The asymptotic number of non-isomorphic rooted trees obtained by rooting a tree*

Xueliang Li, Yiyang Li, Yongtang Shi

Center for Combinatorics and LPMC-TJKLC

Nankai University, Tianjin 300071, China

Email: lxl@nankai.edu.cn, liycldk@mail.nankai.edu.cn, shi@nankai.edu.cn

Abstract

Let $\mathcal{T}_n$ denote the set of trees with n vertices. Suppose that each tree in $\mathcal{T}_n$ is equally likely. We show that the number of non-isomorphic rooted trees obtained by rooting a tree equals $(\mu_r + o(1))n$ for almost every tree of $\mathcal{T}_n$, where μ_r is a constant. As an application, we show that in $\mathcal{T}_n$ the number of any given pattern, which is a fixed small tree with internal vertices specified, is asymptotically normally distributed with mean $\sim \mu_M n$ and variance $\sim \sigma_M n$, where μ_M and σ_M are some constants related to the given pattern. This solves an open question claimed in Kok's thesis.

Keywords: tree; rooted tree; pattern; counting, generating function; limiting distribution; automorphism

AMS Subject Classification (2010): 05C05, 05C30, 05D40, 05A15, 05A16.

1 Introduction

A *pattern* M is a prescribed tree. We say that M *occurs* in a tree T if M is a subtree of T in the sense that the degree of each internal vertex (of degree more than one) of M matches the degree of the corresponding vertex in T, while each external vertex (of degree one) of M matches a vertex of T with an arbitrary degree. Let $\mathcal{T}_n$ denote the set of trees with n vertices. If we use $X_{n,M}(T)$ to denote the number of occurrences of

*Supported by NSFC No.11371205 and 11531011, the "973" program No.2013CB834204, and PC-SIRT.

a given pattern M in $T \in \mathcal{T}_n$, then $X_{n,M}(T)$ is a random variable with probability

$$P(X_{n,M} = k) = \frac{t_{n,k}}{t_n},$$

where $t_{n,k}$ denotes the number of those trees in $\mathcal{T}_n$ that the number of occurrences of the pattern M in each of the trees is k, and $t_n = |\mathcal{T}_n|$.

Moreover, let $\mathcal{R}_n$ denote the set of rooted trees. We can also consider the number of occurrences of a given pattern in $\mathcal{R}_n$. Denote the corresponding random variable by $X_{n,M}(R)$.

The main work of this paper is to show that some random variable Y_n in $\mathcal{T}_n$ (or $\mathcal{R}_n$) satisfies

$$\frac{Y_n - \mathbf{E}(Y_n)}{\sqrt{\mathbf{Var}(Y_n)}} \to_w \mathcal{N}(0,1),$$

where $\mathcal{N}(0,1)$ is the random variable with standard normal distribution and $\to_w$ means weak convergence. We then call this random variable Y_n *asymptotically normal*. Moreover, if

$$\frac{Y_n - \mu n}{\sqrt{\sigma n}} \to_w \mathcal{N}(0,1),$$

then Y_n is asymptotically normal with mean $\sim \mu n$ and variance $\sim \sigma n$. We refer to [10] for details.

In fact, it was shown in [3] that in $\mathcal{R}_n$ the number $X_{n,M}(R)$ of occurrences of any given pattern is asymptotically normal with mean $\sim \mu_M n$ and variance $\sim \sigma_M n$, where μ_M and σ_M are some constants corresponding to the given pattern. But, for the set $\mathcal{T}_n$ there is no such a result on normal distribution. In [9], the authors proved that for any given pattern in $\mathcal{T}_n$ the limiting distribution has a density $(a + bt^2)e^{ct^2}$, where a, b, c are some constants. The mean and variance of the number of occurrences of any given pattern are still asymptotically $\mu_M n$ and $\sigma_M n$ where the constants are the same as in $\mathcal{R}_n$. Clearly, if one shows that $b = 0$, then the distribution is normal. For some special patterns, such as a star pattern (or a node with a given degree) [5], a double-star pattern [11], and a path pattern [10], the corresponding limiting distributions were proved to be normal. For some previous work we refer to Robinson and Schwenk [16]. For more details, we refer to [3, 9, 10, 16]. Moreover, Gittenberger [7], Panagiotou and Sinha [15] considered the case for the growing star pattern (the number of vertices of the star tending to infinity with n), which yields a non-normal limiting distribution. However, Kok claimed in his thesis [10] that for any given pattern it seems much more difficult to demonstrate the normality. In this paper, we will solve this problem from a new point of view which is different from the existing ones. We study the number of non-isomorphic rooted trees obtained by rooting a tree and get that for almost every tree of order n the number of corresponding non-isomorphic rooted trees is $(\mu_r + o(1))n$

(the authors of [2] and [12] computed this number for other sorts of trees). Based on this result and the normal limiting property on the number of occurrences of any given pattern in $\mathcal{R}_n$, we will show that in $\mathcal{T}_n$ the limiting distribution is also normal.

We organize this paper as follows. In Section 2, we will introduce some basic knowledge that will be used in our proofs. In Section 3, we will present the detailed proofs. Our focus is on studying the number of non-isomorphic rooted trees obtained by rooting a tree. Section 4 is devoted to study the limiting distribution for any given pattern.

2 Preliminaries

Let T_n be a tree in $\mathcal{T}_n$. We say that two vertices u and v of T_n are in the same vertex class if u can be mapped to v by an automorphism of T_n. Clearly, this establishes an equivalent relation on the vertex set of T_n, and hence the vertices in T_n are partitioned into some classes. If u and v are in the same vertex class of T_n, then rooting T_n at u and v, respectively, yields the same rooted tree.

