On random trees obtained from permutation graphs

Hüseyin Acan
Department of Mathematics
Rutgers University
Piscataway, NJ 08854
USA

huseyin.acan@rutgers.edu

Paweł Hitczenko
Department of Mathematics
Drexel University
Philadelphia, PA 19104
USA

phitczen@math.drexel.edu

Abstract

A permutation w gives rise to a graph G_w; the vertices of G_w are the letters in the permutation and the edges of G_w are the inversions of w. We find that the number of trees among permutation graphs with n vertices is 2^{n-2} for $n \geq 2$. We then study T_n, a uniformly random tree from this set of trees. In particular, we study the number of vertices of a given degree in T_n, the maximum degree in T_n, the diameter of T_n, and the domination number of T_n. Denoting the number of degree-k vertices in T_n by D_k, we find that $(D_1, \ldots, D_m)$ converges to a normal distribution for any fixed m as $n \to \infty$. The vertex domination number of T_n is also asymptotically normally distributed as $n \to \infty$. The diameter of T_n shifted by -2 is binomially distributed with parameters $n - 3$ and $1/2$. Finally, we find the asymptotic distribution of the maximum degree in T_n, which is concentrated around $\log_2 n$.

Keywords: permutation graph; permutation tree; indecomposable permutation; diameter; maximum degree; domination number

Mathematics Subject Classifications: 05C80, 05C05, 60C05

1   Introduction

A permutation graph is an undirected graph obtained from a permutation by drawing an edge for each inversion in the permutation. For a permutation $w_1, \ldots, w_n$ of n numbers, a pair (w_a, w_b) is an *inversion*, if $a < b$ and $w_a > w_b$. Thus, formally, given a permutation $w = w_1, \ldots, w_n$ of $[n] := \{1, \ldots, n\}$, the permutation graph G_w is defined to be the (undirected) graph with the vertex set $[n]$ and the edge set $\{(w_a, w_b) : (w_a, w_b) \text{ is an inversion}\}$. It follows from the definition that the number of edges in G_w is the same as the number of inversions of w. Since a permutation is uniquely determined by the set of its inversions, two different permutations yield two different graphs. Consequently, we have $n!$ permutation graphs on the vertex set $[n]$, as opposed to $2^{\binom{n}{2}}$ general graphs. Note that our definition of a permutation graph was given by Even et al. [13] and it is different from the one given by Chartrand and Harary [6].

Permutation graphs form a subclass of perfect graphs and hence various NP-complete problems in general graphs, including the coloring problem, the maximum clique problem, and the maximum independent set problem, have polynomial time solutions in permutation graphs. This aspect of permutation graphs has led to many studies that are computational in nature. Frequently, a problem on permutation graphs can be easily translated to a problem on permutations. For example, a clique in a permutation graph corresponds to a decreasing subsequence in the accompanying permutation and likewise an independent set corresponds to an increasing subsequence. Permutation graphs have also been used in the theory of limits of combinatorial structures to relate graphons (i.e. graph limits) with permutons (limits of permutations). We refer the reader to [16] for more information on these concepts and the relation between them.

A variant of permutation graphs with an additional parameter t was studied by Adin and Roichman [2] and by Keevash et al. [21]. For $t = 0$, Adin and Roichman, and for general t, Keevash et al. found the maximum number

of edges in $\Gamma_{t,\boldsymbol{\pi}}$, where $\boldsymbol{\pi}$ is a permutation of $[n]$ and $\Gamma_{t,\boldsymbol{\pi}}$ is the graph obtained from $\boldsymbol{\pi}$ in the following way. The vertex set of $\Gamma_{t,\boldsymbol{\pi}}$ is $[n]$, and two vertices $i < j$ are adjacent if i and j form an inversion and there are at most t numbers k with the property $i < k < j$ but $\boldsymbol{\pi}^{-1}(i) > \boldsymbol{\pi}^{-1}(k) > \boldsymbol{\pi}^{-1}(j)$. It follows that the number of edges in $\Gamma_{0,\boldsymbol{\pi}}$ is the number of permutations that are covered by $\boldsymbol{\pi}$ in the strong Bruhat order.

A parallel notion for connected permutation graphs is indecomposable permutations. A permutation $\boldsymbol{w} = w_1, \ldots, w_n$ of $[n]$ is called *decomposable* at m if $\{w_1, \ldots, w_m\} = \{1, \ldots, m\}$ for $m < n$. If there is no such m, then $\boldsymbol{w}$ is called *indecomposable*. Koh and Ree [22] showed that a permutation $\boldsymbol{w}$ is indecomposable if and only if the graph $G_{\boldsymbol{w}}$ is connected. This connection makes it possible to transfer various results and relations about indecomposable permutations to connected permutation graphs. For instance, the bijection between indecomposable permutations of $[n]$ and pointed hypermaps of size $n-1$, given by Mendez and Rosenstiehl [26], implies a one–to–one correspondence between connected permutation graphs and pointed hypermaps of size $n - 1$. For more information on indecomposable permutations, we refer the reader to Bóna [4] or Flajolet and Sedgewick [14].

Indecomposable permutations were first studied by Lentin [23] and Comtet [9, 10]. Lentin [23] showed that $f(n)$, the number of indecomposable permutations (hence the number of connected permutation graphs) of length n, satisfies the recurrence relation

$$n! - f(n) = \sum_{i=1}^{n-1}(n-i)!f(i), \quad f(1) := 1,$$

and consequently, $f(n)$ is the coefficient of t^n in the series $1 - \left(\sum_{k \geq 0} k! t^k\right)^{-1}$. Comtet [9] showed that a permutation chosen uniformly at random from $\mathcal{S}_n$, the set of permutations of $[n]$, is indecomposable with probability $1 - 2/n + O(n^{-2})$. Later, Cori et al. [12] considered the random permutation of $[n]$ with a given number m of cycles and showed that the probability of indecomposability increases from 0 to 1 as m decreases from n to 1. In a more recent paper [1], the first author and Pittel studied a uniformly random permutation of $[n]$ with a given number m of inversions and found an evolution of this random permutation, where the evolution starts with the identity permutation (no inversions) and reaches the unique permutation with $\binom{n}{2}$ inversions, each time gaining one inversion. In this evolution, they showed that the probability of indecomposability is monotone increasing in m, and they found a threshold value of m around which the transition from 'being decomposable' to 'being indecomposable' occurs with high probability. Asymptotic sizes of the largest and the smallest components of the corresponding graph were also found when the number of inversions is slightly smaller than the threshold value.

The well-known Cayley's formula states that the number of trees on n vertices is n^{n-2}. In this paper, we first find that the number of trees among permutation graphs on n vertices is 2^{n-2}. Then we study the tree T_n that is chosen uniformly at random from all these 2^{n-2} trees as n tends to infinity. In particular, we study the degree distribution of T_n, the maximum degree in T_n, the diameter of T_n, and the size of a minimum dominating set in T_n. We find that the number of leaves and the diameter are binomially distributed. Denoting by D_i the number of degree-i vertices in T_n, we prove that $(D_i)_{i=1}^m$ is asymptotically jointly normal for any m. Furthermore, we find the asymptotic distribution of the maximum degree in T_n as $n \to \infty$. Finally, we show that the size of a minimum dominating set, $\gamma(T_n)$, is also asymptotically normally distributed with mean $n/3 + O(1)$ and variance $0.26n + O(1)$.

The maximum degree and the diameter of a random tree have been studied extensively for various classes of trees. For example, Moon [25] showed that $\Delta_n \log\log n / \log n$ approaches 1 in probability as n tends to infinity, where Δ_n is the maximum degree of a tree chosen uniformly at random from all the trees on n labeled vertices. For the same random tree, Rényi and Szekeres [27] showed that the diameter is of order $\Theta(\sqrt{n})$ with probability approaching 1. Cooper and Zito [11] showed that the size of a minimum dominating set of a random recursive tree is $dn + o(n)$ with probability approaching 1 as $n \to \infty$, where $d \approx 0.3745$.

In Section 2 we find the number of trees and the number of forests with a given number of trees. In Section 3, we study the shape of T_n, the degree distribution of T_n, the maximum degree in T_n, the diameter of T_n, and the domination number of T_n.

2 Number of Permutation Trees

Let $\mathcal{S}_n$ denote the set of permutations of the set $[n]$. In this section we count permutations $\boldsymbol{w} \in \mathcal{S}_n$ such that $G_{\boldsymbol{w}}$ is a tree or a forest. In this work, a *permutation tree* refers to a tree that is also a permutation graph and a *tree permutation* refers to a permutation whose graph is a tree. Recall that the number of edges in $G_{\boldsymbol{w}}$ is the same as the number of inversions of $\boldsymbol{w}$ and $G_{\boldsymbol{w}}$ is connected if and only if $\boldsymbol{w}$ is indecomposable.

Let $n > 1$ and let $\boldsymbol{w} = w_1, \ldots, w_n \in \mathcal{S}_n$ be a permutation such that $G_{\boldsymbol{w}}$ is a tree. Let $m = m(\boldsymbol{w}) = n - w_n$. We must have $m > 0$ since otherwise $w_n = n$ is an isolated vertex in $G_{\boldsymbol{w}}$. Moreover, we have the following simple observations.

