The asymptotic behavior of the Estrada index for trees*

Xueliang Li, Yiyang Li

Center for Combinatorics and LPMC-TJKLC
Nankai University, Tianjin 300071, China
E-mail: lxl@nankai.edu.cn, liycldk@mail.nankai.edu.cn

Abstract

Let $\mathcal{T}_n^\Delta$ denote the set of trees of order n, in which the degree of each vertex is bounded by some integer Δ. Suppose that every tree in $\mathcal{T}_n^\Delta$ is equally likely. For any given subtree H, we first show that the number of occurrences of H in trees of $\mathcal{T}_n^\Delta$ has mean $(\mu_H + o(1))n$ and variance $(\sigma_H + o(1))n$, where μ_H, σ_H are some constants. Then we apply this result to estimate the value of the Estrada index EE for almost all trees in $\mathcal{T}_n^\Delta$, and give a theoretical explanation to the approximate linear correlation between EE and the first Zagreb index obtained by quantitative analysis.

1 Introduction

We denote the set of trees with n vertices and maximum degree at most Δ by $\mathcal{T}_n^\Delta$. Setting $t_n = |\mathcal{T}_n^\Delta|$, we introduce a generating function for these trees:

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

Let H be a given small tree. For a tree $T_n^\Delta \in \mathcal{T}_n^\Delta$, we say that H *occurs* in T_n^Δ if there is a subtree of T_n^Δ isomorphic to H. Denote the number of occurrences of H in a tree T_n^Δ by $t_{T_n^\Delta,H}$. To count the occurrences, we introduce a generating function in two variables as follows:

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

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

*Supported by NSFC No.10831001.

Furthermore, suppose that every tree in $\mathcal{T}_n^\Delta$ is equally likely. Then, we can regard $t_{T_n^\Delta,H}$ as a random variable $X_n(T_n^\Delta)$ in T_n^Δ on the space $\mathcal{T}_n^\Delta$, simply denoted by X_n. Clearly, the probability distribution of X_n is given by

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

If H occurs in a tree and the degrees of the internal vertices (vertices of degrees greater than 1) coincide with those of the corresponding vertices in the tree, then the corresponding subtree of the tree is called a *pattern* of H. If there is no degree restriction on the trees, many results have been established for the number of occurrences of a pattern. Kok [9] showed that the number X_n for any pattern in trees without degree restriction has mean $E(X_n) = (\mu + o(1))n$ 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 for this pattern in a tree is exactly the number of vertices with degrees equal to the degree of the internal vertex of the star. It has been shown that for the number X_n of vertices of a given degree, $\frac{X_n - E(X_n)}{\sqrt{Var(X_n)}}$ is asymptotically normally distributed. We refer the readers to [5, 12] 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], [9], [10]). However, for the number of occurrences of H in general trees, similar results have not been obtained so far. It seems that this is very difficult.

In this paper, we will first show that the number of occurrences of H in planted trees and rooted trees with bounded degree is also asymptotically normally distributed with mean and variance in $\Theta(n)$, but for $\mathcal{T}_n^\Delta$, we can only get a weak result. Then, we will use this result to estimate the Estrada index EE for the trees in $\mathcal{T}_n^\Delta$, and give a theoretical explanation to the approximate linear correlation between EE and the first Zagreb index [7] obtained by quantitative analysis. The definition of EE will be introduced in Section 3, and we refer the readers to a survey [3] for more information on the Estrada index.

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

2 The number of occurrences of a given small tree

In this section, we show that the number of occurrences of H in $\mathcal{T}_n^\Delta$ has mean $(\mu_H + o(1))n$ and variance $(\sigma_H + o(1))n$ for some constants μ_H and σ_H. In the procedure of our discussion, we get related results for planted trees and rooted trees first.

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

Analogous to trees, we introduce the 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 Δ. 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 in each of these trees. 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 with degrees bounded by Δ. 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 in each of these trees. By the definition of planted trees, one can readily see that $p(x,1) = p(x) = r(x,1) = r(x)$.