Hence, the number of non-isomorphic rooted trees obtained by rooting a tree is exactly the number of vertex classes of the tree. Let $X_n(T)$ represent the number of vertex classes of the tree T. Clearly, $X_n(T)$ is also a random variable on $\mathcal{T}_n$. Therefore, we can similarly introduce the random variable $X_n(R)$ of vertex classes in the space of rooted trees $\mathcal{R}_n$.

Notice that $X_n(T) \geq 1$. In analogy to the enumeration of patterns in [9], we introduce the following two functions:

$$t(x) = \sum_{n\geq 1} t_n x^n,$$

$$t(x,u) = \sum_{n\geq 1, k\geq 1} t_{n,k} x^n u^k,$$

where the coefficient $t_{n,k}$ denotes the number of trees each having k vertex classes. Clearly, $\sum_{k\geq 1} t_{n,k} = t_n$. We always assume that every tree of $\mathcal{T}_n$ is equally likely. Then, $P(X_n(T) = k) = \frac{t_{n,k}}{t_n}$.

If we consider $X_n(R)$ in $\mathcal{R}_n$, we also suppose that each tree in $\mathcal{R}_n$ is equally likely. We can define similar generating functions on $\mathcal{R}_n$, and let $r(x)$, $r(x,u)$ be the related functions, respectively. One can see that $r(x,1) = r(x)$. Suppose

$$r(x,u) = \sum_{n\geq 1, k\geq 1} r_{n,k} x^n u^k,$$

where $r_{n,k}$ is the number of rooted trees in $\mathcal{R}_n$ each having k vertex classes. It follows that in $\mathcal{R}_n$

$$P(X_n = k) = \frac{r_{n,k}}{r_n},$$

where $r_n = |\mathcal{R}_n|$.

We should notice that when we count the number of vertex classes of a rooted tree, the root itself always forms a class with a single vertex, since any automorphism on a rooted tree must map the root to itself. That is a bit different from the case for non-rooted trees.

Furthermore, suppose that the radius of convergence of $r(x)$ is x_0. Otter [13] showed that x_0 satisfies that $r(x_0) = 1$ and the asymptotic expansion of $r(x)$ is

$$r(x) = 1 - b_1(x_0 - x)^{1/2} + b_2(x_0 - x) + b_3(x_0 - x)^{3/2} + \cdots , \tag{1}$$

where $x_0 \approx 0.3383219$ and $b_1 \approx 2.6811266$. And $t(x)$ has a similar expansion, namely,

$$t(x) = c_0 + c_1(x_0 - x) + c_2(x_0 - x)^{3/2} + \cdots . \tag{2}$$

Applying the transfer theorems in [6] to Eqs.(1) and (2), we get that

$$t_n \sim \frac{Cx_0^{-n}}{n^{5/2}},$$
$$r_n \sim \frac{Dx_0^{-n}}{n^{3/2}},$$

where C and D are some constants. For this, we refer to [14, 16]. It was shown that $C = 0.5349\ldots$ and $D = 0.4399\ldots$. The book [6] gives us more details on the transfer theorems.

In what follows, we first investigate X_n in $\mathcal{R}_n$. To start with, we need the following two lemmas. We refer to [4, 10] for detailed information.

Lemma 1. *Suppose that $F(x, y, u)$ is an analytic function around $(x_0, y_0, 1)$ such that $F(x_0, y_0, 1) = y_0$, $F_y(x_0, y_0, 1) = 1$, $F_{yy}(x_0, y_0, 1) \neq 0$ and $F_x(x_0, y_0, 1) \neq 0$. Then there exist a neighborhood U_0 of $(x_0, 1)$, a neighborhood U_1 of y_0 and analytic functions $g(x, u)$, $h(x, u)$ and $f(u)$ which are defined on U_0, such that the only solutions $y \in U_1$ with $y = F(x, y, u)$ and $(x, u) \in U_0$ are given by*

$$y(x, u) = g(x, u) \pm h(x, u)\sqrt{1 - \frac{x}{f(u)}}. \tag{3}$$

Furthermore, $g(x_0, 1) = y_0$ and $h(x_0, 1) = \sqrt{\frac{2f(1)F_x(x_0,y_0,1)}{F_{yy}(x_0,y_0,1)}}$. If u is real, then $f(u)$ is the radius of convergence of the power series by fixed u in $y(x, u)$. □

We refer the reader to [4, 16] for more details. In [4], the authors always assume that $F(x, y, u)$ has non-negative Taylor coefficients, which is not a necessary requirement. The above result can be found in [16] where $F(x, y, u)$ is required to be analytic around $(x_0, y_0, 1)$. The proof of it ultimately relies on Weierstrass preparation theorem. In Flajolet-Sedgewick's book [6], there is much information on the singular expansion of a function as Eq.(3).

Furthermore, the following claim will be used in the sequel.

Remark 1. *Here, we point out that if u is real and sufficiently close to 1, then $F_y(f(u), y(f(u), u), u) = 1$. This is because if $F_y(f(u), y(f(u), u), u) \neq 1$, y can be analytically continued around $(f(u), u)$ by implicit function theorem, which contradicts that $x = f(u)$ is a singular point.*

Lemma 2. *Let $y(x, u)$ denote a function defined on a neighborhood U of $(x_0, 1)$, $y(x, u) = \sum y_{n,k} x^n u^k$, where $y_{n,k} \geq 0$ and $y(x, u) = F(x, y(x, u), u) = g(x, u) + h(x, u)\sqrt{1 - \frac{x}{f(u)}}$, where $f(1) = x_0$. Moreover, $f(u)$ satisfies the property as in Lemma 1. If $y(x, 1)$ is aperiodic (a power series $\tilde{y}(x, u)$ is called aperiodic if it satisfies that from $y(x, 1) = x^r \tilde{y}(x^d, 1)$ follows $d = 1$) and if $|F_y(x, y, u)| < 1$ for $|x| \leq f(u)$ and $x \neq f(u)$ where u is real and positive around 1, then there exists an $\eta > 0$ such that $y(x, u)$ can be analytically continued in*

$$\widetilde{U} = \{(x, u) : |x| < x_0 + \eta, |u| < 1 + \eta, \arg(x - f(u)) \neq 0, x \neq f(u)\}.$$

