The asymptotic number of occurrences of a subtree in trees with bounded maximum degree and an application to the Estrada index

Xueliang Li, Yiyang Li
Center for Combinatorics and LPMC-TJKLC
Nankai University, Tianjin 300071, China

Abstract

Let $T_n^\Delta$ denote the set of trees of order $n$, in which the degree of each vertex is bounded by some integer $\Delta$. Suppose that every tree in $T_n^\Delta$ is equally likely. For any given subtree $H$, we show that the number of occurrences of $H$ in trees of $T_n^\Delta$ is with mean $(\mu_H + o(1))n$ and variance $(\sigma_H + o(1))n$, where $\mu_H$, $\sigma_H$ are some constants. As an application, we estimate the value of the Estrada index $EE$ for almost all trees in $T_n^\Delta$, and give an explanation in theory to the approximate linear correlation between $EE$ and the first Zagreb index obtained by quantitative analysis.

Keywords: generating function, planted tree, rooted tree, tree, subtree, bounded degree, asymptotic number, Estrada index, Zagreb index.

AMS subject classification 2010: 05C05, 05C12, 05C30, 05D40, 05A15, 05A16, 92E10

1 Introduction

We denote the set of trees with bounded maximum degree $\Delta$ by $T_n^\Delta$. Setting $t_n = |T_n^\Delta|$, we introduce a generating function for these trees

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

Let $H$ be a given subtree. For a tree $T_n^\Delta$, we say that $H$ occurs in $T_n^\Delta$ if there is a subtree in $T_n^\Delta$ isomorphic to $H$. Denote the number of occurrence of $H$ in a tree $T_n^\Delta$ by $t_n^\Delta$. To count

*Supported by NSFC No.10831001, PCSIRT and the “973” program.
the occurrences, we introduce a generating function in two variables as follows

$$t(x, u) = \sum_{n \geq 1, \Delta \in \mathcal{T}_n} x^n u^\Delta_n.$$  

It can be simplified into

$$t(x, u) = \sum_{n \geq 1, k \geq 0} t_{n,k} x^n u^k,$$

where $t_{n,k}$ denotes the number of trees in $\mathcal{T}_n$ such that the number of occurrences of $H$ is $k$. Note that $t(x, 1) = t(x)$, i.e., $t_n = \sum_{k \geq 0} t_{n,k}$.

Furthermore, suppose that every tree in $\mathcal{T}_n$ is equally likely. We use $X_n(\mathcal{T}_n)$, or simply $X_n$, to denote the number of occurrences of $H$ in $T_n^\Delta$. Clearly, $X_n$ is a random variable in $\mathcal{T}_n^\Delta$. Then, the probability of $X_n$ can be defined as

$$\Pr[X_n = k] = \frac{t_{n,k}}{t_n}.$$  

Without the maximum degree restriction, many results have been established for some special substructures. We further assume that, when $H$ occurs in $T_n^\Delta$, the degrees of the internal vertices (vertices of degrees not equal to 1) should coincide with those of the corresponding vertices in the tree. In this case, $H$ is called a pattern of the tree. Kok [9] showed that the number of occurrences of a pattern in the trees is with mean $E(X_n) = (u + o(1))n$ and variance $Var(X_n) = (\sigma + o(1))n$, and $\frac{X_n - E(X_n)}{\sqrt{Var(X_n)}}$ is asymptotic to a distribution with density $(A + Bx^2)e^{-Cx^2}$ for some constants $A, B, C \geq 0$. Moreover, if the pattern is a star, then the number of occurrences of the pattern in a tree is exactly the number of vertices with a given degree with respect to the internal vertex of the star. It has been shown that for the number of vertices of a given degree $X_n$, $\frac{X_n - E(X_n)}{\sqrt{Var(X_n)}}$ is asymptotically normally distributed. We refer the readers to [5, 11] for more details. And, analogous results have been obtained for other classes of trees, such as simply generated trees, rooted trees, et al. (see [2], [5] and [9]). However, for the number of occurrences of a subtree $H$ in trees without the maximum degree restriction, there is no such a similar result obtained. And, it seems to be much difficult.

