater than $\frac{1}{3}$ which means that Bender's lemma applies. Let $N(x)$ be the numerator of the generating function for $EB(x)$. Asymptotically $EB(x)$ is of the form $N\left(\frac{1}{3}\right)k_n$ where k_n is the x^n term in $\frac{1}{\sqrt{1-2x-3x^2}}$ and since we know the proportion of

k_n to all vertices in all Motzkin trees, this means that the stated proportion exists as n goes to infinity.

Lemma 4.7. *The ratio between the coefficient of x^n in $EB(x)$ and the number of all vertices in all Motzkin trees divided by the proportion of vertices that are balanced is equal to the expected rank of a balanced vertex in a tree of size n .*

Combining Lemma 4.7 with the proof that the asymptotic proportion of $EB(x)$ compared to the number of vertices in all Motzkin trees exists shows us that the expected value exists as n goes to infinity. \square

Theorem 4.8. *The expected rank of a vertex that is balanced is between .6464847301966947 and .64648473019669473*

Proof. We use the standard expected value formula $\mathbb{E}(X) = \sum_{j \geq 0} jP(X = j)$. Getting

upper and lower bounds for this is very similar to the technique we used to find the probability a vertex is balanced. We calculate directly the sum of the first 19 terms and then we use the fact that $\sum_{j \geq 20} jr^{j-20} = \frac{20 - 19r}{(-1 + r)^2}$ with r the common ratio so to get a

lower bound we use $\frac{1}{3}$ and to get an upper bound we get $\frac{1}{3} + b_{20}$. Solving this and dividing by the probability a vertex is balanced gives us the stated expectation. \square

5. OPEN QUESTIONS

We were able to prove that the expected value existed for balanced vertices and calculated fairly accurate bounds for it but the question remains can you show that the expected value exist for general vertices and if so what is the expected value?

The sequence for the number of leaves in all Motzkin trees with $n - 1$ steps is a rather interesting sequence in that it is the number of all paths from $(0,0)$ to (n,n) that avoid 3 right steps in a row that starts with a right step and ends with an up step. This is interesting because there is an easy bijection from the number of Motzkin trees to the number of Dyck paths that avoid 3 right steps. Is there a similar bijection for leaves? If so does it translate easily to rooted planar trees with more than 2 children?

Another question would be what proportion of vertices are leaves, or more generally k -protected, for rooted planar trees where the out degree of every vertex is between 0 and n , and if we take these probabilities as sequences do the sequences converge to the proportion of vertices that are k -protected in all rooted planar trees?

REFERENCES

- [1] Gi-Sang Cheon and Louis W. Shapiro, *Protected points in ordered trees*, Applied Mathematics Letters vol. 21 no. 5 pp. 516 – 520
- [2] Toufik Mansour, *Protected points in k -ary trees*, Applied Mathematics Letters vol. 24 number 4 pp.478 – 480

- [3] Miklós Bóna, *On the number of vertices of each rank in phylogenetic trees and their generalizations*, Discrete Mathematics & Theoretical Computer Science, Vol. 18 no. 3, Combinatorics (April 11, 2016) dmtcs:1431
- [4] Philippe Flajolet and Robert Sedgewick, *Analytic Combinatorics*, Cambridge University Press, Cambridge (2009)
- [5] Luc Devroye and Svante Janson, *Protected nodes and fringe subtrees in some random trees*, Electronic Communications in Probability, **19**(2014), no 6, 1–10
- [6] Rosena R. X. Du and Helmut Prodinger, *Notes on protected nodes in digital search trees*, Appl. Math. Lett., **25** (2012), 1025–1028
- [7] Hosam M. Mahmoud and Mark D. Ward, *Asymptotic properties of protected nodes in random recursive trees*, Journal of Applied Probability, **52**(2015) 290–297

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF FLORIDA, GAINESVILLE, FLORIDA 32601

Current address: Department of Mathematics, University of Florida, Gainesville, Florida 32611

E-mail address: avanduzer@ufl.edu