Moreover, let $y(x, u) = \sum y_n(u) x^n$, then

$$y_n(u) = \frac{h(f(u), u)}{2\sqrt{\pi} n^{3/2}} f(u)^{-n} + O(\frac{f(u)^{-n-1}}{n^{5/2}}).$$

And if $h(f(1), 1) \neq 0$, then the corresponding random variable X_n determined by $y(x, u)$ (like $r(x, u)$ or $t(x, u)$) is asymptotically normal with mean $\sim \mu n$ and variance $\sim \sigma n$. □

Remark 2. *In [3, 4, 10], the authors always assumed that all the Taylor coefficients of $F(x, y, u)$ are non-negative. But from the proof procedure in [4], we find that $|F_y(x, y, u)| < 1$ is sufficient to get Lemma 2. Because this condition is an initial must when using implicit function theorem. And we can absolutely follow the entire proof in [4] to illustrate this lemma. Hence, we do not repeat the procedure here and refer the reader to the papers [3, 4].*

3 The number of non-isomorphic rooted trees obtained by rooting a tree

Now we concentrate on the number of vertex classes of a rooted tree. Recall that an automorphism of a rooted tree must map the root to itself, which is a bit different from an automorphism of a non-rooted tree. We shall show that $X_n(R)$ is asymptotically normal with mean $(\mu_r + o(1))n$ and variance $(\sigma_r + o(1))n$ in $\mathcal{R}_n$.

In what follows, there appears an expression of the form $Z^*(S_n; f(x,u))$ (or $Z(S_n; f(x))$), which is the substitution of the counting series $f(x,u)$ (or $f(x)$) into the cycle index $Z(S_n)$ of the symmetric group S_n. This involves replacing each variable s_i in $Z(S_n)$ by $f(x^i,u)$ (or $f(x^i)$); see [6]. For instance, if $n=3$, then $Z(S_3) = (1/3!)(s_1^3 + 3s_1s_2 + 2s_3)$ and $Z(S_3; f(x)) = (1/3!)(f(x)^3 + 3f(x)f(x^2) + 2f(x^3))$, $Z^*(S_3; f(x,u)) = (1/3!)(f(x,u)^3 + 3f(x,u)f(x^2,u) + 2f(x^3,u))$. We refer to [8] for details, where it was shown that

$$r(x) = x \cdot \sum_{n\geq 0} Z(S_n; r(x)) = x \cdot \mathrm{e}^{\sum_{k\geq 1} \frac{r(x^k)}{k}}. \tag{4}$$

The coefficient of x^p in $Z(S_n; r(x))$ is the number of rooted trees of order $p+1$ whose roots have degree n. Multiplication of $Z(S_n; r(x))$ by x corrects the power of x so that x^p in $xZ(S_n; r(x))$ is the number of those trees with p vertices. This expression $Z(S_n; r(x))$ follows from the Pólya Enumeration Theorem; see [8]. For the case of double variables, we refer the reader to [3, 5] on the enumeration of patterns.

Analogously, we take the same procedure for $r(x,u)$ in this paper. But, here we should notice that if the same two copies of a rooted tree with k vertex classes connect to a root, then the number of vertex classes of the new rooted tree is $k+1$, because there is only one new class, i.e., the new root, which is different from the procedure for calculating the number of occurrences of a star patten [5]. In other words, the number of vertex classes is not an additive parameter any more. Hence, we must use $r(x^k,u)$ to denote the generating function for rooted trees with k copies of a branch. Moreover, when we apply Pólya Enumeration Theorem to get the expression of $r(x,u)$ as the form of Eq.4, we need to consider the non-additive property further.

For instance, suppose that the tree has a root of degree 2. Then, $xu \cdot Z^*(S_2, r(x,u)) = xu \cdot \frac{1}{2}(r(x,u)^2 + r(x^2,u))$. We have $r_{n,k}$ choices to form a rooted tree with the same two branches. In $r(x,u)^2$, suppose the coefficient of $x^{2n}u^{2k}$ is $r_{n,k,2}$. We should notice that if the two branches of the rooted tree are the same, the $r_{n,k}$ trees counted by error since the number of vertex classes is k rather than $2k$. We should modify it into $(r_{n,k,2} - r_{n,k})x^{2n}u^{2k} + r_{n,k}x^{2n}u^k$. Hence, the generating function of trees with root degree 2 is $xu(Z^*(S_2, r(x,u)) - \frac{1}{2}r(x^2,u^2) + \frac{1}{2}r(x^2,u))$. Notice that the modification only happens to the power of u which represents the number of vertex classes.

In general case, we need to modify $xu \cdot Z^*(S_n, r(x,u))$ to get the correct expression of the generating function of trees with root degree n. The n branches of the rooted tree with root degree n are classified by automorphisms into k_i classes each having i branches, $1 \le i \le n$. By the non-additive property of vertex class, similar to the case of $n = 2$ (i.e., trees with root degree 2), the power of u in the term $\frac{n!}{1^{k_1}2^{k_2}\cdots n^{k_n}k_1!\cdots k_n!} r(x,u)^{k_1}\cdots r(x^n,u)^{k_n}$ must be modified. For instance, if there is a branch class of length $s+2$ ($s \le k_1$), the enumerating series must be $\cdots r(x^{s+2},u)\cdots$. But in $\frac{n!}{1^{k_1}2^{k_2}\cdots n^{k_n}k_1!\cdots k_n!} r(x,u)^{k_1}\cdots r(x^n,u)^{k_n}$, the coefficient contains the case that s classes of length 1 (and $\binom{k_1}{s}$ ways to choose the s branches) and one classes of length 2 are the same, but the corresponding series becomes $\cdots r(x^{s+2},u^2)\cdots$, where the vertex classes are counted twice (in the s branch classes with length 1, the vertex classes are counted once, but in the other class with two branches they are counted again). Hence, the correction term is $-\cdots r(x^{s+2},u^2)\cdots + \cdots r(x^{s+2},u)\cdots$. Clearly, if $u = 1$, the correction term equals 0. For general cases we can deal with them similarly, for example, the correction to the case having $s-1$ branch classes of length 1 and one branch class of length 3, or at least three kinds of branch classes with different length. Then, all the correction terms can be concluded into the form