In this paper, we get that the number of occurrences of a subtree $H$ in the planted trees and rooted trees with bounded maximum degree is also asymptotically normally distributed with mean and variance in $\Theta(n)$. And for $\mathcal{T}_n^\Delta$, we get a weak result which does not show that the distribution is also asymptotically normal. And, as an application, we use this result to estimate the Estrada index $EE$ for these trees in $\mathcal{T}_n^\Delta$, and give an explanation in theory to the the approximate linear correlation between $EE$ and the first Zagreb index [7] obtained by quantitative analysis.

In this paper, Section 2 is devoted to a systematic treatment of the number of occurrences of a given subtree $H$. And in Section 3, we investigate the Estrada index for the trees in $\mathcal{T}_n^\Delta$.  

2
2 The number of occurrences of a given subtree

In this section, we proceed to show that the number of occurrences of a subtree $H$ in $\mathcal{T}_n^\Delta$ is with mean $(\mu_H + o(1))n$ and variance $(\sigma_H + o(1))n$ for some constants $\mu_H$ and $\sigma_H$. And in the procedure of proof, we get the related results for planted trees and rooted trees.

In what follows, we introduce some terminology and notation which will be used in the sequel. For the others not defined here, we refer to book [8].

Analogous to trees, we introduce generating functions for rooted trees and planted trees. Let $\mathcal{R}_n^\Delta$ denote the set of rooted trees of order $n$ with degrees bounded by an integer $\Delta$. Setting $r_n = |\mathcal{R}_n^\Delta|$, we have

$$r(x) = \sum_{n \geq 1} r_n x^n$$

and

$$r(x, u) = \sum_{n \geq 1, k \geq 0} r_{n,k} x^n u^k,$$

where $r_{n,k}$ denotes the number of trees in $\mathcal{R}_n^\Delta$ such that $H$ occurs $k$ times. A planted tree is formed by adding a vertex to the root of a rooted tree. The new vertex is called the plant, and we never count it in the sequel. Analogously, let $\mathcal{P}_n^\Delta$ denote the set of planted trees of order $n$ and bounded maximum degree $\Delta$. Setting $p_n = |\mathcal{P}_n^\Delta|$, we have

$$p(x) = \sum_{n \geq 1} p_n x^n$$

and

$$p(x, u) = \sum_{n \geq 1, k \geq 0} p_{n,k} x^n u^k,$$

where $p_{n,k}$ denotes the number of trees in $\mathcal{P}_n^\Delta$ such that $H$ occurs $k$ times. By the definition of planted trees, one can readily see that $p(x, 1) = p(x) = r(x, 1) = r(x)$.

Moreover, in [10], it has been showed that there exists a number $x_0$ such that

$$p(x) = b_1 + b_2 \sqrt{x_0 - x} + b_3 (x_0 - x) + \cdots,$$  \hspace{1cm} (1)

where $b_1, b_2, b_3$ are some constants not equal to zero. Evidently, $p(x_0) = b_1$ and for any $|x| \leq x_0$, $p(x)$ is convergent. For any $\Delta$, $x_0 \leq 1/2$, particularly, if $\Delta = 4$, $x_0 \approx 0.3551817$ and $p(x_0) \approx 1.117421$.

Let $p^{(\Delta-1)}(x)$ be the generating function of trees such that the degrees of the roots are not more than $\Delta - 1$, while the degrees of the other vertices are still bounded by $\Delta$. Then, we have (see [10])

$$p^{(\Delta-1)}(x_0) = 1.$$ \hspace{1cm} (2)

And, this fact will play an important role in the following proof.
Let \( y(x, u) = (y_1(x, u), \ldots, y_N(x, u))^T \) be a column vector. We suppose that \( G(x, y, u) \) is an analytic function with non-negative Taylor coefficients. \( G(x, y, u) \) can be expanded as
\[
G(x, y, u) = \sum_{n \geq 1, k \geq 0} g_{n,k} x^n u^k.
\]
Let \( X_n \) denote a random variable with probability
\[
\Pr[X_n = k] = \frac{g_{n,k}}{g_n},
\] (3)
where \( g_n = \sum_k g_{n,k} \). First, we introduce a useful lemma [2, 4].