(i) The numbers $n - m + 1, \ldots, n$ appear in increasing order in $\boldsymbol{w}$. Otherwise, we have $w_i > w_j > w_n = n - m$ for some $i < j < n$ and hence these three vertices form a triangle in $G_{\boldsymbol{w}}$. (Similarly, the numbers $1, \ldots, w_1 - 1$ must be in increasing order.)

(ii) There is no number smaller than $n - m$ appearing after $n - m + 2$ in $\boldsymbol{w}$. Otherwise, using the previous observation, we have a 4-cycle on the vertices $n - m + 1$, $n - m + 2$, j, and $n - m$ for some $j < n - m$.

From these observations we get

$$(w_{n-m+1}, \ldots, w_n) = (n - m + 2, \ldots, n, n - m). \tag{2.1}$$

Note that the vertices $n - m + 2, \ldots, n$ are leaves adjacent to $n - m$ in $G_{\boldsymbol{w}}$. The only other vertex adjacent to $n - m$ is $n - m + 1$, which lies in $w_1, \ldots, w_{n-m}$ and replacing $n - m + 1$ with $n - m$ in $w_1, \ldots, w_{n-m}$, we obtain a tree permutation in $\mathcal{S}_{n-m}$.

We now describe a two-case insertion algorithm to produce tree permutations recursively. For $n = 1$, there is a unique permutation, which is a tree permutation. For $n = 2$, the only tree permutation is $2, 1$. Suppose $n \geq 2$ and let $\boldsymbol{w} \in \mathcal{S}_n$ be a tree permutation. We produce two tree permutations by inserting the number $n + 1$ to $\boldsymbol{w}$ in two ways as follows.

Insertion algorithm.

I_1: Insert $n + 1$ between w_{n-1} and w_n in $\boldsymbol{w}$. In this case $n + 1$ becomes a leaf adjacent to w_n. Denoting the new permutation by $\boldsymbol{w}'$, we have
$$\boldsymbol{w}' = w_1, \ldots, w_{n-1}, n + 1, w_n.$$

This operation increases m by 1, that is, $m(\boldsymbol{w}') = m(\boldsymbol{w}) + 1$.

I_2: Substitute n with $n + 1$ in $\boldsymbol{w}$ and put the number n at the end of the new permutation $\boldsymbol{w}''$. Hence, if $n = w_j$ in $\boldsymbol{w}$, then we have
$$\boldsymbol{w}'' = w_1, \ldots, w_{j-1}, n + 1, w_{j+1}, \ldots, w_n, n.$$

In this case we have $m(\boldsymbol{w}'') = 1$. After this operation, $\{n, n + 1\}$ becomes an edge in $G_{\boldsymbol{w}''}$ and vertex n loses its neighbors in $G_{\boldsymbol{w}}$ to vertex $n + 1$ in $G_{\boldsymbol{w}''}$. All other adjacency relations are preserved.

Clearly, for any given $\boldsymbol{w}$, the pair $(\boldsymbol{w}', \boldsymbol{w}'')$ is obtained uniquely. It is also easy to see that given a tree permutation $\boldsymbol{\tau}$ in $\mathcal{S}_{n+1}$, there is a unique tree permutation $\boldsymbol{w}$ in $\mathcal{S}_n$ such that $\boldsymbol{\tau} = \boldsymbol{w}'$ or $\boldsymbol{\tau} = \boldsymbol{w}''$. If $m(\boldsymbol{\tau}) > 1$, then $\boldsymbol{\tau} = \boldsymbol{w}'$, where $\boldsymbol{w}$ is the permutation obtained from $\boldsymbol{\tau}$ simply by removing $n + 1$ (reverse I_1). If $m(\boldsymbol{\tau}) = 1$, then $\boldsymbol{\tau} = \boldsymbol{w}''$, where $\boldsymbol{w}$ is the permutation obtained from $\boldsymbol{\tau}$ by first substituting $n + 1$ with n and then removing the last element τ_{n+1} (reverse I_2). Hence, for $n \geq 3$, there is a bijection between tree permutations of length n and $\{I_1, I_2\}^{n-2}$. Combining this with the fact that there is only one tree permutation of length 1 and one of length 2, we obtain the following result.

Theorem 1. *Let t_n be the number of tree permutations of length n. We have $t_1 = 1$ and $t_n = 2^{n-2}$ for $n \geq 2$.*

2.1 Shape of a tree permutation

A *caterpillar* is a tree such that all the nonleaves lie on a single path, which we call the *central path*. We will show that if $\boldsymbol{w}$ is a tree permutation, then $G_{\boldsymbol{w}}$ is a caterpillar.

Let $n \geq 3$ and let $w_1, \ldots, w_n$ be a tree permutation. Note that n and w_n are adjacent in $G_{\boldsymbol{w}}$. Also, it follows from (2.1) that exactly one of the vertices n and w_n is a leaf. Similarly, 1 and w_1 are adjacent and exactly one of them is a leaf.

Theorem 2. *Let $n \geq 3$. Let $\boldsymbol{\tau}$ be a tree permutation of length n and $G_{\boldsymbol{\tau}}$ be the corresponding tree. Then $G_{\boldsymbol{\tau}}$ is a caterpillar. If $\tau_1 = n$ (respectively $\tau_n = 1$), then the central path consists of the unique vertex n (respectively 1). If $\tau_1 \neq n$ and $\tau_n \neq 1$, then one endpoint of the central path lies in $\{1, \tau_1\}$ and the other one lies in $\{n, \tau_n\}$.*

Proof. The proof is by induction on n. It is easy to verify that the claim holds for $n = 3$. Now suppose $n > 3$ and let $\boldsymbol{\tau} = \tau_1, \ldots, \tau_n$ be a tree permutation of length n. Let $\boldsymbol{w}$ be the unique tree permutation in $\mathcal{S}_{n-1}$ that produces $\boldsymbol{\tau}$ with one of the operations in the insertion algorithm. By induction hypothesis, $G_{\boldsymbol{w}}$ is a caterpillar with the properties given in the statement of the theorem. Let $P_{\boldsymbol{w}}$ denote the central path in $G_{\boldsymbol{w}}$.

First suppose that $\boldsymbol{\tau}$ is produced from $\boldsymbol{w}$ with the first operation. Hence, the graph $G_{\boldsymbol{\tau}}$ is obtained from $G_{\boldsymbol{w}}$ by adding the vertex n and the edge $\{n, w_{n-1}\}$. We have three cases.

(i) $w_{n-1} = 1$. In this case $G_{\boldsymbol{w}}$ is a star with the central vertex 1 and so is $G_{\boldsymbol{\tau}}$.

(ii) $w_1 = n - 1$. In this case, $\boldsymbol{w} = n - 1, 1, \ldots, n - 2$ and $\boldsymbol{\tau} = n - 1, 1, \ldots, n - 3, n, n - 2$. Consequently, the only nonleaves in $G_{\boldsymbol{\tau}}$ are $\tau_1 = n - 1$ and $\tau_n = n - 2$.

(iii) $w_{n-1} \neq 1$ and $w_1 \neq n - 1$. Here $P_{\boldsymbol{w}}$ has at least two vertices and it starts with one of the vertices 1 or w_1. If w_{n-1} is an end vertex of $P_{\boldsymbol{w}}$, then adding the vertex n and the edge $\{n, w_{n-1}\}$ we get another caterpillar with the same central path $P_{\boldsymbol{w}}$. If w_{n-1} is not an end vertex of $P_{\boldsymbol{w}}$, then by the induction hypothesis, $n - 1$ is an end vertex of $P_{\boldsymbol{w}}$ and w_{n-1} is a leaf adjacent to $n - 1$. Adding the vertex n and the edge $\{n, w_{n-1}\}$ to $G_{\boldsymbol{w}}$ we obtain a caterpillar with the central path $P_{\boldsymbol{w}} \cup \{n - 1, n\}$, so that n becomes an end vertex in the new central path.

The analysis when $\boldsymbol{\tau}$ is produced from $\boldsymbol{w}$ with the second operation is very similar and we skip the details. $\square$

2.2 Forest Permutations

Now we turn our attention to *forest permutations*, which are the permutations whose graphs are acyclic. Only induced cycles in permutation graphs are triangles and cycles of length 4. These two cycles correspond to patterns of 321 and 3412, respectively. In other words, forest permutations correspond to those permutations avoiding the patterns 321 and 3412. The sequence enumerating such permutations (along with several other interpretations of these numbers) are given in [29, Sequence A001519]. Namely, denoting the number of length-n permutations avoiding the patterns 321 and 3412 by f_n, we have the following recurrence relation:

$$f_n = 3f_{n-1} - f_{n-2}; \quad f_1 = 1, \quad f_2 = 2.$$

Solving this recurrence relation we find the number of forest permutations

$$f_n = \frac{\sqrt{5} - 1}{2\sqrt{5}} \left(\frac{3 + \sqrt{5}}{2} \right)^n + \frac{\sqrt{5} + 1}{2\sqrt{5}} \left(\frac{3 - \sqrt{5}}{2} \right)^n.$$

Here we find $f(n, m)$, the number of forest permutations with m trees and a total of n vertices.