$$\frac{n!}{1^{k_1}2^{k_2}\cdots n^{k_n}k_1!\cdots k_n!}\sum_{s=0}^{k_1}\binom{k_1}{s} r(x,u)^{k_1-s}(\phi_{n,s}(x,u)-\varphi_{n,s}(x,u)),$$

where $\phi_{n,s}(x,u)$, $\varphi_{n,s}(x,u)$ are polynomial functions of $r(x^l,u^m)$, $2 \le l \le n$, $1 \le m \le n$, and $\phi_{n,s}(x,1) - \varphi_{n,s}(x,1) = 0$. In what follows, we will find that we do not need to know the exact forms of $\phi_{n,s}(x,u)$ and $\varphi_{n,s}(x,u)$.

Therefore, the generating function of $r(x,u)$ is as follows:

$$\begin{aligned} r(x,u) &= xu \cdot \Big(\sum_{n\ge 0} Z^*(S_n, r(x,u)) \\ &\quad + xu \cdot \sum_{1k_1+\cdots+nk_n=n,\, n\ge 0} \frac{1}{n!}\Big(\frac{n!}{1^{k_1}2^{k_2}\cdots n^{k_n}k_1!\cdots k_n!}\sum_{s=0}^{k_1}\binom{k_1}{s} r(x,u)^{k_1-s}(\phi_{n,s}(x,u)-\varphi_{n,s}(x,u))\Big) \qquad (5) \\ &= xu \cdot \mathrm{e}^{\sum_{k\ge 1}\frac{1}{k}r(x^k,u)} \\ &\quad + xu \cdot \sum_{1k_1+\cdots+nk_n=n,\, n\ge 0} \frac{1}{n!}\Big(\frac{n!}{1^{k_1}2^{k_2}\cdots n^{k_n}k_1!\cdots k_n!}\sum_{s=0}^{k_1}\binom{k_1}{s} r(x,u)^{k_1-s}(\phi_{n,s}(x,u)-\varphi_{n,s}(x,u))\Big), \end{aligned}$$

Let $y = y(x,u) = r(x,u)$, and $y = F(x,y,u)$. Here, we notice that the Taylor coefficients of x, y and u need not be non-negative any more. However by Remark 2, Lemma 1 and Lemma 2 can still be applied.

Recall that $r(x,1)=r(x)$ and there exists a real number x_0 such that $r(x_0)=1$, i.e., $y(x_0,1)=0$. Frequently, it is easy to see that $F(x_0,y_0,1)=1$ and $F_x(x_0,y_0,1)\neq 0$. In order to use Lemma 1, we must verify the conditions on the derivative of y. From Eq. (5), it follows that

$$\begin{aligned}
F_y(x,y,u) &= xu\cdot \ \mathrm{e}^{y+\sum_{k\geq 2}\frac{1}{k}r(x^k,u)}\\
&+ xu\cdot \sum_{1k_1+\cdots+nk_n=n,n\geq 0}\frac{1}{n!}\Big(\frac{n!}{1^{k_1}2^{k_2}\cdots n^{k_n}k_1!\cdots k_n!}\sum_{s=0}^{k_1}\binom{k_1}{s}(k_1-s)y^{k_1-s-1}(\phi_{n,s}(x,u)-\varphi_{n,s}(x,u))\Big)
\end{aligned} \tag{6}$$

$$\begin{aligned}
&= xu\cdot \ \mathrm{e}^{y+\sum_{k\geq 2}\frac{1}{k}r(x^k,u)}\\
&+ xu\cdot \sum_{1k_1-1+\cdots+nk_n=n-1,n\geq 1}\frac{1}{n!}\Big(\frac{n!}{1^{k_1}2^{k_2}\cdots n^{k_n}(k_1-1)!\cdots k_n!}\sum_{s=0}^{k_1}\binom{k_1-1}{s}y^{k_1-s-1}(\phi_{n,s}(x,u)-\varphi_{n,s}(x,u))\Big)\\
&= F(x,y(x,u),u)\\
&= F(x,r(x,u),u)\\
&= r(x,u)\\
&= \sum_{n\geq 1,k\geq 1} r_{n,k}x^n u^k.
\end{aligned}$$

Then, we have $F_y(x_0,y(x_0,1),1)=r(x_0,1)=1$ and $F_{yy}(x_0,y(x_0,1),1)\neq 0$ which implies that all the conditions in Lemma 1 hold. That is, for the generating function $y=F(x,y,u)$, there exist a neighborhood U_0 of $(x_0,1)$, a neighborhood U_1 of y_0 and analytic functions $g(x,u)$, $h(x,u)$ and $f(u)$ which are defined on U_0, such that the only solutions $y\in U_1$ with $y=F(x,y,u)$ and $(x,u)\in U_0$ are given by $y(x,u)=g(x,u)+h(x,u)\sqrt{1-\frac{x}{f(u)}}$. Clearly, this expression is coincident with Eq.(1) if we set $u=1$.