**Lemma 1.** Let \( F(x, y, u) = (F_1(x, y, u), \ldots, F_N(x, y, u))^T \) be functions analytic around \( x = 0, y = (y_1, \ldots, y_N)^T = 0, u = 0 \), with Taylor coefficients all are non-negative. Suppose \( F(0, y, u) = 0, F(x, 0, u) \neq 0, F_x(x, y, u) \neq 0 \), and for some \( j \), \( F_{y^j}(x, y, u) \neq 0 \). Furthermore, assume that \( x = x_0 \) together with \( y = y_0 \) is a non-negative solution of the system of equations
\[
y = F(x, y, 1)
\] (4)
\[
0 = \det(I - F_y(x, y, 1))
\] (5)
inside the region of convergence of \( F \), \( I \) is the unit matrix. Let \( y = (y_1(x, u), \ldots, y_N(x, u))^T \) denote the analytic solution of the system
\[
y = F(x, y, u)
\] (6)
with \( y(0, u) = 0 \).

If the dependency graph \( G_F \) of the function system Equ.(6) is strongly connected, then there exist functions \( f(u) \) and \( g_i(x, u), h_i(x, u) \) \((1 \leq i \leq N)\) which are analytic around \( x = x_0, u = 1 \), such that
\[
y_i(x, u) = g_i(x, u) - h_i(x, u) \sqrt{1 - \frac{x}{f(u)}}
\] (7)
is analytically continued around \( u = 1, x = f(u) \) with \( \arg(x - f(u)) \neq 0 \), where \( x = f(u) \) together with \( y = y(f(u), u) \) is the solution of the extended system
\[
y = F(x, y, u)
\] (8)
\[
0 = \det(I - F_y(x, y, u)).
\] (9)

Moreover, let \( G(x, y, u) \) be an analytic function with non-negative Taylor coefficients such that the point \((x_0, y(x_0, 1), 1)\) is contained in the region of convergence. Finally, let \( X_n \) be the random variable defined in Equ.(3). Then the random variable \( X_n \) is asymptotically normal with mean
\[
E(X_n) = \mu n + O(1) \ (n \to \infty),
\]
and variance

$$Var(X_n) = \sigma_n + O(1) \ (n \to \infty)$$

with $\mu = -\frac{f''(1)}{f'(1)}$.

**Remark 1:** We say that the dependency graph $G_F$ of $y = F(x, y, u)$ is strongly connected if there is no subsystem of equations that can be solved independently from others. If $G_F$ is strongly connected, then $I - F_y(x_0, y_0, 1)$ has rank $N - 1$. Suppose that $\nu^T$ is a vector with $\nu^T(I - F_y(x_0, y_0, 1)) = 0$. Then, $\mu = \frac{\nu^T(F_u(x_0, y_0, 1))}{x_0 \nu^T(F_x(x_0, y_0, 1))}$. We refer the readers to [2, 4] for more details.

Now, we concentrate on considering the generating function $p(x, u)$.

For the subtree $H$, we suppose the diameter of $H$ is $h$. The depth of a vertex in a planted tree is the distance from the vertex to the root. The depth of a planted tree is the largest distance from the vertices to the root. We split up $P_n^\Delta$ into two sets $W_0$ and $W$, which denote the trees with depth not more than $h - 1$ and the trees with depth greater than $h - 1$, respectively. We can see that if $H$ occurs in the planted tree and the corresponding subtree takes the root, then the depth of the subtree is not more than $h$. Moreover, since we mainly consider the asymptotic number of subtree, the trees in $W_0$ will contribute nothing to the coefficient of $x^u u^k$ when $n$ is large enough. Therefore, in this paper, we do not need to know the exactly expression of the generating function for the trees in $W_0$, and denote it by $\phi(x, u)$. Moreover, the counting function of a subset in $W_0$ is also denoted by $\phi(x, u)$.

In what follows, we shall see that this assumption is reasonable. Then, we focus on the trees in $W$.

First, we introduce some conceptions. For a planted tree in $W$, the planted subtree formed by the vertices with depth not more than $\ell$ is called $\ell$-depth subtree of this tree. Now, we split up $W$ according to the $h$-depth subtree. That is, the trees in $W$ having the same $h$-depth subtree $w_i$ form a subset $H_i$ of $W$. Since the degrees of the vertices in $W$ are bounded by $\Delta$, there are finite number of different $h$-depth subtrees, i.e., for some $N_\Delta$, $1 \leq i \leq N_\Delta$. Therefore, we obtain that

$$p(x, u) = \phi(x, u) + \sum_{i=1}^{N_\Delta} a_{w_i, h}(x, u),$$

where $a_{w_i, h}(x, u)$ denotes the generating function of $H_i$.