Let $T(y)$ be the generating function of tree permutations. By Theorem 1 we have

$$T(y) = y + \sum_{n \geq 2} 2^{n-2} y^n = y + \frac{y^2}{1 - 2y}.$$

Note that $f(n,n) = 1$ since the identity permutation is the only inversion-free permutation. Let $\boldsymbol{w}$ be a permutation and $i < j < k$. If $w_i > w_k$, then either $w_i > w_j$ or $w_j > w_k$. In other words, if w_i and w_k are neighbors in $G_{\boldsymbol{w}}$, then w_j has at least one neighbor in $\{w_i, w_k\}$. It follows from this observation that if C is a connected component of $G_{\boldsymbol{w}}$, the vertex set of C is $\{w_a, w_{a+1}, \ldots, w_b\}$ for some integers a and b. In other words, a connected component of $G_{\boldsymbol{w}}$ consists of consecutive terms in the permutation $\boldsymbol{w}$. Thus, we have

$$
\begin{aligned}
f(n,m) &= [y^n] T(y)^m = [y^{n-m}] \left(1 + y(1-2y)^{-1}\right)^m \\
&= [y^{n-m}] \sum_{k \geq 1} \binom{m}{k} y^k (1-2y)^{-k} \\
&= \sum_{k \geq 1} \binom{m}{k} [y^{n-m-k}](1-2y)^{-k} \\
&= \sum_{k=1}^{\min\{m,n-m\}} \binom{m}{k}\binom{n-m-1}{k-1} 2^{n-m-k}.
\end{aligned}
$$

Hence we have the following theorem.

Theorem 3. *The number of permutation forests with n vertices and m trees is given by the formula*

$$
f(n,m) = \begin{cases} 1 & \text{if } n = m, \\ \sum_{k=1}^{\min\{m,n-m\}} \binom{m}{k}\binom{n-m-1}{k-1} 2^{n-m-k} & \text{if } m < n. \end{cases}
$$

3 The permutation graph of a random tree permutation

Let $\mathcal{C}_n$ be the subset of $\mathcal{S}_n$ consisting of tree permutations and let $\mathcal{GC}_n$ denote the set of trees corresponding to these tree permutations, i.e.

$$
\mathcal{GC}_n = \{G_{\boldsymbol{w}} : \boldsymbol{w} \in \mathcal{C}_n\}.
$$

In Section 2, we proved $|\mathcal{C}_n| = |\mathcal{GC}_n| = 2^{n-2}$ for $n \geq 2$. In this section we turn $\mathcal{GC}_n$ into a probability space by equipping it with the uniform probability measure. We denote by T_n a random element of $\mathcal{GC}_n$ and study various graph properties of T_n. In this section we denote by $\boldsymbol{\tau} = \tau_1 \ldots \tau_n$ the random permutation corresponding to T_n, i.e. $G_{\boldsymbol{\tau}} = T_n$.

3.1 The number of leaves

Let L_n denote the number of leaves in T_n. Here we find the distribution of L_n.

Lemma 4. *Let $n \geq 3$. The vertex τ_n is a leaf in T_n with probability $1/2$. Consequently, the vertex n is a leaf with probability $1/2$.*

Proof. Recall from Section 2.1 that exactly one of $\{n, \tau_n\}$ is a leaf in T_n. Further, the vertex τ_n is a leaf in $G_{\boldsymbol{\tau}}$ if and only if $\boldsymbol{\tau}$ is produced from a tree permutation in $\mathcal{C}_{n-1}$ with the second operation in the insertion algorithm. Since $\boldsymbol{\tau}$ is generated by a uniformly random sequence in $\{I_1, I_2\}^{n-2}$, in probabilistic language it means that the last insertion is performed with the second operation with probability $1/2$, which proves the lemma. $\square$

Lemma 5. *Let $n \geq 3$. The number of leaves L_n in T_n is distributed as $2 + \text{Bin}(n-3, 1/2)$.*

Proof. Consider the insertion algorithm applied to $\boldsymbol{w} = w_1, \ldots, w_n$. When the first operation is applied, the vertex $n+1$ becomes a leaf, the degree of w_n increases by 1, and the other degrees do not change. Hence, if w_n is a leaf in $G_{\boldsymbol{w}}$, then the number of leaves stays the same after the operation, and if w_n is not a leaf, then the number of leaves increases by 1.

Similarly, when the second operation is applied, the vertex n becomes a leaf in $G_{w''}$, the degree of $n + 1$ in $G_{w''}$ becomes one more than the degree of n in G_w, and all other degrees stay the same. Hence, if n is a leaf in G_w, then the number of leaves stays the same after the second operation, and if n is not a leaf, then the number of leaves increases by 1.

The lemma now follows easily from Lemma 4 and induction on n. $\qquad\square$

3.2 The diameter

The diameter of a tree T is the length (number of edges) of the longest path in T. In a caterpillar, a longest path starts and ends with leaves and contains all the vertices in the central path. Thus, the diameter of a tree in $\mathcal{GC}_n$ is two more than the length of the central path. Consequently, denoting the diameter of T_n by $\mathrm{diam}(T_n)$, we have

$$\mathrm{diam}(T_n) = n - L_n + 1. \tag{3.1}$$

Combining (3.1) with Lemma 5 gives the next result.

Lemma 6. *For $n \geq 3$, $\mathrm{diam}(T_n)$ is distributed as $2 + \mathrm{Bin}(n - 3, 1/2)$.*

3.3 The highest degree and the number of vertices of a given degree in T_n

In this section we study $D_k = D_k(T_n)$, the number of vertices of degree k in T_n. Note that D_1 denotes the number of leaves, which was separately studied in Section 3.1. We start with some observations on G_w for a tree permutation w.

Let $w = w_1, \ldots, w_n$ be a tree permutation. We say that w_k is a left-to-right maximum if there is no $i \in [k-1]$ such that $w_i > w_k$. Let W_1 be the set of left-to-right maxima and W_0 be the rest of the numbers in w.

Lemma 7. *The graph G_w has the bipartition (W_0, W_1).*

Proof. Clearly, two elements w_i and w_j of W_1 cannot be neighbors since both of them are left-to-right maxima. Now let $i < j$ and suppose $w_i, w_j \in W_0$. Since $w_i \in W_0$, there is some k such that $k < i$ and $w_k > w_i$. Now, we must have $w_i < w_j$ since otherwise w_k, w_i, w_j would form a triangle. Hence $w_i w_j$ is not an edge, which finishes the proof. $\qquad\square$

This lemma says that elements of W_0 as well as elements of W_1 appear in increasing order in w. This fact implies that (i) if $w_k \in W_1$, then the neighbors of w_k lie in the set $\{w_i \in W_0 : i > k\}$ and (ii) if $w_k \in W_0$, then the neighbors of w_k lie in the set $\{w_i \in W_1 : i < k\}$. For $k \in [n]$, let $\deg(w_k)$ and $N(w_k)$ denote the degree of w_k and the set of neighbors of w_k in G_w. To formulate our result about $N(w_k)$, we define the block decomposition $B_1, \ldots, B_{2\ell}$ of w as follows:

 (i) each B_j consists of vertices with consecutive indices,

 (ii) the indices of the vertices in B_j are smaller than those of B_{j+1},

 (iii) $B_{2j-1} \subseteq W_1$ and $B_{2j} \subseteq W_0$ for any integer $j \in [\ell]$.

Note that there must be an even number of blocks since $w_1 \in W_1$ and $w_n \in W_0$. Let $b_j = |B_j|$. For any j, we denote the smallest and largest elements of B_j by f_j and l_j, respectively. These are the elements with the smallest and largest indices in B_j, respectively, as well.

Lemma 8. *For $k \in [n]$, the following hold for $\deg(w_k)$ and $N(w_k)$.*

 (a) *If $w_k \in B_{2i-1}$, and $w_k \neq l_{2i-1}$, then $\deg(w_k) = 1$ and $N(w_k) = \{f_{2i}\}$.*

 (b) *If $w_k \in B_{2i-1}$, $w_k = l_{2i-1}$, and B_{2i} is not the last block, then $\deg(w_k) = b_{2i} + 1$ and $N(w_k) = B_{2i} \cup \{f_{2i+2}\}$.*

 (c) *If $w_k \in B_{2i-1}$, $w_k = l_{2i-1}$, and B_{2i} is the last block, then $\deg(w_k) = b_{2i}$ and $N(w_k) = B_{2i}$.*

 (d) *If $w_k \in B_{2i}$ and $w_k \neq f_{2i}$, then $\deg(w_k) = 1$ and $N(w_k) = \{l_{2i-1}\}$.*

 (e) *If $w_k \in B_{2i}$ for some $i \geq 2$ and $w_k = f_{2i}$, then $\deg(w_k) = b_{2i-1} + 1$ and $N(w_k) = B_{2i-1} \cup \{l_{2i-3}\}$.*

6

(f) If $w_k \in B_2$ and $w_k = f_2$, then $\deg(w_k) = b_1$ and $N(w_k) = B_1$.

In words, this lemma says the following. If w_k is a left-to-right maximum, to find $N(w_k)$, we start reading $\boldsymbol{w}$ from w_{k+1} and keep record of all the non-left-to-right maxima until we see the first left-to-right maximum followed by a non-left-to-right maximum. (This last non-left-to-right maximum following a left-to-right-maximum is also recorded.) These recorded vertices will be the neighbors of w_k. Similarly, if w_k is not a left-to-right maximum, we start reading $\boldsymbol{w}$ backwards from w_{k-1} and keep record of all the left-to-right maxima until we see the first non-left-to-right maximum followed by a left-to-right maximum.