Moreover, in [11], it has been shown that there exists a number x_0 such that

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

where b_1, b_2, b_3 are some constants not equal to zero; for any $|x| \leq x_0$, $p(x)$ is convergent (evidently, $p(x_0) = b_1$); and for any $\Delta \geq 2$, $x_0 \leq 1/2$.

Let $p^{(\Delta-1)}(x)$ be the generating function of planted 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 Δ. Then, we have (see [11])

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

And, this fact will play an important role in the following proof.

Let $\mathbf{y}(x,u) = (y_1(x,u), \ldots, y_N(x,u))^T$ be a column vector. We suppose that $G(x, \mathbf{y}, u)$ is an analytic function with non-negative Taylor coefficients. $G(x, \mathbf{y}, u)$ can be expanded as

$$G(x, \mathbf{y}, u) = \sum_{n\geq 1, k\geq 0} g_{n,k} x^n u^k.$$

Let X_n denote a random variable with probability

$$\Pr[X_n = k] = \frac{g_{n,k}}{g_n}, \tag{3}$$

where $g_n = \sum_k g_{n,k}$. First, we introduce a useful lemma [2, 4].

Lemma 2.1. *Let $\boldsymbol{F}(x, \boldsymbol{y}, u) = (F^1(x, \boldsymbol{y}, u), \ldots, F^N(x, \boldsymbol{y}, u))^T$ be functions analytic around $x = 0$, $\boldsymbol{y} = (y_1, \ldots, y_N)^T = \boldsymbol{0}$, $u = 0$, with Taylor coefficients all are non-negative. Suppose $\boldsymbol{F}(0, \boldsymbol{y}, u) = \boldsymbol{0}$, $\boldsymbol{F}(x, \boldsymbol{0}, u) \neq \boldsymbol{0}$, $\boldsymbol{F}_x(x, \boldsymbol{y}, u) \neq \boldsymbol{0}$, and for some j, $\boldsymbol{F}_{y_j y_j}(x, \boldsymbol{y}, u) \neq \boldsymbol{0}$. Furthermore, assume that $x = x_0$ together with $\boldsymbol{y} = \boldsymbol{y}_0$ is a non-negative solution of the system of equations*

$$\boldsymbol{y} = \boldsymbol{F}(x, \boldsymbol{y}, 1) \tag{4}$$

$$0 = det(\boldsymbol{I} - \boldsymbol{F}_{\boldsymbol{y}}(x, \boldsymbol{y}, 1)) \tag{5}$$

inside the region of convergence of $\boldsymbol{F}$, $\boldsymbol{I}$ is the unit matrix. Let $\boldsymbol{y} = (y_1(x, u), \ldots, y_N(x, u))^T$ denote the analytic solution of the system

$$\boldsymbol{y} = \boldsymbol{F}(x, \boldsymbol{y}, u) \tag{6}$$

with $\boldsymbol{y}(0, u) = \boldsymbol{0}$.

If the dependency graph $G_{\boldsymbol{F}}$ 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

$$y_i(x, u) = g_i(x, u) - h_i(x, u)\sqrt{1 - \frac{x}{f(u)}} \tag{7}$$

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

$$\boldsymbol{y} = \boldsymbol{F}(x, \boldsymbol{y}, u) \tag{8}$$

$$0 = det(\boldsymbol{I} - \boldsymbol{F}_{\boldsymbol{y}}(x, \boldsymbol{y}, u)). \tag{9}$$