To establish the functions system of $a_{w_i, h}(x, u)$, we need other functions $a_{w_i', h-1}(x, u)$ as follows. For some $(h - 1)$-depth subtree $w_i'$, we denote the subset of the planted trees in $W$ having $w_i'$ by $H_i'$. Note that $w_i' \notin H_i$. Then, the set of the planted trees having $w_i'$ consists of $H_i'$ and $w_i'$. If we use $a_{w_i', h-1}(x, u)$ to denote the generating function with respect to $H_i' \cup w_i'$, it follows that

$$a_{w_i', h-1}(x, u) = \sum_{w_i \in H_i'} a_{w_i, h}(x, u) + w_i'(x, u).$$

(11)
There will appear an expression of the form $Z(S_n, f(x, u))$ (or $f(x)$), 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 $f(x^i, u^i)$ (or $f(x^i)$). For instance, if $n = 3$, then $Z(S_3) = (1/3!)(s_1 + 3s_1s_2 + 2s_3)$, and $Z(S_3, f(x, u)) = (1/3!)(f(x, u)^3 + 3f(x, u)f(x^2, u^2) + 2f(x^3, u^3))$. We refer the readers to [8] for details.

Note that a planted tree can be seen as a root attached by some planted subtrees. Employing the classic Pólya enumeration theorem, we have $Z(S_j; p(x))$ as the counting series of the planted trees whose roots have degree $p$, and the coefficient of $x^r$ in $x \cdot Z(S_j; p(x))$ is the number of planted trees with $r$ vertices (see [8] p.51–54). Therefore,

$$p(x) = x \cdot \sum_{j=0}^{\Delta-1} Z(S_j; p(x)),$$

and

$$p^{(\Delta-1)}(x) = x \cdot \sum_{j=0}^{\Delta-2} Z(S_j; p(x)).$$

By means of the same method, $a_{w_i, h}(x, u)$ can be expressed in $a_{w_i', h-1}(x, u)$ and $\phi(x, u)$. Suppose that the roots of the trees in $H_i$ have degree $j$, and each has $j'$ planted subtrees with depth at least $h - 1$ attached to it. Clearly, $j'$ belongs to $\{1, \ldots, j - 1\}$, and some of these subtrees may have the same $w'_i$. Denote these different $(h - 1)$-depth subtrees by $\{w'_s\}$ and suppose $w'_s$ happens $\ell_s$ times. Evidently, $\sum \ell_s = j'$. It follows that

$$a_{w_i, h}(x, u) = x \cdot \prod Z(\ell_s; a_{w'_i, h-1}) \cdot \phi(x, u) \cdot u^{k(\ell_s, \phi)} \cdot (1 \leq i \leq N_\Delta). \quad (12)$$

Here, $\phi(x, u)$ denotes the counting function of the other $j - 1 - j'$ planted subtrees, since these subtrees belong to $W_0$. And the factor $u^{k(\ell_s, \phi)}$ serves to count the number of occurrences of $H$ using the root of the new tree. In this case, all these vertices of the new tree corresponding to the vertices of $H$ have depth not more than $h$. And, since we know that the $h$-depth subtree of the new tree is $w_i$, the number of occurrences taking the root can be calculated, that is, the upper index $k(\ell, \phi)$ can be calculated. Therefore, combining with Equ. (11), the functions system of $a_{w_i, h}(x, u)$ has been established.

Now, we start to show that all the conditions of Lemma [11] hold for $a_{w_i, h}(x, u).$ For convenience, we still use $F$ to denote the functions system. Set vector $a(x, u) = (a_{w_1, h}, \ldots, a_{w_{N_\Delta}, h}).$ We suppose $a_{w_i, h}(x, u) = F^i(x, a, u).$ Since $p(x, 1) = p(x)$ and $p(x_0) = b_1$, one can see that $a_{w_i, h}(x_0, 1)$ is convergent. So, $x_0$ and $a(x_0, 1)$ are inside the region of convergence of $F$. Apparently, the other conditions are easy to verify except for Equ. (11). In what follows, we shall show that the sum $S_{a_{w_i, h}}$ of every column of $F_a(x_0, a(x_0, 1), 1)$ equals 1. Consequently, the equation $\det(I - F_a(x_0, a(x_0, 1), 1)) = 0$ holds.