Moreover, by the definition of $r(x,u)$, we know $r_{n,k}\geq 0$. By Remark 1, for $|x|\leq f(u)$ and $x\neq f(u)$ where u is real and sufficiently close to 1,

$$\begin{aligned}
|F_y(x,y,u)| = |F(x,y,u)| = |F(x,r(x,u),u)| = |r(x,u)| &= |\sum_{n\geq 1,k\geq 1} r_{n,k}x^n u^k|\\
&< \sum_{n\geq 1,k\geq 1} r_{n,k}|x|^n|u|^k\\
&\leq \sum_{n\geq 1,k\geq 1} r_{n,k}f(u)^n u^k\\
&= F_y(f(u),y(f(u),u),u)\\
&= 1.
\end{aligned}$$

Thus, by Lemma 2, we have that the random variable $X_n(R)$ is asymptotically normal with mean

$$\mathbf{E}(X_n(R)) \sim \mu_r n \ (n \to \infty)$$

and variance

$$\mathbf{Var}(X_n(R)) \sim \sigma_r n \ (n \to \infty),$$

where μ_r and σ_r are some constants. Here, we just concentrate on the rooted trees. Some researchers considered the number of vertex classes of other kind of trees, for example phylogenetic trees [2], and the conclusion also points to an asymptotically normal distribution.

In this paper, we mainly focus on the overall property of a probability space. Following the book [1], we will say that *almost every* (a.e.) graph in a graph space $\mathcal{G}_n$ has a certain property Q if the probability $P(Q)$ in $\mathcal{G}_n$ converges to 1 as n tends to infinity. Occasionally, we will say *almost all* instead of almost every.

From Chebyshev inequality

$$P\left[\left|X_n - \mathbf{E}(X_n)\right| > n^{3/4}\right] \le \frac{\mathbf{Var}X_n}{n^{3/2}} \to 0 \text{ as } n \to \infty,$$

it follows that for almost all rooted trees, $\mathbf{E}(X_n) - n^{3/4} \le X_n \le \mathbf{E}(X_n) + n^{3/4}$, namely, $X_n = (1 + o(1))\mathbf{E}(X_n)$. We can get the following conclusion.

Theorem 3. *For almost all rooted trees in $\mathcal{R}_n$, the number of vertex classes under automorphisms is $(\mu_r + o(1))n$.* □

Therefore, we can study the number of vertex classes in a tree. To get the final result, we need another property as follows. We have defined the number of vertex classes of a tree. We call a vertex *fixed* if this single vertex forms a class.

Lemma 4. *Almost every tree in $\mathcal{T}_n$ has more than $\lfloor \frac{1}{24}n \rfloor$ fixed vertices.*

Proof. We prove this result by contradiction. Suppose that $\mathcal{T}'_n$ is a subset of $\mathcal{T}_n$ such that every tree T'_n in $\mathcal{T}'_n$ has at most $\lfloor \frac{1}{24}n \rfloor$ fixed vertices. We first show that these fixed vertices form a subtree in T'_n. In fact, for any two fixed vertices v_1 and v_2, they can only map to v_1 and v_2 among themselves, respectively. Thus, any (v_1, v_2)-path maps to the (v_1, v_2)-path under any automorphism. So, all vertices in the (v_1, v_2)-path are fixed ones, that is, all the fixed vertices form a connected subgraph of T'_n. Consequently, the fixed vertices induce a subtree T''_n of T'_n and $|T''_n| \le \lfloor \frac{1}{24}n \rfloor$.

Case 1: If $|T''_n| = 0$, then the structure of T'_n is determined by one half of the vertices in T'_n. Hence, the number of trees in $\mathcal{T}_n$ having a symmetrical edge is at most $|\mathcal{T}_{\frac{n}{2}}|$, and $\frac{|\mathcal{T}_{\frac{n}{2}}|}{|\mathcal{T}_n|} \to 0$, which completes the proof.

Case 2: We suppose $|T''_n| > 0$. Let u be a vertex in T''_n. Suppose that H_u is a subtree of T'_n attached to u such that all the vertices in H_u are not in T''_n. Suppose there are m copies of H_u after deleting u. We have $m \geq 2$; otherwise the vertex in H_u connecting to u is also a fixed vertex, a contradiction. If m is even, we get rid of $m/2$ copies of H_u, and if m is odd, we get rid of $(m+1)/2$ copies of H_u. We repeat this operation on all vertices in T''_n. At the end, this produces a new tree A with at most $\lfloor \frac{1}{2}(n + 1/24 \cdot n) \rfloor$ vertices, and we denote the set of these new trees by $\mathcal{A}_{\frac{25}{48}n}$. Moreover, we replace these $\lfloor \frac{m}{2} \rfloor$ copies of H_u by a vertex, that is, we add some vertices to u and different copies of H_u correspond to different vertices. Thus, we constructed another tree A'. An example of this is shown in Figure 1. Observe that T''_n is a subtree of A'. We shall show that A' has at most $\lfloor n/3 \rfloor$ vertices. Color the vertices in A' corresponding to T''_n black and the others gray. We already know that there are at most $\lfloor \frac{1}{24} n \rfloor$ black vertices. Let u be a black vertex. Recall that a gray vertex being neighbor of u in A' corresponds to a set of subtrees in A. We claim that if the set contains only one subtree with a single vertex of A, then for each u this kind of gray vertex must be unique. Otherwise, if for a black vertex u there are two different such kinds of gray vertices v'_1, v'_2 in A', then these two vertices correspond to two leaves in A, and so they should be contained in a same set of subtrees, and hence the two leaves will be replaced by one gray vertex by the construction of A' from A, which contradicts the assumption that v'_1 and v'_2 are two different gray vertices. Therefore, in A' there are at most $\lfloor \frac{1}{24} n \rfloor$ gray vertices such that each of them corresponds to a set containing a single vertex of A. Moreover, there are at most $\lfloor \frac{1}{2} \cdot \frac{1}{2} n \rfloor$ gray vertices such that each of them corresponds to a set containing at least two single vertices or subtrees (or forests) having at least two vertices. Consequently, we get that A' has at most $\lfloor n/3 \rfloor$ vertices.