Example 9. Let $\boldsymbol{w} = 2, 5, 1, 3, 6, 7, 11, 4, 8, 9, 10$. Here $W_1 = \{w_1, w_2, w_5, w_6, w_7\}$ and $W_0 = \{w_3, w_4, w_8, w_9, w_{10}, w_{11}\}$. We have $N(w_2) = N(5) = \{w_3, w_4, w_8\}$ and $N(w_8) = N(4) = \{w_2, w_5, w_6, w_7\}$.

Proof of Lemma 8. We prove only the first three parts as the others follow immediately from the first three parts combined with Lemma 7.

(a) If $\deg(w_k) > 1$, then any two neighbors of w_k together with w_k and w_{k+1} form a 4-cycle, which is a contradiction. Hence $\deg(w_k) = 1$ and w_k must be a neighbor of the smallest number appearing after w_k, which is f_{2i}.

(b) In this case w_k is the largest element of $\cup_{j \leq 2i} B_j$, which means that it is a neighbor of each element in B_{2i}. Moreover, if w_k is not larger than any element of $\cup_{j > 2i} B_j$, then there is no edge from $\cup_{j \leq 2i} B_j$ to $\cup_{j > 2i} B_j$, a contradiction. If w_k is greater than both of w_a and w_b for some $w_a, w_b \in \cup_{j > 2i}$, then w_k, w_a, w_b and any element of B_{2i+1} form a 4-cycle, a contradiction. Thus, w_k is greater than exactly one element in $\cup_{j > 2i}$, which is f_{2i+2}.

(c) This is similar to (b) but since B_{2i+2} does not exist in this case, all the neighbors of w_k are in B_{2i}. $\qquad\square$

Corollary 10. *Let* $\boldsymbol{d} = (\deg(w_1), \ldots, \deg(w_n))$. *If there are $2k$ blocks in the block decomposition of $\boldsymbol{w}$, then*

$$\boldsymbol{d} = (\ldots, 1^{b_{2i-1}-1}, b_{2i} + 1 - \mathbb{1}_{(i=k)}, b_{2i-1} + 1 - \mathbb{1}_{(i=1)}, 1^{b_{2i}-1}, \ldots)$$

where $\mathbb{1}_A$ denotes the indicator of A.

Since $w_1 \in W_1$ and $w_n \in W_0$, there are at most 2^{n-2} pairs (W_0, W_1). A pair (W_0, W_1) can be encoded by a vector $(a_1 = 1, a_2, \ldots, a_{n-1}, a_n = 0) \in \{0, 1\}^n$: we have $a_k = 1$ if and only if $w_k \in W_1$.

Lemma 11. *Distinct tree permutations correspond to distinct pairs (W_0, W_1).*

Proof. Consider the insertion algorithm given in Section 2. Let $\boldsymbol{a} = (1, a_2, \ldots, a_{n-1}, 0)$ be the described encoding of the bipartition of a tree permutation $\boldsymbol{w}$ of $[n]$. If we insert $n + 1$ to $\boldsymbol{w}$ via I_1, then the new sequence is updated to $(1, a_2, \ldots, a_{n-1}, 1, 0)$. If we insert $n + 1$ to $\boldsymbol{w}$ via I_2, then the new sequence is updated to $(1, a_1, \ldots, a_{n-1}, 0, 0)$. The only difference between these two updated sequences is on their next-to-last components. Hence, for $k \geq 3$, if k is inserted to the permutation via I_1, we have $a_{k-1} = 1$, otherwise $a_{k-1} = 0$. This finishes the proof of the lemma since a tree permutation is obtained via a unique sequence of I_1's and I_2's. $\qquad\square$

In the proof of the following lemma and later, we mean by a *block of a 0-1 sequence* a maximal run of 0's or 1's in the sequence.

Lemma 12. *Let $\boldsymbol{y}$ denote a random 0-1 sequence of length $n - 2$. Let Y_i denote the number of blocks whose length is equal to i in $\boldsymbol{y}$ and let $Y = \sum Y_i$. Then,*

$$(D_1, D_2, D_3 \ldots) \stackrel{d}{=} (n - Y, Y_1, Y_2 \ldots),$$

where $\stackrel{d}{=}$ means 'equal in distribution'.

Proof. By the previous lemma and the encoding of the pairs (W_0, W_1), there is a canonical bijection between 0-1 sequences of length $n - 2$ and tree permutations of $[n]$. Let $W_0' = W_0 \setminus \{w_n\}$ and $W_1' = W_1 \setminus \{w_1\}$ and consider the block decomposition $B_1', \ldots, B_k'$ of $w_2, \ldots, w_{n-1}$. Unlike the former case (where w_1 and w_n were also considered),

now we may have $B_1' \subseteq W_0'$, $B_k' \subseteq W_1'$, or k odd. It is easy to verify using Lemma 8 that corresponding to each block B_j', there is a unique element w_k of degree $|B_j'| + 1$. All the remaining elements are of degree 1.

Now the lemma follows immediately from coupling the random tree permutation $\boldsymbol{\tau}$ with $\boldsymbol{y}$, i.e. by letting $\boldsymbol{\tau}$ be the permutation represented by $\boldsymbol{y}$. $\qquad\square$

Let H_n denote the highest degree in T_n. An immediate consequence of Lemma 12 is the following.

Corollary 13. *If $\boldsymbol{y}$ and Y_i are as defined in Lemma 12, then*

$$H_n \stackrel{d}{=} 1 + \max\{i : Y_i > 0\}.$$

In other words, $H_n - 1$ is distributed as the size of the largest block in a random 0-1 sequence of length $n - 2$.

A random 0-1 sequence of length n can be generated by first choosing the first element randomly and then by flipping a coin $n - 1$ times. During this flipping process, if a tail (T) comes up, then we put the same symbol as the previous one and hence extend the size of the current block by one, and if a head (H) comes up, we put a different symbol than the previous one and hence start a new block. For instance, for $n = 8$, if the first symbol of the sequence is 0 and the outcome of coin flips is $THHTTTHT$, then we have the sequence 00100011. Note that the number of blocks in the 0-1 sequence obtained this way is one more than the number of heads and the size of the largest block is one more than the longest run of tails. Now let LT_n denote the longest run of tails in a sequence of n coin flips. Combining this fact with Corollary 13, we get

$$H_n \stackrel{d}{=} 2 + LT_{n-3}. \tag{3.2}$$

Földes [15] proved the following result regarding the distribution of LT_n (see also earlier, unpublished work by Boyd [5]).

Theorem 14 (Földes)**.** *For any integer k, we have*

$$\mathbb{P}(LT_n - \lfloor \log_2 n \rfloor < k) = \exp\left(-2^{-k-1+\{\log_2 n\}}\right) + o(1),$$

where $\{x\} := x - \lfloor x \rfloor$ for any positive real number x.

Combining Theorem 14 with (3.2) gives the following corollary.

Corollary 15. *For any integer k, we have*

$$\mathbb{P}(H_n - \lfloor \log_2(n-3) \rfloor < k) = \exp\left(-2^{-k+1+\{\log_2(n-3)\}}\right) + o(1).$$

In view of this discussion and Theorem 12, for the distribution of the degree sequence (D_i), it is enough to study the number of blocks of a given size in a random 0-1 sequence of length $n - 2$ or the runs of tails in flipping a coin $n - 3$ times. Let $\boldsymbol{s} = (s_1, \ldots, s_{n-3})$ denote a random H-T sequence which represents the outcome of $n - 3$ coin flips. Let $\boldsymbol{y} = (y_1, \ldots, y_{n-2})$ denote the random 0-1 sequence corresponding to $\boldsymbol{s}$, where y_1 is chosen randomly and independently of $\boldsymbol{s}$. Let Y_k denote the number of blocks of size k in $\boldsymbol{y}$. Each run of tails of length $k - 1$ corresponds to a block of size k in $\boldsymbol{y}$. Let Y_k^* count the strings of length $k + 1$ equal to $HT^{k-1}H$ in $\boldsymbol{s}$. Denoting the sizes of the first and last blocks in $\boldsymbol{y}$ by b_1 and b_l, respectively , we have

$$Y_k^* = Y_k - \mathbb{1}_{(b_1=k)} - \mathbb{1}_{(b_l=k)} \quad \text{and} \quad \sum_{k \geq 2}(Y_k - Y_k^*) \in \{0, 1, 2\}. \tag{3.3}$$

Letting $\xi_i = \xi_i(k)$ be the indicator of $\{(s_i, \ldots, s_{i+k}) = (HT^{k-1}H)\}$ and using the linearity of the expectation, we get

$$\mathbb{E}Y_k^* = \sum_{i=1}^{n-k-3} \mathbb{E}\xi_i = (n-k-3)2^{-k-1}. \tag{3.4}$$

8

Also, routine calculations yield

$$\text{var}(Y_k^*) = \frac{2^{k+1} + 1 - 2k}{2^{2k+2}} n + O(k2^{-k}). \tag{3.5}$$

Together with Chebyshev's Inequality, these two equations imply that Y_k^* (and hence Y_k) is concentrated around its expected value, which is roughly $n/2^{k+1}$. In fact, we have the joint normality of $(Y_1, \ldots, Y_m)$ and hence of the degrees $(D_1, \ldots, D_m)$ for any fixed m.