We consider the derivative on $a_{w_i, h}$. If $F^i(x, a, u)$ is not the function of $a_{w_{i'}, h}$, then $F^i_{a_{w_{i'}, h}}(x, a, u)$ will contribute nothing to the sum $S_{a_{w_i, h}}$. Thus, we just need to consider the
functions $F^i(x, a, u)$ with some $a_{x', h-1}$ having the term $a_{w_{0}, h}$. In Equ.(12), if both $a_{w_{1}', h-1}$ and $a_{w_{2}', h-1}$ have the term $a_{w_{0}, h}$, which implies that the trees corresponding to $a_{w_{1}', h-1}$, $a_{w_{2}', h-1}$ have the same $(h-1)$-depth subtree, then by the definition of $a_{w_{1}', h-1}$, we get that $a_{w_{1}', h-1} = a_{w_{2}', h-1}$. Therefore, there exists exactly one product factor, say $Z(l_{s_{0}}; a_{w_{0}, h-1})$, in the expression of $a_{w_{0}, h}$.

Moreover, it is well-known that the partial derivative of $Z(S_{n}; \cdot)$ enjoys (see [5])

$$\frac{\partial}{\partial s_{1}} Z(S_{n}; s_{1}, \ldots, s_{n}) = Z(S_{n-1}; s_{1}, \ldots, s_{n-1}). \quad (13)$$

For the planted tree, we have $\frac{\partial Z(S_{n}; p(x, 1))}{\partial p(x, 1)} = Z(S_{n-1}; p(x, 1))$, which equals the generating function obtained by deleting one subtree from the root. Analogously, we have

$$F^i_{a_{w_{0}, h}} = x \cdot \prod_{s \neq s_{0}} Z(\ell_{s}; a_{w_{0}', h-1}) \cdot Z(\ell_{s_{0}} - 1; a_{w_{0}', h-1}) \cdot \phi(x, u) \cdot u^{k(\ell_{s}, \phi)},$$

and it is exactly the new generating function produced by deleting one planted subtree of $\mathcal{H}_{0}' \cup w_{0}'$. Clearly, the root of the new planted tree is of degree $j - 1$. Particularly, if $\ell_{s_{0}} = 1$, after taking the derivative, the yielded function corresponds to the trees with roots of degree $j - 1$ such that every planted subtree does not belong to $\mathcal{H}_{0}' \cup w_{0}'$. Set $u = 1$. It follows that $S_{a_{w_{1}, h}}(x, a(x, 1), 1)$ is the generating function of the planted trees with roots of degree not more than $\Delta - 1$, i.e., $p^{(\Delta-1)}(x, 1)$. Combining with the fact $p^{(\Delta-1)}(x_{0}, 1) = 1$, we obtain $S_{a_{w_{1}, h}}(x_{0}, a(x_{0}, 1), 1) = 1$. Immediately, the Equ.(5)

$$\det(I - F_{a}(x_{0}, a(x_{0}, 1), 1)) = 0$$

follows.

Employing Lemma [1] we have that $a_{w_{1}, h}(x, u)$ is in the form of Equ.(7), namely, for some $f(u)$ and $g_{w_{1}, h}(x, u)$, $h_{w_{1}, h}(x, u)$ which are analytic around $x = x_{0}$, $u = 1$, it follows that

$$a_{w_{1}, h}(x, u) = g_{w_{1}, h}(x, u) - h_{w_{1}, h}(x, u) \sqrt{1 - \frac{x}{f(u)}}$$

is analytically continued around $u = 1$, $x = f(u)$ with arg$(x - f(u)) \neq 0$. From Equ.(10), we can see that $p(x, u)$ can be written into a function of $a(x, u)$, and denote it by $P(x, a(x, u), u)$. Clearly, all the coefficients of $P(x, a(x, u), u)$ are non-negative. Therefore, $p(x, u)$ is also in the form of Equ.(7). Moreover, recalling Equ.(11), we can see that $f(1) = x_{0}$. Apply Lemma [1] to $P(x, a(x, u), u)$, the following result is obtained.

**Theorem 2.** For any given subtree $H$, the number $X_{n}$ of occurrences of $H$ in $\mathcal{P}_{n}^{H}$ is asymptotical to be normal with mean $E(X_{n}) = \mu_{H}n + O(1)$ and variance $Var(X_{n}) = \sigma_{H}^{2}n + O(1)$ for some constants $\mu_{H}$ and $\sigma_{H}^{2}$.