Moreover, let $G(x, \boldsymbol{y}, u)$ be an analytic function with non-negative Taylor coefficients such that the point $(x_0, \boldsymbol{y}(x_0, 1), 1)$ 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* $G_{\mathbf{F}}$ of $\mathbf{y} = \mathbf{F}(x, \mathbf{y}, u)$ is strongly connected if there is no subsystem of equations that can be solved independently from others. If $G_{\mathbf{F}}$ is strongly connected, then $\mathbf{I} - \mathbf{F}_{\mathbf{y}}(x_0, \mathbf{y}_0, 1)$ has rank $N-1$. Suppose that $\mathbf{v}^T$ is a vector with $\mathbf{v}^T(\mathbf{I} - \mathbf{F}_{\mathbf{y}}(x_0, \mathbf{y_0}, 1)) = 0$. Then, $\mu = \frac{\mathbf{v}^T(\mathbf{F}_u(x_0,\mathbf{y_0},1))}{x_0\mathbf{v}^T(\mathbf{F}_x(x_0,\mathbf{y_0},1))}$. We refer the readers to [2, 4] for more details.

Now, we focus our attention on the generating function $p(x, u)$.

For the subtree H, we suppose that the diameter of H is h. The *height* of a vertex in a planted tree is the distance from the vertex to the root. The *height of a planted tree* is the largest distance from the vertices to the root. We split up $\mathcal{P}_n^{\Delta}$ into two sets $\mathcal{W}_0$ and $\mathcal{W}$, which denotes the set of trees with height not more than $h-1$ and the trees with height greater than $h-1$, respectively. We can see that if H occurs in the planted tree and the corresponding subtree contains the root, then the height of the subtree is not more than h. Moreover, since we mainly consider the asymptotic number of subtrees, the trees in $\mathcal{W}_0$ will contribute nothing to the coefficient of $x^n u^k$ for any k when n is large enough. Therefore, in this paper, we do not need to know the exact expression of the generating function for the trees in $\mathcal{W}_0$, and we denote it by $\phi(x, u)$. Now, we focus on the trees in $\mathcal{W}$.

First, we introduce some concepts. For a planted tree in $\mathcal{W}$, the planted subtree formed by the vertices with height not more than ℓ is called *ℓ-height subtree* of this tree. Now, we split up $\mathcal{W}$ according to the h-height subtree. That is, the trees in $\mathcal{W}$ having the same h-height subtree w_i form a subset $\mathcal{H}_i$ of $\mathcal{W}$. Since the degrees of the vertices in $\mathcal{W}$ are bounded by Δ, there are finite number N_Δ of different h-height subtrees. So, $1 \le i \le N_\Delta$. Therefore, we obtain that

$$p(x, u) = \phi(x, u) + \sum_{i=1}^{N_\Delta} a_{w_i,h}(x, u), \tag{10}$$

where $a_{w_i,h}(x, u)$ denotes the generating function of $\mathcal{H}_i$.

To establish the system of functional equations for $a_{w_i,h}(x, u)$, we need other functions $a_{w_i',h-1}(x, u)$ as follows. For some tree w_i' of height $h-1$, we denote $\mathcal{H}_i'$ to be the subset of $\mathcal{W}$ such that the $(h-1)$-height subtree of each planted tree in $\mathcal{H}_i'$ is w_i'. Note that $w_i' \notin \mathcal{H}_i'$. Then, we use $a_{w_i',h-1}(x, u)$ to denote the generating function of $\mathcal{H}_i' \cup \{w_i'\}$, it follows that

$$a_{w_i',h-1}(x, u) = \sum_{w_i \in \mathcal{H}_i'} a_{w_i,h}(x, u) + w_i'(x, u), \tag{11}$$

where $w_i'(x, u)$ serves to count the occurrences of H on w_i'.