Theorem 16. *For any $m \geq 1$, as $n \to \infty$ one has*

$$\frac{1}{\sqrt{n}} (Y_k^* - \mathbb{E}Y_k^*)_{k=1}^m \overset{d}{\to} N(0, \Sigma),$$

where $\Sigma = [\sigma_{i,j}]$ with

$$\sigma_{i,i} = \frac{1}{2^{i+1}} \left(1 - \frac{2i-3}{2^{i+1}}\right), \quad \text{and } \sigma_{i,j} = -\frac{i+j-3}{2^{i+j+2}}, \quad i \neq j, \quad i, j = 1, \ldots, m.$$

Remark 17. Since the joint convergence in $\mathbb{R}^\infty$ is defined through the joint convergence of any finitely many components, the above theorem can be re-stated as the convergence of the infinite–dimensional vector in $\mathbb{R}^\infty$

$$\frac{1}{\sqrt{n}} (Y_k^* - \mathbb{E}Y_k^*)_{k=1}^\infty \overset{d}{\to} N(0, \Sigma),$$

with $\sigma_{i,j}$ given above for all $i, j \geq 1$. Furthermore, if $G = (G_1, G_2, \ldots)$ is a mean–zero Gaussian vector in $\mathbb{R}^\infty$ with covariance matrix Σ and $A : \mathbb{R}^\infty \to \mathbb{R}^\infty$ is an infinite dimensional matrix then AG is Gaussian with the covariance matrix $A\Sigma A^T$. Applying this with matrix

$$A = \begin{bmatrix} -1 & -1 & -1 & -1 & \ldots \\ 1 & 0 & 0 & 0 & \ldots \\ 0 & 1 & 0 & 0 & \ldots \\ 0 & 0 & 1 & 0 & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \end{bmatrix}$$

we obtain

Corollary 18. *As $n \to \infty$*

$$\frac{1}{\sqrt{n}} (D_k - \mathbb{E}D_k)_{k=1}^\infty \overset{d}{\to} N(0, A\Sigma A^T) \quad in \ \mathbb{R}^\infty,$$

where A is as above and Σ is as in Theorem 16. In particular, for any $m \geq 1$

$$\frac{1}{\sqrt{n}} (D_k - \mathbb{E}D_k)_{k=1}^m \overset{d}{\to} N(0, (A\Sigma A^T)_{m \times m}),$$

where $(A\Sigma A^T)_{m \times m}$ is an $m \times m$ northwest corner of $A\Sigma A^T$.

Proof of Theorem 16. We will apply the following Hoeffding–Robbins Central Limit Theorem [20]. There are stronger versions of this theorem, see e.g. [28] and references therein, but the original version of Hoeffding–Robbins is enough for our purpose. Interestingly, while most of the later papers concentrate on 1–dimensional random variables, Hoeffding and Robbins actually give a version for random vectors. To be precise they state and prove the vector–valued version for stationary sequences of 2–dimensional, m–dependent random vectors (see Theorem 3 in [20]) but state after the proof that "The extension of Theorem 3 to the case $N > 2$, as well as to the non–stationary case, is evident and will be left to the reader." We recall that a sequence (X_n) of random variables is m–dependent if for all k and l in $\mathbb{N}$, $(X_1, \ldots, X_k)$ and $(X_{k+n}, \ldots, X_{k+n+l})$ are independent whenever $n > m$, and it is stationary if, for any $j \in \mathbb{N}$, the distribution of random vector $(X_n, X_{n+1}, \ldots, X_{n+j})$ does not depend on n.

Theorem 19. *(Hoeffding–Robbins) Let $(X_{i,1}, X_{i,2}, \ldots, X_{i,N})$, $i = 1, 2, \ldots$ be a stationary and m–dependent sequence of random vectors in $\mathbb{R}^N$ such that*

$$\mathbb{E}X_{1,k} = 0, \quad \mathbb{E}|X_{1,k}|^3 < \infty, \quad k = 1, \ldots, N.$$

Then as $n \to \infty$ the random vector

$$\frac{1}{\sqrt{n}} \left(\sum_{i=1}^{n} X_{i,1}, \ldots, \sum_{i=1}^{n} X_{i,N} \right), \tag{3.6}$$

has a limiting normal distribution with mean zero and covariance matrix $\Sigma = [\sigma_{j,k}]$, where

$$\sigma_{j,k} = \mathbb{E}X_{1,j}X_{1,k} + \sum_{l=1}^{m} \left(\mathbb{E}X_{1,j}X_{l+1,k} + \mathbb{E}X_{l+1,j}X_{1,k} \right) \quad j, k = 1, \ldots, N. \tag{3.7}$$

We wish to apply this theorem to the random vectors

$$(\xi_i(1) - \mathbb{E}\xi_i(1), \xi_i(2) - \mathbb{E}\xi_i(2), \ldots, \xi_i(N) - \mathbb{E}\xi_i(N)), \quad i = 1, 2, \ldots$$

with $N = m$. A minor nuisance is that unless the $\xi_i(k)$ are defined based on the *infinite* sequence of coin tosses they are not stationary (this is because in n tosses, $\xi_i(k)$'s are 0 for $i > n - k$ and thus their distribution is different from that of $\xi_i(k)$ for $1 \le i \le n - k$). We will deal with this issue later and for now we assume that the $\xi_i(k)$ are defined based on an infinite sequence of coin tosses. Then the sequence $(X_{i,1}, \ldots, X_{i,m})$ is stationary. Note also that for every $k \ge 1$, $\xi_i(k)$, $i \ge 1$, are k-dependent because $\xi_i(k)$ involves positions $(s_i, \ldots, s_{i+k})$. Therefore, the random vectors $(\xi_i(1), \ldots, \xi_i(m))$, $i \ge 1$, are m–dependent and hence $(X_{i,1}, \ldots, X_{i,m})$ are m–dependent, too. Since $\xi_i(k)$ are indicator random variables, it is evident that $\mathbb{E}|X_{i,k}|^3 \le 2^3$. Therefore, the asymptotic normality (3.6) holds and it remains to evaluate the covariance matrix (3.7).

Since $\xi_1(j)$ and $\xi_{l+1}(k)$ involve positions $(s_1, \ldots, s_{j+1})$ and $(s_{l+1}, \ldots, s_{l+1+k})$, respectively, they are independent if $l > j$, impossible to happen simultaneously if $l < j$, and correspond to $(s_1, \ldots, s_{k+l+1}) = HT^{k-1}HT^{l-1}H$ if $l = j$. Hence, for $l \ge 1$,

$$\mathbb{E}X_{1,j}X_{l+1,k} = \mathrm{cov}(\xi_1(j), \xi_{l+1}(k)) = \begin{cases} 0, & \text{if } l > j; \\ 2^{-j-k-1} - 2^{-j-k-2} = 2^{-j-k-2}, & \text{if } l = j; \\ -2^{-j-k-2}, & \text{if } l < j. \end{cases}$$

Consequently,

$$\sum_{l=1}^{m} \left(\mathbb{E}X_{1,j}X_{l+1,k} + \mathbb{E}X_{l+1,j}X_{1,k} \right) = -\frac{j-1}{2^{k+j+2}} + \frac{1}{2^{k+j+2}} - \frac{k-1}{2^{k+j+2}} + \frac{1}{2^{k+j+2}} = -\frac{j+k-4}{2^{j+k+2}}.$$

This holds regardless of whether $j = k$ or not. However,

$$\mathbb{E}X_{1,j}X_{1,k} = \mathrm{cov}(\xi_1(j), \xi_1(k)) = \begin{cases} 2^{-j-1} - 2^{-2(j+1)}, & \text{if } k = j; \\ -2^{-j-k-2}, & \text{if } k \ne j. \end{cases}$$

Hence,

$$\sigma_{j,j} = \frac{1}{2^{j+1}} - \frac{2j-3}{2^{2(j+1)}} = \frac{1}{2^{j+1}} \left(1 - \frac{2j-3}{2^{j+1}} \right), \quad \sigma_{j,k} = -\frac{j+k-3}{2^{2(j+1)}}, \quad j \ne k.$$

This proves the central limit theorem in the case of the infinite number of coin tosses. In Theorem 16 we formally have a triangular array of random vectors

$$\left(X_{i,1}^{(n)}, \ldots, X_{i,m}^{(n)} \right) = \left(\xi_i^{(n)}(1) - \mathbb{E}\xi_i^{(n)}(1), \ldots, \xi_i^{(n)}(m) - \mathbb{E}\xi_i^{(n)}(m) \right), \quad i = 1, 2, \ldots, n, \quad n \ge 1,$$

where $\xi_i^{(n)}(k)$ is the indicator of the event $\{(s_i, \ldots, s_{i+k}) = (HT^{k-1}H)\}$ and $(s_1, \ldots, s_n)$ is a sequence of the first n tosses in an infinite sequence of a coin toss. But for $k \leq m$, this does not affect the distribution of $(X_{i,k}^{(n)})$ as long as $i \leq n - m$. Thus

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_{i,k}^{(n)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_{i,k} + \frac{1}{\sqrt{n}} \sum_{i=n-m+1}^{n} (X_{i,k}^{(n)} - X_{i,k}).$$

Since m is fixed and $X_{i,k}^{(n)} - X_{i,k}$ are uniformly bounded it follows that

$$\frac{1}{\sqrt{n}} \sum_{i=n-m+1}^{n} (X_{i,k}^{(n)} - X_{i,k}) \xrightarrow{P} 0, \quad \text{as } n \to \infty$$

and hence Theorem 16 is proved. $\qquad\square$

3.4 Domination number of T_n

A *dominating set* of a graph G is a subset S of the vertex set of G such that each edge of G is incident to a vertex in S. The *domination number* of G, denoted $\gamma(G)$, is the minimum size of a dominating set, i.e.

$$\gamma(G) = \min\{\, |S| : S \text{ is a dominating set of } G\}.$$

Let T be a tree whose vertices are labeled with integers. A smallest dominating set S of T can be found with the following recursive algorithm, similar to the one given by Cockayne, Goodman, and Hedetniemi [8]. As long as a tree contains at least three vertices, first mark (simultaneously) the neighbors of the leaves of the tree and then delete all the edges incident to these marked vertices. Repeat this process as long as there is a tree of size at least 3. At the end, we end up with trees of size 1 and 2. At this point, mark the vertex with smaller value in each tree of size 2. Finally, put the marked vertices into S.