A rooted tree in $\mathcal{R}_{n}^{H}$ can also be seen as a root attached by some planted trees. That is, by the classic Pólya enumeration theorem, analogous to Equ.(12), the generating function
of $R_n^\Delta$ is also a function in $a(x, u)$. We denote the function by $R(x, a(x, u), u)$, and $r(x, u) = R(x, a(x, u), u)$. By means of the above analysis, it is not difficult to see that the Taylor coefficients of $R(x, a(x, u), u)$ are non-negative. Thus, $r(x, u)$ also has the form of Equ.(7).

And, apply Lemma 1 to $R(x, a(x, u), u)$, the following result is obtained.

**Theorem 3.** For any given subtree $H$, the number $X_n$ of occurrences of $H$ in $R_n^\Delta$ is asymptotically normally distributed with mean $E(X_n) = \mu_H n + O(1)$ and variance $\text{Var}(X_n) = \sigma^{2}_H n + O(1)$ for some constants $\mu_H$ and $\sigma^{2}_H$.

**Remark 2:** Since $r(x, u)$ and $p(x, u)$ correspond to the same function $f(u)$, by Lemma 1 we can see that the means of $X_n$ with respect to $R_n^\Delta$ and $P_n^\Delta$ are with the same constant $\mu_H$. Moreover, it has been showed that the sum of each column of $F_a(x_0, a(x_0, 1), 1)$ equals 1, then we have $v^T = (1, \ldots, 1)$ such that $v^T (I - F_y(x_0, y_0, 1)) = 0$. Therefore, it is easy to see that $\mu_H$ is positive by Remark 1.

In what follows, we investigate the generating function of trees. Two edges in a tree are **similar**, if they are the same under some automorphism of the tree. To **join** two planted trees is to connect the two roots with a new edge and get rid of the two plants. If the two planted trees are the same, we say that the new edge is **symmetric**. Then, we have the following lemma due to [10].

**Lemma 4.** For any tree, the number of rooted trees corresponding to this tree minus the number of nonsimilar edges (except for the symmetric edge) is the number 1.

Note that, if we delete any one edge from a similar set in a tree, the yielded trees are the same two trees. Hence, different pairs of planted trees correspond to nonsimilar edges. Now, we have

$$t(x, u) = r(x, u) - \frac{1}{2} \left( \sum_{1 \leq i_1, i_2 \leq N^\Delta} a_{w_{i_1}, h}(x, u) a_{w_{i_2}, h}(x, u) \cdot u^{k_{w_{i_1}, w_{i_2}}} \right)$$

$$+ \frac{1}{2} \sum_{1 \leq i \leq N^\Delta} a_{w_i, h}(x^2, u^2) \cdot u^{k_{w_i, w_i}},$$

(14)

where $k(w_{i_1}, w_{i_2})$ serves to count the subtrees taking vertices both in $w_{i_1}$ and $w_{i_2}$. Consequently, we obtain that $t(x, u)$ is also in the form of Equ.(7), i.e., there exist some functions $\overline{g}(x, u), \overline{h}(x, u)$ which are analytic around $x = x_0$, $u = 1$, such that

$$t(x, u) = \overline{g}(x, u) - \overline{h}(x, u) \sqrt{1 - \frac{x}{f(u)}}$$

is analytically continued around $u = 1$, $x = f(u)$ with $\text{arg}(x - f(u)) \neq 0$. Here, we could not show that $t(x, u)$ likes $P(x, a(x, u), u)$ and $R(x, a(x, u), u)$ that have non-negative Taylor coefficients, so Lemma 1 fails in this case. However, we can use the following result due to [9] to get a weak result for $t(x, u)$.
Lemma 5. Suppose that $t(x, u)$ has the form
\[
t(x, u) = \overline{g}(x, u) - \overline{h}(x, u) \sqrt{1 - \frac{x}{f(u)}},
\]
where $\overline{g}(x, u)$, $\overline{h}(x, u)$ and $f(u)$ are analytic functions around $x = f(1)$ and $u = 1$ that satisfy $\overline{h}(f(1), 1) = 0$, $\overline{h}_x(f(1), 1) \neq 0$, $f(1) > 0$ and $f'(1) < 0$. Furthermore, $x = f(u)$ is the only singularity on the cycle $|x| = |f(u)|$ for $u$ is close to 1. Suppose that $X_n$ is defined as Equ. (3) to $y(x, u)$. Then, $E(X_n) = (\mu + o(1))n$ and $\text{Var}(X_n) = (\sigma + o(1))n$, where $\mu = -f'(1)/f(1)$ and $\sigma = \mu^2 + \mu - f''(1)/f(1)$.