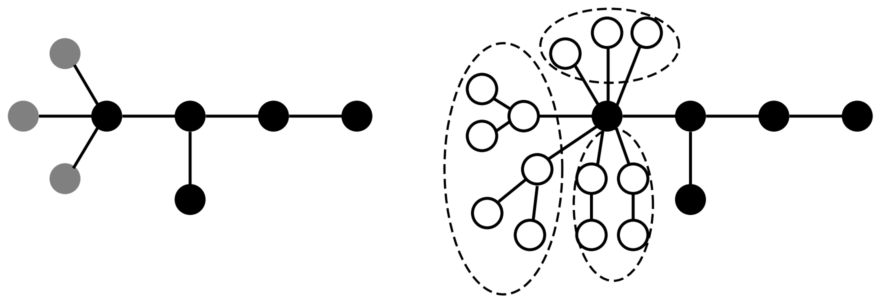

Figure 1: An example of A' and A.

In the above, we have built a map from $\mathcal{T}'_n$ to $\mathcal{A}_{\frac{25}{48}n}$. Suppose A is a tree in $\mathcal{A}_{\frac{25}{48}n}$. Then $|A| \leq \lfloor \frac{25}{48} n \rfloor$. There are at most $\binom{\lfloor \frac{25}{48} n \rfloor}{k}$ ways of choosing k vertices to form a subtree of A. We color these vertices in A by black. Notice that any tree in $\mathcal{T}''_n$ has at most $\lfloor \frac{1}{24}n \rfloor$ vertices. Then, the number of all subtrees of A with order k is less than $\binom{\lfloor \frac{25}{48}n \rfloor}{k} \leq \binom{\lfloor \frac{25}{48}n \rfloor}{\lfloor \frac{1}{24}n \rfloor}$.

We select one subtree T''_n, and color the vertices by black. Suppose that A' is the corresponding tree defined as above. For $u \in T''_n$, each gray vertex in A' connecting to u corresponds to a kind of subtree H_u. Moreover, the number of H_u can be odd or even in T'_n. From the structure of A, we reconstruct the tree T'_n from A by deciding the number of H_u to be odd or even. Since the number of gray vertices is less than $|V(A')| \leq \lfloor n/3 \rfloor$, we can get that there exist at most $2^{\lfloor \frac{n}{3} \rfloor}$ different T'_n's mapped to the same A.

Therefore, for trees in $\mathcal{T}''_n$ with k vertices, at most $2^{\lfloor \frac{n}{3} \rfloor} \cdot 2C \cdot x_0^{-\frac{25}{48}n} \cdot \binom{\lfloor \frac{25}{48}n \rfloor}{k}$ trees in $\mathcal{T}'_n$ map to them. Recall that each T'_n corresponds to some T''_n. Moreover, we need the fact that the order of $\mathcal{T}_n$ is asymptotically $\frac{C \cdot x_0^{-n}}{n^{5/2}}$. So, the number of trees with at most n vertices is asymptotically less than $2 \cdot C x_0^{-n}$. Then we have

$$
\begin{aligned}
|\mathcal{T}'_n| &\leq \sum_{k=1}^{\lfloor 1/24n \rfloor} 2^{\frac{n}{3}} \cdot 2C \cdot x_0^{-\frac{25}{48}n} \cdot \binom{\lfloor \frac{25}{48}n \rfloor}{k} \\
&\leq \frac{1}{24} n 2^{\frac{n}{3}} \cdot 2C \cdot x_0^{-\frac{25}{48}n} \cdot \binom{\lfloor \frac{25}{48}n \rfloor}{\lfloor \frac{1}{24}n \rfloor} \\
&= \frac{C}{12} n \cdot 2^{\frac{n}{3}} \cdot x_0^{-\frac{25}{48}n} \cdot \binom{\lfloor \frac{25}{48}n \rfloor}{\lfloor \frac{1}{24}n \rfloor}.
\end{aligned}
$$

By Stirling's approximation, i.e., $\frac{n!}{\sqrt{2\pi n}(\frac{n}{e})^n} \to 1$ as $n \to \infty$, we can get that when n is large enough,

$$
\binom{\lfloor \frac{25}{48}n \rfloor}{\lfloor \frac{1}{24}n \rfloor} < \frac{C_0}{\sqrt{n}} 1.2^n,
$$

where C_0 is a constant. Then,

$$
|\mathcal{T}'_n| < C_1 n^{1/2} \cdot 2^{\frac{n}{3}} x_0^{-\frac{25}{48}n} 1.2^n,
$$

where C_1 is some real number for large n. It is known that $|\mathcal{T}_n| \sim \frac{C \cdot x_0^{-n}}{n^{\frac{5}{2}}}$. Recall that $x_0 \approx 0.3383219$. Consequently, $\frac{|\mathcal{T}'_n|}{|\mathcal{T}_n|} \to 0$.

Hence, in conclusion we get that almost all trees do not belong to $\mathcal{T}'_n$. The proof is thus complete. □

Next, we proceed to estimate the number of non-isomorphic rooted trees obtained by rooting a tree from Theorem 3 and Lemma 4. The following theorem is established.