Now suppose that T is a caterpillar, in which case, the dominating set S produced by the algorithm we described above is a subset of the central path. Let $S_1 \subset S$ be the set of marked vertices produced by the first iteration of the algorithm, that is, the set of neighbors of the leaves. Note that S_1 contains all the endpoints of the central path as well as all vertices of degree at least 3, but nothing more. Let $S_1 = \{v_1, \ldots, v_k\}$, where the unique path between v_i and v_{i+1} does not contain any other vertex from S_1 for $1 \leq i \leq k-1$. The remaining vertices (if any) of the central path are of degree-2 and they are scattered between the vertices of S_1. If there are n_i internal vertices of degree-2 on the path between v_i and v_{i+1}, then

$$\gamma(T) = k + \sum_{i=1}^{k-1} \lfloor n_i/2 \rfloor. \tag{3.8}$$

In view of this discussion and the adjacency relation given in Lemma 8, to find an asymptotic distribution of $\gamma(T_n)$, we need to analyze the structure of the block decomposition of $\boldsymbol{\tau} = \tau_1, \ldots, \tau_n$. In particular, S_1 differs by at most 2 from the number of vertices of degree at least 3, which is given by $\sum_{k \geq 2} Y_k$ by Lemma 12. The rest of the smallest dominating set described above consists of degree-2 vertices. For this part, (i.e. degree-2 vertices in the dominating set) we need to analyze the maximal runs of blocks of size 1 in the block decomposition of $\boldsymbol{\tau} = \tau_1, \ldots, \tau_n$. More specifically, letting $B_0 = \emptyset$, it follows from Lemma 8 that for each pair $i \geq 0$ and $j \geq 1$ such that

$$b_i \neq 1, \quad b_{i+1} = \cdots = b_{i+j} = 1, \quad b_{i+j+1} \neq 1, \tag{3.9}$$

where b_t denotes the size of B_t, there is a unique set of j vertices, each of degree 2, that connect two vertices of a longest path in T_n. Necessarily, a longest path in a caterpillar consists of the central path and two leaves.

In fact, all we need is the information of runs of single-vertex blocks in the block decomposition of $\tau_2, \ldots, \tau_{n-1}$. Recalling that $W_0' = W_0 \setminus \{w_n\}$ and $W_1' = W_1 \setminus \{w_1\}$, and using the bijection between the pairs (W_0', W_1') and 0-1 sequences of length $n - 2$, it is enough to study the block decomposition of 0-1 sequences. Again we will couple the random tree permutation $\tau_1, \ldots, \tau_n$ with $\boldsymbol{y}$, a random 0-1 sequence of length $n - 2$ and we will generate $\boldsymbol{y}$ with

the coin-flip algorithm. For the runs of blocks of size one, we now need to analyze runs of heads in the random sequence $\boldsymbol{s}$, where $\boldsymbol{s}$ represents the output of $n-3$ coin flips.

Analogous to Y_k^*, we define Z_k^* as the number of strings equal to $TH^{k+1}T$ in $\boldsymbol{s}$. (Note that this string is of size $k+3$ as opposed to $k+1$ in the case of Y_k^*.) Each such string adds 1 to a run of k-consecutive blocks of size one in the block decomposition of $\boldsymbol{\tau}$. Note that Z_k^* has the same distribution as Y_{k+2}^*.

Theorem 20. *As $n \to \infty$,*

$$\frac{\gamma(T_n) - \mathbb{E}\gamma(T_n)}{\sqrt{n}} \xrightarrow{d} N(0, \sigma^2),$$

where $\mathbb{E}\gamma(T_n) = n/3 + O(1)$ and $\sigma^2 = \frac{13}{50}$.

Proof. It follows from (3.8) that

$$\gamma(T_n) = \sum_{i \geq 2} Y_i + \sum_{k \geq 2} \lfloor k/2 \rfloor Z_k^* + \sum_{k \geq 1} \lceil k/2 \rceil (\mathbb{1}_{A_k} + \mathbb{1}_{B_k}),$$

where A_k and B_k denote the events that $\boldsymbol{s}$ starts with $H^k T$ and ends with TH^k, respectively. (Note that in the last sum we have the ceiling of $k/2$ because if the central path starts (ends) with a degree-2 vertex, this vertex is included in S by the algorithm.) Since $\mathbb{P}(A_k) = \mathbb{P}(B_k) = 1/2^{k+1}$, the last sum above is bounded in probability. Using this and (3.3), for the asymptotic distribution of $\gamma(T_n)$, we may only consider the sum

$$\sum_{k \geq 2} \left(\lfloor k/2 \rfloor Z_k^* + Y_k^* \right).$$

Let

$$V_k := \lfloor k/2 \rfloor Z_k^* + Y_k^* - \mathbb{E}(\lfloor k/2 \rfloor Z_k^* + Y_k^*), \quad k \geq 2.$$

For $x \in \mathbb{R}$, any $\varepsilon > 0$ and m to be chosen later we have

$$\mathbb{P}\Big(\frac{1}{\sqrt{n}} \sum_{k \geq 2} V_k \leq x \Big) \leq \mathbb{P}\Big(\frac{1}{\sqrt{n}} \sum_{k=1}^m V_k \leq x + \varepsilon, \frac{1}{\sqrt{n}} \Big| \sum_{k > m} V_k \Big| \leq \varepsilon \Big) + \mathbb{P}\Big(\frac{1}{\sqrt{n}} \Big| \sum_{k > m} V_k \Big| > \varepsilon \Big).$$

Also,

$$\mathbb{P}\Big(\frac{1}{\sqrt{n}} \sum_{k \geq 2} V_k \leq x \Big) \geq \mathbb{P}\Big(\frac{1}{\sqrt{n}} \sum_{k=1}^m V_k \leq x, \frac{1}{\sqrt{n}} \Big| \sum_{k > m} V_k \Big| \leq \varepsilon \Big) \geq \mathbb{P}\Big(\frac{1}{\sqrt{n}} \sum_{k=1}^m V_k \leq x - \varepsilon, \frac{1}{\sqrt{n}} \Big| \sum_{k > m} V_k \Big| \leq \varepsilon \Big)$$

$$\geq \mathbb{P}\Big(\frac{1}{\sqrt{n}} \Big| \sum_{k \leq m} V_k \Big| \leq x - \varepsilon \Big) - \mathbb{P}\Big(\frac{1}{\sqrt{n}} \Big| \sum_{k > m} V_k \Big| > \varepsilon \Big).$$

Using the inequality $\mathrm{var}(\sum_{k>m} V_k) \leq (\sum_{k>m} \sqrt{\mathrm{var}V_k})^2$ and the fact that $\mathrm{var}(V_k) = O(n/2^k)$ uniformly in k, (see (3.5)), by Chebyshev's inequality we get

$$\mathbb{P}\Big(\frac{1}{\sqrt{n}} \Big| \sum_{k > m} V_k \Big| > \varepsilon \Big) \leq \frac{1}{\varepsilon^2 n} \mathrm{var}\Big(\sum_{k > m} V_k \Big) \leq \frac{Cn}{\varepsilon^2 n 2^{m/2}} = \frac{C}{\varepsilon^2 2^{m/2}},$$

for an absolute constant C. Pick m so that $C/2^{m/2} < \varepsilon^3/2$ and consider

$$\frac{1}{\sqrt{n}} \sum_{k \leq m} V_k.$$

We let $\eta_i(k)$ be the indicator of the event $\{(s_i, \ldots, s_{i+k+2}) = TH^{k+1}T\}$ and write

$$\sum_{k=2}^m (\lfloor k/2 \rfloor Z_k^* + Y_k^*) = \sum_{k=2}^m \sum_{i=1}^n (\lfloor k/2 \rfloor \eta_i(k) + \xi_i(k)) = \sum_{i=1}^n \left(\sum_{k=2}^m (\lfloor k/2 \rfloor \eta_i(k) + \xi_i(k)) \right).$$

12

Just as in the proof of Theorem 16, random variables $\{\sum_{k=2}^{m}\lfloor k/2\rfloor\eta_i(k)+\xi_i(k)\}_{i\geq 1}$ are $(m+2)$–dependent. They are also stationary (if based on an infinite number of coin tosses). So, when centered and normalized by $\sqrt{n}$ they satisfy the CLT. Therefore,

$$\mathbb{P}(N(0,\sigma^2)\leq x-\varepsilon)-\varepsilon\leq\liminf_n\mathbb{P}\Big(\frac{1}{\sqrt{n}}\sum_{k\geq 2}V_k\leq x\Big)\leq\limsup_n\mathbb{P}\Big(\frac{1}{\sqrt{n}}\sum_{k\geq 2}V_k\leq x\Big)\leq\mathbb{P}(N(0,\sigma^2)\leq x+\varepsilon)+\varepsilon.$$