Remark 3: This result does not tell us that the limiting distribution is asymptotically normal. If $h(f(1), 1) \neq 0$, this lemma is trivial by Lemma 4 and if $h(f(u), u) = 0$, we can still get that the limiting distribution is normal by further analysis (see [5]).

For $t(x)$, it has been obtained that [10]
\[
t(x) = c_0 + c_1(x_0 - x) + c_2(x_0 - x)^{3/2} + \cdots,
\]
where $c_0$, $c_1$, $c_2$ are some constants not equal to 0. Combining with the fact $t(x, 1) = t(x)$, we can see that $\overline{h}(f(1), 1) = 0$ and $\overline{h}_x(f(1), 1) \neq 0$. Moreover, the other conditions in Lemma 5 are easy to verify. Then, we formulate the following theorem.

Theorem 6. Let $X_n$ be the number of occurrences of a given subtree $H$ in the trees of $\mathcal{T}_n^\Delta$. Then it follows that
\[
E(X_n) = (\mu_H + o(1))n
\]
and
\[
\text{Var}(X_n) = (\sigma_H^t + o(1))n,
\]
where $\mu_H$ and $\sigma_H^t$ are some constants with respect to the subtree $H$.

Following book [1], we will say that almost every (a.e.) graph in a random graph space $\mathcal{G}_n$ has a certain property $Q$ if the probability $\Pr(Q)$ in $\mathcal{G}_n$ converges to 1 as $n$ tends to infinity. Occasionally, we shall write almost all instead of almost every.

By Chebyshev inequality one can get that
\[
\Pr \left[ \left| X_n - E(X_n) \right| > n^{3/4} \right] \leq \frac{\text{Var}(X_n)}{n^{3/2}} \to 0 \text{ as } n \to \infty.
\]
Therefore, for any subtree $H$, $X_n = (\mu_H + o(1))n$ a.e. in $\mathcal{T}_n^\Delta$. Then, an immediate consequence is the following.

Corollary 7. For almost all trees in $\mathcal{T}_n^\Delta$, the number of occurrences of $H$ equals $(\mu_H + o(1))n$. 

9
3 The Estrada index

In this section, we explore the Estrada index for trees in $T_n^\Delta$. Let $G$ be a simple graph with $n$ vertices. The eigenvalues of the adjacency matrix of $G$ are said to be the eigenvalues of $G$ and to form the spectrum. Suppose that the eigenvalues of $G$ are $\lambda_i$, $1 \leq i \leq n$. The Estrada index is defined as

$$EE(G) = \sum_{i=1}^{n} e^{\lambda_i}.$$ 

This index is invented in year 2000, and nowadays widely accepted and used in the information-theoretical and network-theoretical applications. And for this graph invariant, many results have been established. We refer the readers to a survey [3] for more details.

Furthermore, for trees with $n$ vertices, it has been showed that the path has the minimum Estrada index and the star has the maximum. And by quantitative analysis, there is an approximate linear correlation between $EE$ and the first Zagreb index, i.e., $\sum d_i^2$ for trees. Denote $\sum d_i^2$ by $D$. That is,

$$EE \approx aD + b,$$

(15)

where $a$ and $b$ are some constants. We refer the readers to [3] and [7].

In what follows, we shall get the estimate of $EE$ for almost all trees in $T_n^\Delta$ and give an explanation to the correlation (15) in theory.