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 $Z(S_3) = (1/3!)(s_1^3 + 3s_1s_2 + 2s_3)$, 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 to some branches, and each branch is also a planted subtree. Employing the classic Pólya enumeration theorem, we have $Z(S_{j-1}; p(x))$ as the counting series of the planted trees whose roots have degree j, and the coefficient of x^p in $x \cdot Z(S_{j-1}; p(x))$ is the number of planted trees with p 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 terms of $a_{w_i',h-1}(x, u)$. Suppose that the roots of the trees in $\mathcal{H}_i$ have degree j, and each has j' planted subtrees with height 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 $(h - 1)$-height subtrees by $\{w_s'\}$ and suppose w_s' happens ℓ_s times. Evidently, $\sum \ell_s = j'$. It follows that

$$a_{w_i,h}(x, u) = x \cdot \prod_s Z(S_{\ell_s}; a_{w_s',h-1}) \cdot \phi_{w_i}(x, u) \cdot u^{k(\ell_s,\phi_{w_i})}, \ (1 \leq i \leq N_\Delta). \tag{12}$$

Here, $\phi_{w_i}(x, u)$ denotes the counting function of the other $j - 1 - j'$ branches of w_i. The factor $u^{k(\ell_s,\phi_{w_i})}$ serves to count the number of occurrences of H using the root of the new tree, and $k(\ell_s, \phi_{w_i})$ denotes the corresponding number. In this case, all the vertices of the new tree corresponding the vertices of H have height not more than h. And, since we know that the h-height subtree of the new tree is w_i, the number of occurrences including the root can be calculated, that is, the exponent $k(l_s, \phi_{w_i})$ 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 2.1 hold for $a_{w_i,h}(x, u)$. For convenience, we still use $\mathbf{F}$ to denote the functions system. Set vector $\mathbf{a}(x, u) = (a_{w_1,h}, \ldots, a_{w_{N_\Delta},h})^T$. We suppose that the i-th component $F^i(x, \mathbf{a}, u)$ of $\mathbf{F}$ equals $a_{w_i,h}(x, u)$. 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 $\mathbf{a}(x_0, 1)$ are inside the region of convergence of $\mathbf{F}$. Clearly, the other conditions are easy to verify except for Equ.(5). In what follows, we shall show that the sum $S_{a_{w_i,h}}$ of every column of $\mathbf{F_a}(x_0, \mathbf{a}(x_0, 1), 1)$ equals 1. Consequently, the equation $\det(\mathbf{I} - \mathbf{F_a}(x_0, \mathbf{a}(x_0, 1), 1)) = 0$ holds.

We consider the derivative to $a_{w_{i_0},h}$. Suppose the degree of the root of w_{i_0} is j. If $F^i(x,\mathbf{a},u)$ is not a function of $a_{w_{i_0},h}$, then $F^i_{a_{w_{i_0},h}}(x,\mathbf{a},u)$ will contribute nothing to the sum $S_{a_{w_{i_0},h}}$. Thus, we just need to consider the functions $F^i(x,\mathbf{a},u)$ with some $a_{w'_s,h-1}$ having the term $a_{w_{i_0},h}$. In Equ.(12), if both $a_{w'_{s_1},h-1}$ and $a_{w'_{s_2},h-1}$ have the term $a_{w_{i_0},h}$, which implies that the trees corresponding to $a_{w'_{s_1},h-1}$, $a_{w'_{s_2},h-1}$ have the same $(h-1)$-height subtree, then by the definition of $a_{w'_s,h-1}$, we get that $a_{w'_{s_1},h-1} = a_{w'_{s_2},h-1}$. Therefore, there exists exactly one product factor, say $Z(S_{\ell_{s_0}}; a_{w'_{s_0},h-1})$, that is a function of $a_{w_{i_0},h}$.

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

$$\frac{\partial}{\partial s_1} Z(S_n; s_1, \ldots, s_n) = Z(S_{n-1}; s_1, \ldots, s_{n-1}). \tag{13}$$