The proof of asymptotic normality is completed by letting $\varepsilon\to 0$ and dealing with the infinite versus finite sequence of tosses issue in the same way as in the proof of Theorem 16. We close by computing the expected value and the variance of $\gamma(T_n)$. By (3.4) we already know that $\mathbb{E}Y_k^*=(n-k-3)/2^{k+1}$ so that

$$\sum_{k=2}^{\infty}\mathbb{E}Y_k^*=\frac{n}{4}+O(1).$$

Since Z_k^* has the same distribution as Y_{k+2}^*, we get

$$\sum_{k\geq 2}\lfloor k/2\rfloor\mathbb{E}Z_k^*=n\sum_{k\geq 1}k\left(\frac{1}{2^{2k+3}}+\frac{1}{2^{2k+4}}\right)+O(1)=\frac{n}{2^3\cdot 4}\sum_{k\geq 1}\frac{k}{4^{k-1}}\left(1+\frac{1}{2}\right)+O(1)=\frac{n}{12}+O(1)$$

and hence $\mathbb{E}\gamma(T_n)=n/3+O(1)$. Furthermore,

$$\mathrm{var}\Big(\sum_{k\geq 2}(\lfloor k/2\rfloor Z_k^*+Y_k^*)\Big)=\mathrm{var}\Big(\sum_{i=1}^{n}\sum_{k\geq 2}(\lfloor k/2\rfloor\eta_i(k)+\xi_i(k))\Big)$$

$$=\sum_{i=1}^{n}\mathrm{var}\Big(\sum_{k\geq 2}(\lfloor k/2\rfloor\eta_i(k)+\xi_i(k))\Big)+2\sum_{i<j}\mathrm{cov}\Big(\sum_{k\geq 2}(\lfloor k/2\rfloor\eta_i(k)+\xi_i(k)),\sum_{k\geq 2}(\lfloor k/2\rfloor\eta_j(k)+\xi_j(k))\Big). \qquad (3.10)$$

Now, for $1\leq i\leq n-k-2$

$$\mathrm{var}\Big(\sum_{k\geq 2}(\lfloor k/2\rfloor\eta_i(k)+\xi_i(k))\Big)=\sum_{k\geq 2}\big\{\lfloor k/2\rfloor^2\mathrm{var}(\eta_i(k))+\mathrm{var}(\xi_i(k))+2\lfloor k/2\rfloor\mathrm{cov}(\eta_i(k),\xi_i(k))\big\}$$

$$+2\sum_{k<l}\big\{\lfloor k/2\rfloor\lfloor l/2\rfloor\mathrm{cov}(\eta_i(k),\eta_i(l))+\mathrm{cov}(\xi_i(k),\xi_i(l))+\lfloor k/2\rfloor\mathrm{cov}(\eta_i(k),\xi_i(l))+\lfloor l/2\rfloor\mathrm{cov}(\xi_i(k),\eta_i(l))\big\}$$

$$=\sum_{k\geq 2}\left\{\frac{\lfloor k/2\rfloor^2}{2^{k+3}}\left(1-\frac{1}{2^{k+3}}\right)+\frac{1}{2^{k+1}}\left(1-\frac{1}{2^{k+1}}\right)-2\frac{\lfloor k/2\rfloor}{2^{k+3}2^{k+1}}\right\}$$

$$-2\sum_{k<l}\left\{\frac{\lfloor k/2\rfloor\lfloor l/2\rfloor}{2^{k+l+6}}+\frac{1}{2^{k+l+2}}+\frac{\lfloor k/2\rfloor+\lfloor l/2\rfloor}{2^{l+3+k+1}}\right\}=\sum_{k\geq 2}\left(\frac{\lfloor k/2\rfloor^2}{2^{k+3}}+\frac{1}{2^{k+1}}\right)-\left(\sum_{k\geq 2}\frac{\lfloor k/2\rfloor}{2^{k+3}}+\frac{1}{2^{k+1}}\right)^2$$

$$=\frac{5}{4\cdot 3^2}+\frac{1}{4}-\frac{1}{3^2}=\frac{5}{18},$$

by the same calculation as in the proof of Theorem 16. To compute the second term in (3.10), we first use

$$\mathrm{cov}\Big(\sum_{k\geq 2}\big(\lfloor k/2\rfloor\eta_i(k)+\xi_i(k)\big),\sum_{k\geq 2}\big(\lfloor k/2\rfloor\eta_j(k)+\xi_j(k)\big)\Big)$$

$$=\sum_{k\geq 2}\Big(\lfloor k/2\rfloor^2\mathrm{cov}(\eta_i(k),\eta_j(k))+\lfloor k/2\rfloor\big(\mathrm{cov}(\eta_i(k),\xi_j(k))+\mathrm{cov}(\xi_i(k),\eta_j(k))\big)+\mathrm{cov}(\xi_i(k),\xi_j(k))\Big)$$

$$+2\sum_{k<l}\Big(\lfloor k/2\rfloor\lfloor l/2\rfloor\mathrm{cov}(\eta_i(k),\eta_j(l))+\lfloor k/2\rfloor\mathrm{cov}(\eta_i(k),\xi_j(l))+\lfloor l/2\rfloor\mathrm{cov}(\xi_i(k),\eta_j(l))+\mathrm{cov}(\xi_i(k),\xi_j(l))\Big)$$

and then note that for $i < j$ and $k \leq l$ we have

$$\text{cov}(\xi_i(k), \xi_j(l)) = \begin{cases} 0, & \text{if } j > i + k; \\ -2^{-k-1}2^{-l-1}, & \text{if } i < j < i + k; \\ 2^{-k-l-1} - 2^{-k-l-2}, & \text{if } j = i + k. \end{cases}$$

$$\text{cov}(\eta_i(k), \eta_j(l)) = \begin{cases} 0, & \text{if } j > i + k + 2; \\ -2^{-k-3}2^{-l-3}, & \text{if } i < j < i + k + 2; \\ 2^{-k-l-5} - 2^{-k-3}2^{-l-3}, & \text{if } j = i + k + 2. \end{cases}$$

$$\text{cov}(\eta_i(k), \xi_j(l)) = \begin{cases} 0, & \text{if } j > i + k + 2; \\ 2^{-k-3-l+1} - 2^{-k-3}2^{-l-1}, & \text{if } j = i + k + 1; \\ -2^{-k-3}2^{-l-1}, & \text{if } i < j < i + k + 1 \text{ or } j = i + k + 2, \end{cases}$$

$$\text{cov}(\xi_i(k), \eta_j(l)) = \begin{cases} 0, & \text{if } j > i + k; \\ 2^{-k+1-l-3} - 2^{-k-1}2^{-l-3}, & \text{if } j = i + k - 1; \\ -2^{-k-1}2^{-l-3}, & \text{if } i < j < i + k - 1 \text{ or } j = i + k. \end{cases}$$

Consequently,

$$2\sum_{i<j}\sum_{k\geq 2}\lfloor k/2\rfloor^2\text{cov}(\eta_i(k), \eta_j(k)) = 2\sum_{k\geq 2}\left(\lfloor k/2\rfloor^2\sum_i\left(-\frac{k+1}{2^{2(k+3)}} + \frac{1}{2^{2(k+3)}}\right)\right)$$

$$\sim -\frac{2n}{2^6}\sum_{k\geq 2}\frac{k\lfloor k/2\rfloor^2}{4^k} = -\frac{n}{2^5}\sum_{m\geq 1}\left(\frac{m^2(2m)}{4^{2m}} + \frac{m^2(2m+1)}{4^{2m+1}}\right) = -\frac{77}{9000}n.$$

Similarly,

$$2\sum_{i<j}\sum_{k\geq 2}\lfloor k/2\rfloor\text{cov}(\eta_i(k), \xi_j(k)) \sim -\frac{n}{2^5}\sum_{m\geq 1}\frac{m(10m-9)}{4^{2m}} = -\frac{7}{1350}n$$

$$2\sum_{i<j}\sum_{k\geq 2}\lfloor k/2\rfloor\text{cov}(\xi_i(k), \eta_j(k)) \sim -\frac{n}{2^5}\sum_{m\geq 1}\frac{m(10m-19)}{4^{2m}} = \frac{23}{1350}n$$

$$2\sum_{i<j}\sum_{k\geq 2}\text{cov}(\xi_i(k), \xi_j(k)) \sim -\frac{n}{2}\sum_{k\geq 2}\frac{k-2}{2^{2k}} = -\frac{n}{72}$$

$$4\sum_{i<j}\sum_{k\leq l}\text{cov}(\xi_i(k), \xi_j(l)) = 4\sum_{k<l}\sum_i\left(-\frac{k-1}{2^{k+l+2}} + \frac{1}{2^{k+l+2}}\right) \sim -4n\sum_{k\geq 2}\frac{k-2}{2^{k+2}}\sum_{l>k}\frac{1}{2^l} = -\frac{n}{36}$$

$$4\sum_{i<j}\sum_{k\leq l}\lfloor k/2\rfloor\lfloor l/2\rfloor\text{cov}(\eta_i(k), \eta_j(l)) \sim -4n\sum_{k\geq 2}\frac{k\lfloor k/2\rfloor}{2^{k+6}}\sum_{l>k}\frac{\lfloor l/2\rfloor}{2^l} = -\frac{43}{1500}n$$

$$4\sum_{i<j}\sum_{k\leq l}\lfloor k/2\rfloor\text{cov}(\eta_i(k), \xi_j(l)) \sim -\frac{n}{4}\sum_{k\geq 2}\frac{\lfloor k/2\rfloor(k-2)}{2^k}\sum_{l>k}\frac{1}{2^l} \sim -\frac{7}{675}n$$

$$4\sum_{i<j}\sum_{k\leq l}\lfloor l/2\rfloor\text{cov}(\xi_i(k), \eta_j(l)) \sim -\frac{n}{4}\sum_{k\geq 2}\frac{k-4}{2^k}\sum_{l>k}\frac{\lfloor l/2\rfloor}{2^l} \sim \frac{161}{2700}n.$$