Denoting by $M_k = M_k(G) = \sum_{i=1}^{n} \lambda_i^k$ the $k$-th spectral moment of $G$, and bearing in mind the power-series expansion of $e^x$, we have

$$EE(G) = \sum_{k=0}^{\infty} \frac{M_k(G)}{k!}.$$ 

Note that $M_k(G)$ is equal to the number of closed walks of length $k$. For trees, one can readily see that

$$EE(T) = \sum_{k=0}^{\infty} \frac{M_{2k}}{(2k)!},$$

(16)

Then, in a tree, the closed walk of length $2k$ forms a subtree with at most $k + 1$ vertices. We have got that, for any given subtree, the number of occurrences of the subtree in $T_n^\Delta$ equals $(\mu_H + o(1))n$ a.e. Since there are finite different subtrees with at most $k + 1$ vertices, and each subtree corresponds to finite numbers of $2k$ closed walks, we can obtain that there exists a constant $\mu_{2k}$ such that the number of $2k$ closed walk is $(\mu_{2k} + o(1))n$ a.e., namely,

$$M_{2k} = (\mu_{2k} + o(1))n$$

in $T_n^\Delta$. Moreover, we introduce a lemma due to Fiol and Garriga [6].

**Lemma 8.** For any graph $G$, $M_{2k} \leq \sum_{i=1}^{n} d_i^{2k}$.
Recall that the degrees of a tree in $T_n^\Delta$ are bounded by $\Delta$. So, $\sum_{i=1}^n d_i^\Delta \leq \Delta^2 n$ and thus $EE(T_n^\Delta) \leq e^\Delta n$. Moreover, since $\sum_{k=0}^{\Delta/2} \frac{\Delta^2}{(2k)!}$ is convergent, for any positive number $\varepsilon$, there exists an integer $j_0$ such that for any $j > j_0$, $\sum_{k=j+1}^{\Delta/2} \frac{M_k}{(2k)!} < \varepsilon n$. Evidently, it is uniform for all the trees in $T_n^\Delta$. Therefore, we have

$$\sum_{k=0}^j \frac{M_{2k}}{(2k)!} \leq EE(T_n^\Delta) \leq \sum_{k=0}^j \frac{M_{2k}}{(2k)!} + \varepsilon n.$$ 

Hence, we just contribute to consider the closed walks of length at most $j_0$.

For any integer $j$, we have $\sum_{k=0}^j \frac{\mu_{2k}}{(2k)!} \leq e^\Delta$. Therefore, $\sum_{k=0}^j \frac{\mu_{2k}}{(2k)!}$ is convergent, and denote the limit by $\mu_\Delta$. It follows that

$$(\mu_\Delta - \varepsilon)n < \sum_{k=0}^j \frac{M_{2k}}{(2k)!} = \sum_{k=0}^j \frac{(\mu_{2k} + o(1))n}{(2k)!} \leq (\mu_\Delta + o(1))n \text{ a.e.}$$

Then, we have that $(\mu_\Delta - \varepsilon)n < EE(T_n^\Delta) < (\mu_\Delta + \varepsilon)n \text{ a.e.}$ Now, we can formulate the following theorem.

**Theorem 9.** For any $\varepsilon > 0$, the Estrada index of a tree in $T_n^\Delta$ enjoys

$$(\mu_\Delta - \varepsilon)n < EE(T_n^\Delta) < (\mu_\Delta + \varepsilon)n \text{ a.e.,}$$

where $\mu_\Delta$ is some constant.

If we suppose that the given subtree $H$ is a path $L$ of length 2, then there exists some constant $u_L$ such that in $T_n^\Delta$, the number of occurrences $X_n$ of $L$ is $(u_L + o(1))n$ a.e. In this case, it is easy to see that for each tree $T_n^\Delta$, $X_n(T_n^\Delta) = \sum_i \binom{d_i}{2} = \frac{1}{2}D(T_n^\Delta) - n + 1$. Therefore, the value of $D$ also enjoys $(u_D + o(1))n$ a.e. for some constant $u_D$. Then, combining with Theorem [2] we can see that, for trees in $T_n^\Delta$, the correlation between $EE$ and $D$ is approximate to be linear.

**References**

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

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

[3] H. Deng, S. Radenković, I. Gutman, The Estrada index, in: D. Cvetkovic, I. Gutman (Eds.), *Applications of Graph Spectra*, Math. Inst., Belgrade, (2009), 123–140.

[4] M. Drmota, Systems of functional equations, *Random Struct. Alg.*, 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] M.A. Fiol, E. Garriga, Number of walks and degree powers in a graph, Discrete Math., 309 (2009), 2613–2614.

[7] I. Gutman, B. Furtula, B. Glišić, V. Marković, A. Vesel, Estrada index of acyclic molecules, Indian J. Chem., 46 (2007), 1321–1327.

[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 2005 International Conference on Analysis of Algorithms, 223–230.

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

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