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 corresponds to the generating function obtained by deleting one branch from the root. Analogously, we have

$$F^i_{a_{w_{i_0},h}} = x \cdot \prod_{s\neq s_0} Z(S_{\ell_s}; a_{w'_s,h-1}) \cdot Z(S_{\ell_{s_0}-1}; a_{w'_{s_0},h-1}) \cdot \phi_{w_{i_0}}(x,u) \cdot u^{k(\ell_s,\phi_{w_{i_0}})},$$

and it is exactly the new generating function produced by deleting one brahch of $\mathcal{H}'_{s_0} \cup w'_{s_0}$. 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 branch does not belong to $\mathcal{H}'_{s_0} \cup \{w'_{s_0}\}$. Hence, $S_{a_{w_{i_0},h}}(x,\mathbf{a}(x,u),u)$ counts the number of occurrences in all planted trees with roots of degree not more than $\Delta - 1$. Set $u = 1$. Generally, it follows that $S_{a_{w_i,h}}(x,\mathbf{a}(x,1),1)$ equals the generating function $p^{(\Delta-1)}(x,1)$. Combining with the fact $p^{(\Delta-1)}(x_0,1) = 1$, we obtain $S_{a_{w_i,h}}(x_0,\mathbf{a}(x_0,1),1) = 1$. Immediately, the Equ.(5)

$$\det(\mathbf{I} - \mathbf{F_a}(x_0,\mathbf{a}(x_0,1),1)) = 0$$

follows.

Employing Lemma 2.1, we have that $a_{w_i,h}(x,u)$ is in the form of Equ.(7), namely, for some $f(u)$ and $g_{w_i,h}(x,u)$, $h_{w_i,h}(x,u)$ which are analytic around $x = x_0$, $u = 1$, it follows that

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

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 $\mathbf{a}(x,u)$, and denote it by $P(x,\mathbf{a}(x,u),u)$. Clearly, all the coefficients of $P(x,\mathbf{a}(x,u),u)$ are non-negative. Therefore, $p(x,u)$ is also in the form of Equ.(7). Moreover, recalling Equ.(1), we can see that $f(1) = x_0$. Apply Lemma 2.1 to $P(x,\mathbf{a}(x,u),u)$, the following result is obtained.

Theorem 2.2. *For any given subtree H, the number X_n of occurrences of H in $\mathcal{P}_n^{\Delta}$ is asymptotical to be normal with mean $E(X_n) = \mu_H n + O(1)$ and variance $Var(X_n) = \sigma_H^p n + O(1)$ for some constants μ_H and σ_H^p.*

A rooted tree in $\mathcal{R}_n^{\Delta}$ 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 $\mathcal{R}_n^{\Delta}$ is also a function in $\mathbf{a}(x,u)$. We denote the function by $R(x,\mathbf{a}(x,u),u)$, and $r(x,u) = R(x,\mathbf{a}(x,u),u)$. By means of the above analysis, it is not difficult to see that the Taylor coefficients of $R(x,\mathbf{a}(x,u),u)$ are non-negative. Thus, $r(x,u)$ also has the form of Equ.(7). And, apply Lemma 2.1 to $R(x,\mathbf{a}(x,u),u)$, the following result is obtained.

Theorem 2.3. *For any given subtree H, the number X_n of occurrences of H in $\mathcal{R}_n^{\Delta}$ is asymptotically normally distributed with mean $E(X_n) = \mu_H n + O(1)$ and variance $Var(X_n) = \sigma_H^r n + O(1)$ for some constants μ_H and σ_H^r.*

Remark 2: Since $r(x,u)$ and $p(x,u)$ correspond to the same function $f(u)$, by Lemma 2.1 we can see that the means of X_n with respect to $\mathcal{R}_n^{\Delta}$ and $\mathcal{P}_n^{\Delta}$ are with the same constant μ_H. Moreover, it has been shown that the sum of each column of $\mathbf{F_a}(x_0,\mathbf{a}(x_0,1),1)$ equals 1, then we have $\mathbf{v}^T = (1,\ldots,1)$ such that $\mathbf{v}^T(\mathbf{I} - \mathbf{F_y}(x_0,\mathbf{y_0},1)) = 0$. Therefore, it is easy to see that μ_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 panted trees are the same, we say that the new edge is *symmetric*. Then, we have the following lemma due to [11].