Combining all of these calculations we finally get

$$\text{var}\left(\sum_{k\geq 2}(\lfloor k/2\rfloor Z_k^* + Y_k^*)\right) \sim n\left(\frac{5}{8} - \frac{77}{9000} - \frac{7}{1350} + \frac{23}{1350} - \frac{1}{72} - \frac{1}{36} - \frac{43}{1500} - \frac{7}{675} + \frac{161}{2700}\right) = \frac{13}{50}n$$

which completes the proof of Theorem 20. $\qquad\square$

4 Concluding Remarks

Runs of patterns in $0-1$ sequences have many other connections. For example, they can be used to construct threshold graphs (see e.g. [7, 24]) or represent random compositions of integers (see [3] or e.g. [18, 19] for probabilistic interpretation). Thus, our results are directly applicable to such situations. In particular, Theorem 16 gives the joint asymptotic normality of multiplicities of part sizes in random compositions of an integer n, as $n \to \infty$. Random compositions are often studied in conjunction with samples $(\Gamma_1, \ldots, \Gamma_n)$ of iid $\mathrm{Geom}(p)$ random variables. For example, Grabner, Knopfmacher, and Prodinger [17] considered runs (of the same values) in such samples and using arguments based on generating functions derived, among other things, the expressions for the expected value, the variance, and the limiting distribution of the number of runs in such samples (see, Propositions 1, Proposition 2, and Theorem 2, respectively in [17]). We would like to mention that these results are also available by probabilistic arguments we used in this paper. Note that the number of runs is given by

$$R_n = 1 + \sum_{j=1}^{n-1} I_j, \quad \text{where} \quad I_j := \mathbb{1}_{\Gamma_{j+1} \neq \Gamma_j}.$$

Since Γ_j's are iid and I_j involves only Γ_j and Γ_{j+1} it follows that $(I_1, \ldots, I_{n-1})$ are identically distributed 1-dependent random variables. Hence, we can recover the results of Grabner, Knopfmacher, and Prodinger:

$$
\begin{aligned}
\mathbb{E}R_n &= 1 + (n-1)\mathbb{P}(\Gamma_2 \neq \Gamma_1) = 1 + (n-1)\frac{2q}{1+q} = \frac{2q}{1+q}n + \frac{1-q}{1+q}, \\
\mathrm{var}(R_n) &= (n-1)\mathrm{var}(I_1) + 2(n-2)\mathrm{cov}(I_1, I_2) \\
&= (n-1)\frac{2q}{1+q}\left(1 - \frac{2q}{1+q}\right) + 2(n-2)\left(\mathbb{E}I_1 I_2 - \left(\frac{2q}{1+q}\right)^2\right) \\
&= (n-1)\frac{2q(1-q)}{(1+q)^2} + 2(n-2)\frac{q(1-q)^3}{(1+q)^2(1-q^3)} \\
&= \frac{2q(1-q)^2(2+q^2)}{(1+q)^2(1-q^3)}n - \frac{2q(1-q^2)(3-q+q^2)}{(1+q)^2(1-q^3)},
\end{aligned}
$$

where in the penultimate step we have used the fact that

$$
\begin{aligned}
\mathbb{E}I_1 I_2 &= \mathbb{P}(\Gamma_1 \neq \Gamma_2 \neq \Gamma_3) = \sum_{j=1}^{\infty} q^{j-1}p(1 - q^{j-1}p)^2 = 1 - 2\frac{p^2}{1-q^2} + \frac{p^3}{1-q^3} \\
&= \frac{q(1-q)(4q^2 + q + 1)}{(1+q)(1-q^3)}.
\end{aligned}
$$

Furthermore, $(R_n - \mathbb{E}R_n)/\sqrt{n}$ is asymptotically normal by a special case of Theorem 19 (or Theorems 1 or 2 in [20]).

Acknowledgement

References

[1] H. Acan and B. Pittel, *On the connected components of a random permutation graph with a given number of edges*, J. Combin. Theory Ser. A **120** (2013), 1947–1975.

[2] R. Adin and Y. Roichman, *On degrees in the Hasse diagram of the strong Bruhat order*, Sém. Lothar. Combin. **53** (2006), B53g.

[3] G. E. Andrews. *The Theory of Partitions*, Addison – Wesley, Reading, MA, 1976.

[4] M. Bóna, *Combinatorics of permutations*, Chapman and Hall/CRC, 2004.

[5] D. W. Boyd, *Losing runs in Bernoulli trials*, unpublished manuscript, 1972.

[6] G. Chartrand and F. Harary, *Planar permutation graphs*, Ann. Inst. H. Poincaré Sect. B (N.S.) **3** (1967), 433–438.

[7] V. Chvátal and P. L. Hammer, *Aggregation of inequalities in integer programming*. In: Studies in Integer Programming (Proc. Worksh. Bonn 1975). Annals of Discrete Mathematics, vol. 1, pp. 145162. North-Holland, Amsterdam (1977).

[8] E. Cockayne, S. Goodman, and S. Hedetniemi, *A linear algorithm for the domination number of a tree*, Inform. Process. Lett. **4** (1975), no. 2, 41–44.

[9] L. Comtet, *Sur les coefficients de l'inverse de la série formelle $\sum n! t^n$*, C. R. Acad. Sci. (Paris) **A275** (1972), 569–572.

[10] L. Comtet, *Advanced combinatorics,the art of finite and infinite expansions*, Reidel Publishing, 1974.

[11] C. Cooper and M. Zito, *An analysis of the size of the minimum dominating sets in random recursive trees, using the Cockayne–Goodman–Hedetniemi algorithm*, Discrete Appl. Math. **157** (2009), no. 9, 2010-2014.

[12] R. Cori, C. Matthieu, and J. M. Robson, *On the number of indecomposable permutations with a given number of cycles*, Electronic J. Comb. **19** (2012), P49.

[13] S. Even, A. Pnueli, and A. Lempel, *Permutation graphs and transitive graphs*, J. Assoc. Comput. Mach. **19** (1972), 400-410.

[14] P. Flajolet and R. Sedgewick, *Analytic combinatorics*, Cambridge University Press, 2009.

[15] A. Földes, *The limit distribution of the length of the longest head-run*, Period. Math. Hungar. **10** (1979), no. 4, 301–310.

[16] R. Glebov, A. Grzesik, T. Klimošová, and D. Král, *Finitely forcible graphons and permutons*, J. Combin. Theory Ser. B **110** (2015), 112-135.

[17] P. Grabner, A. Knopfmacher and H. Prodinger, *Combinatorics of geometrically distributed random variables: Run statistics*, LATIN 2000, Lecture Notes in Computer Science **1776** (2000), 457–462.

[18] P. Hitczenko and G. Louchard. *Distinctness of compositions of an integer: a probabilistic analysis*, Random Structures Algorithms, **19** (2001), 407–437.

[19] P. Hitczenko and C. D. Savage. *On the multiplicity of parts in a random composition of a large integer*, SIAM J. Discrete Math., **18** (2005), 418–435.

[20] W. Hoeffding and H. Robbins, *The central limit theorem for dependent random variables*, Duke Math. J. **15** (1948), 773–780.

[21] P. Keevash, P. Loh, and B. Sudakov, *Bounding the Number of Edges in Permutation Graphs*, Electronic J. Comb. **13** (2006), no. 1, R44.

[22] Y. Koh and S. Ree, *Connected permutation graphs*, Discrete Math. **307** (2007), 2628–2635.

[23] A. Lentin, *Équations dans les monoïdes libres*, Gauthier-Villars, Paris, 1972.

[24] N. V. R. Mahadev and U. N. Peled, *Threshold graphs and related topics*, Annals of Discrete Mathematics, vol. 56. Elsevier Science Publishers (1995).

[25] J.W. Moon, *On the maximum degree in a random tree*, Michigan Math. J. **15** (1968), 429–432.

[26] P. Ossona de Mendez and P. Rosenstiehl, *Transitivity and connectivity of permutations*, Combinatorica, **24** (2004), 487–502.

[27] A. Rényi and G. Szekeres, *On the height of trees*, J. Austral. Math. Soc. **7** (1967), 497–507.

[28] J. P. Romano and M. Wolf, *A more general central limit theorem for m-dependent random variables with unbounded m*, Statist. Probab. Lett. **47** (2000), 115–124.

[29] N. J. E. Sloane, *On–line Encyclopedia of Integer Sequences*, `http://oeis.org`.