Theorem 5. *For almost all trees in $\mathcal{T}_n$, the number of non-isomorphic rooted trees obtained by rooting a tree is $(\mu_r + o(1))n$.*

Proof. By Lemma 4, we know that almost every tree has at least $\frac{1}{24}n$ fixed vertices, and denote the set of these trees by $\mathcal{T}_n^*$. Clearly, $\mathcal{T}_n^* \subseteq \mathcal{T}_n$ and $\frac{|\mathcal{T}_n^*|}{|\mathcal{T}_n|} \to 1$. Let T be a tree in $\mathcal{T}_n^*$. If we pick up one of the fixed vertices as the root, we can get a rooted tree having the same number of vertex classes. There are at least $\frac{1}{24}n$ rooted trees in which the roots of the rooted trees correspond to the fixed vertices of T. And the number of vertex classes equals that in T. Hence, there are at least $|\mathcal{T}_n^*| \cdot \frac{1}{24}n$ rooted trees in $\mathcal{R}_n$ such that the roots are fixed vertices in the associated tree. These rooted trees form a set $\mathcal{R}_n^*$. Notice that $|\mathcal{R}_n| \sim \frac{D \cdot x_0^{-n}}{n^{\frac{3}{2}}}$ and $|\mathcal{T}_n^*| \sim \frac{C \cdot x_0^{-n}}{n^{\frac{5}{2}}}$. We get $\frac{|\mathcal{R}_n^*|}{|\mathcal{R}_n|} \nrightarrow 0$. Combining this with Theorem 3, we have that the number of vertex classes is $(\mu_r + o(1))n$ for almost all rooted trees in $\mathcal{R}_n^*$.

According to whether the number of vertex classes is $(\mu_r + o(1))n$ or not, we divide $\mathcal{R}_n^*$ into two parts $\mathcal{R}_{n,1}^*$ and $\mathcal{R}_{n,2}^*$. There are at most $\frac{|\mathcal{R}_{n,2}^*|}{\frac{1}{24}n}$ trees in $\mathcal{T}_n^*$ corresponding to $\mathcal{R}_{n,2}^*$. Since $|\mathcal{R}_{n,2}^*| = o(|\mathcal{R}_n^*|) = o(|\mathcal{R}_n|)$, then $\frac{|\mathcal{R}_{n,2}^*|}{\frac{1}{24}n} = o(|\mathcal{T}_n|) = o(|\mathcal{T}_n^*|)$.

Therefore, almost all trees in $\mathcal{T}_n^*$ correspond to the rooted trees in $\mathcal{R}_{n,1}^*$. And recall that the root of the tree in $\mathcal{R}_{n,1}^*$ is a fixed vertex. That is, almost all trees in $\mathcal{T}_n^*$ also have $(\mu_r + o(1))n$ vertex classes. Consequently, almost every tree in $\mathcal{T}_n$ has $(\mu_r + o(1))n$ vertex classes. The proof is complete. □

From Theorem 5, we have an intuitive grasp that the rooted tree space is just the tree space with a scale $(\mu_r + o(1))n$. Not rigorously to say, if we consider any special structure in trees, the case that this structure will appear $(\mu_r + o(1))n$ times in rooted trees is in a large probability, and the probabilities of appearances in tree space and rooted tree space seem to be the same. Moreover, by the asymptotical values of $|\mathcal{R}_n|$ and $|\mathcal{T}_n|$, we can get that $\mu_r \approx 0.8210$.

4 The distribution for any pattern in $\mathcal{T}_n$

In this section, we shall focus on the distribution of the number of occurrences $X_{n,M}(T)$ of a pattern M on tree space $\mathcal{T}_n$. It is known that the distribution of the number of occurrences of a pattern in $\mathcal{R}_n$ is asymptotically normal. We refer to [9] for this. We show that the corresponding distribution in $\mathcal{T}_n$ is also asymptotically normal. It has been shown that $X_{n,M}(T)$ has mean $(\mu_M + o(1))n$ and variance $(\sigma_M + o(1))n$ and for almost every tree, and the number of non-isomorphic rooted trees obtained by rooting a tree is $(\mu_r + o(1))n$. The constants μ and σ are the same as those for the case of rooted trees, namely, $\mathbf{E}(X_{n,M}(T)) \sim \mu_M n \sim \mathbf{E}(X_{n,M}(R))$ and $\mathbf{Var}(X_{n,M}(T)) \sim \sigma_M n \sim \mathbf{Var}(X_{n,M}(R))$. Based on these two results, we proceed to get our final result.

Theorem 6. *For any given pattern, the number of occurrences of the pattern in trees* *is asymptotically normally distributed.*

Proof. Recall that for each given pattern M, $\mathbf{E}(X_{n,M}(T)) \sim \mu_M n$ and $\mathbf{Var}(X_{n,M}(T)) \sim \sigma_M n$, where μ_M and σ_M are some constants. Let $\mathcal{T}_n^1$ be the subset of $\mathcal{T}_n$ such that the number of occurrences $X_{n,M}(T)$ satisfies that $\frac{X_{n,M}(T)-\mu_M n}{\sqrt{\sigma_M n}} \leq t$, where t is some real number. Then, the probability

$$P\left(\frac{X_{n,M}(T)-\mu_M n}{\sqrt{\sigma_M n}} \leq t\right) = \frac{|\mathcal{T}_n^1|}{|\mathcal{T}_n|},$$

where $\mathcal{T}_n^1$ is the subset of $\mathcal{T}_n$. For $\mathcal{R}_n$, we shall try to show that

$$\lim_{n\to\infty} P\left(\frac{X_{n,M}(R)-\mu_M n}{\sqrt{\sigma_M n}} \leq t\right) = \lim_{n\to\infty} P\left(\frac{X_{n,M}(T)-\mu_M n}{\sqrt{\sigma_M n}} \leq t\right).$$

We knew that

$$\lim_{n\to\infty} P\left(\frac{X_{n,M}(R)-\mu_M n}{\sqrt{\sigma_M n}} \leq t\right) = N(0,1,t),$$

where $N(0,1,t)$ denotes the probability value of the normal distribution at t. Denote by $\mathcal{R}_n^1$ the set of rooted trees satisfying $\frac{X_{n,M}(R)-\mu_M n}{\sqrt{\sigma_M n}} \leq t$. The last equation holds from the fact that any pattern in $\mathcal{R}_n$ is asymptotically normally distributed.

If R is a rooted tree in $\mathcal{R}_n^1$ corresponding to $T \in \mathcal{T}_n$, then $X_{n,M}(R) = X_{n,M}(T)$. So, a tree T is in $\mathcal{T}_n^1$ if and only if all the associated rooted trees are in $\mathcal{R}_n^1$. We split $\mathcal{T}_n^1$ into two subsets, $\mathcal{T}_n^{1'}$ and $\mathcal{T}_n^{1''}$, one is the collection of trees corresponding to $(\mu_r + o(1))n$ rooted trees, and the other is not, respectively. By Theorem 5, the number of rooted trees corresponding to $\mathcal{T}_n^{1'}$ is $|\mathcal{T}_n^{1'}|\cdot(\mu(R)+o(1))n$, and $|\mathcal{T}_n^{1''}| = o(|\mathcal{T}_n^1|)$, i.e., the number of rooted trees associated with $\mathcal{T}_n^{1''}$ is at most $o(|\mathcal{T}_n^1|)\cdot n$. Then, it follows that

$$|\mathcal{T}_n^{1'}|\cdot(\mu_r + o(1))n \leq |\mathcal{R}_n^1| \leq |\mathcal{T}_n^{1'}|\cdot(\mu_r + o(1))n + o(|\mathcal{T}_n^1|)\cdot n.$$

Since $o(|\mathcal{T}_n^1|)\cdot n = o(|\mathcal{R}_n^1|)$ and $\frac{|\mathcal{T}_n^{1'}|}{|\mathcal{T}_n^1|} \sim 1$, we have $|\mathcal{R}_n^1| = (\mu_r + o(1))n \cdot |\mathcal{T}_n^1|$. Therefore, we get that

$$\begin{aligned}
P\left(\frac{X_{n,M}(R)-\mu_M n}{\sqrt{\sigma_M n}} \leq t\right) &= \frac{|\mathcal{R}_n^1|}{|\mathcal{R}_n|} \\
&\sim \frac{|\mathcal{T}_n^1|}{|\mathcal{T}_n|} \\
&= P\left(\frac{X_{n,M}(T)-\mu_M n}{\sqrt{\sigma_M n}} \leq t\right).
\end{aligned}$$

Consequently,

$$\begin{aligned}
\lim_{n\to\infty} P\left(\frac{X_{n,M}(T)-\mu_M n}{\sqrt{\sigma_M n}} \leq t\right) &= \lim_{n\to\infty} P\left(\frac{X_{n,M}(R)-\mu_M n}{\sqrt{\sigma_M n}} \leq t\right) \\
&= N(0,1,t).
\end{aligned}$$

Then the variable $X_{n,M}(T)$ is also asymptotically normal with mean $\sim \mu_M n$ and variance $\sim \sigma_M n$. The proof is now complete. $\square$

Now, we have established that for any pattern the limiting distribution of the number of occurrences in $\mathcal{T}_n$ is also normal, which solves an open question raised in [10].

5 Conclusion

In this paper, we explored the limiting distribution on the number of different rooted trees obtained by rooting a tree, and get the mean $\sim \mu_r n$ and variance $\sim \sigma_r n$. By the asymptotical values of $|\mathcal{R}_n|$ and $|\mathcal{T}_n|$, one can readily see that $\mu_r \approx 0.8210$. But we do not focus on how to calculate the two constants μ_r and σ_r in detail. We refer the readers to [16, 12] for more information, in which the authors did the computation for some other cases. Surely, in this paper, μ_r and σ_r can be expressed by the derivatives of $f(u)$ or $F(x, y, u)$ as in [3, 10], but we think that it is still much more complicated to get the numerical value.

Acknowledgement: The authors are very grateful to the referees for useful suggestions and comments, which are helpful for improving the presentation of the paper.

References

[1] B. Bollobás, *Random Graphs (2nd Ed.),* Cambridge Studies in Advanced Math., Vol.73, Cambridge University Press, Cambridge, 2001.

[2] M. Bóna, P. Flajolet, Isomorphism and symmetries in random phylogenetic trees, *J. Appl. Probab.*, **46**(2009), 1005–1019.

[3] F. Chyzak, M. Drmota, T. Klausner, G. Kok, The distribution of patterns in random trees, *Comb. Probab. Comp.* **17**(2008), 21–59.

[4] M. Drmota, Systems of functional equations, *Random Struct. Algor.* **10**(1997), 103–124.

[5] M. Drmota, B. Gittenberger, The distribution of nodes of given degree in random trees, *J. Graph Theory* **31**(1999), 227–253.

[6] P. Flajolet, R. Sedgewick, *Analytic combinatorics*, Cambridge University press, 2009.

[7] B. Gittenberger, Nodes of large degree in random trees and forests, *Random Struct. Algor.* **28**(2006), 374–385.

[8] F. Harary, E.M. Palmer, Graphical Enumeration, Academic Press, New York and London, 1973.

[9] G. Kok, Pattern distribution in various types of random trees, In: Proc. 2005 International Conference on Analysis of Algorithms, 223–230.

[10] G. Kok, The distribution of patterns in random trees. Thesis, Institut für Diskrete Mathematik und Geometrie, TU Wien, Austria, 2005.

[11] X. Li, Y. Li, The asymptotic value of the Randić index for trees, *Adv. Appl. Math.* **47**(2011), 365–378.

[12] K.A. McKeon, The expected number of symmetries in locally restricted trees II, *Discrete Appl. Math.* **66**(1996), 245–253.

[13] R. Otter, The number of trees, *Ann. Math* **49**(1948), 583–599.

[14] G. Pólya, Kombinatorische Anzahlbestimmungen fur Cruppcn, Graphen und chemische, *Acta Math* **68**(1937),145–254.

[15] K. Panagiotou, M.Sinha, Vertices of degree k in random unlabeled trees, *J. Graph Theory* **69**(2012), 114–130.

[16] R.W. Robinson, A.J. Schwenk, The distribution of degrees in a large random tree, *Discrete Math.* **12**(1975), 359–372.