Lemma 2.4. *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

$$\begin{aligned} t(x,u) =& r(x,u) - \frac{1}{2}\Big(\sum_{1\le i_1,i_2\le N_\Delta} a_{w_{i_1},h}(x,u) a_{w_{i_2},h}(x,u)\cdot u^{k(w_{i_1},w_{i_2})}\Big) \\ &+ \frac{1}{2}\sum_{1\le i\le N_\Delta} a_{w_i,h}(x^2,u^2)\cdot u^{k(w_i,w_i)}, \end{aligned} \tag{14}$$

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 $\arg(x - f(u)) \neq 0$. Here, we could not get the result of trees likes planted trees and rooted trees. Some instances show that $t(x, u)$ does not have non-negative Taylor coefficients of $a_{w_{i_1},h}$ and $a_{w_{i_2},h}$, so Lemma 2.1 fails in this case. However, we can use the following result due to [9] to get a weak result for $t(x, u)$.

Lemma 2.5. *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 circle $|x| = |f(u)|$ 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 $Var(X_n) = (\sigma + o(1))n$, where $\mu = -f'(1)/f(1)$ and σ is some constant.

Remark 3: If $\overline{h}(f(1), 1) \neq 0$, this lemma is trivial by Lemma 2.1. But if $\overline{h}(f(u), u) = 0$, we can still get that the limiting distribution of X_n is normal by further analysis (see [5]).

For $t(x)$, it has been obtained that [11]

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

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 2.5 are easy to verify. Then, we formulate the following theorem.

Theorem 2.6. *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

$$Var(X_n) = (\sigma_H^t + o(1))n,$$

where μ_H and σ_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

$$\Pr\big[|X_n - E(X_n)| > n^{3/4}\big] \leq \frac{Var(X_n)}{n^{3/2}} \to 0 \text{ as } n \to \infty.$$

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

3 The Estrada index

In this section, we investigate the Estrada index for trees in $\mathcal{T}_n^{\Delta}$. 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 λ_i, $1 \leq i \leq n$. The Estrada index was defined as

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

This index was invented in year 2000, and is nowadays widely accepted and used in the information-theoretical and network-theoretical applications. 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 shown that the path has the minimum Estrada index and the star has the maximum. 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,

$$EE \approx aD + b, \tag{15}$$

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 $\mathcal{T}_n^{\Delta}$ and give theoretical explanation to the correlation (15).

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

$$EE(T) = \sum_{k=0}^{\infty} \frac{M_{2k}}{(2k)!}. \tag{16}$$

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 $\mathcal{T}_n^{\Delta}$ equals $(\mu_H + o(1))n$ a.e. Since there are finitely many 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 μ_{2k} such that the number of $2k$ closed walks is $(\mu_{2k}+o(1))n$ a.e., namely,

$$M_{2k} = (\mu_{2k}+o(1))n \text{ a.e.}$$

in $\mathcal{T}_n^{\Delta}$. Moreover, we introduce a lemma due to Fiol and Garriga [6].

Lemma 3.1. *For any graph G, $M_{2k} \leq \sum_{i=1}^{n} d_i^{2k}$*

Recall that the degrees of a tree in $\mathcal{T}_n^{\Delta}$ are bounded by Δ. So, $\sum_{i=1}^{n} d_i^{2k} \leq \Delta^{2k} n$ and thus $EE(T_n^{\Delta}) \leq e^{\Delta} n$. Moreover, since $\sum_{k=0} \frac{\Delta^{2k}}{(2k)!}$ is convergent, for any positive number ε, there exists an integer j_0 such that for any $j > j_0$, $\sum_{k=j+1} \frac{M_{2k}}{(2k)!} < \varepsilon n$. Evidently, it is uniform for all the trees in $\mathcal{T}_n^{\Delta}$. 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 have 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, $\sum_{k=0} \frac{\mu_{2k}}{(2k)!}$ is convergent, and denote the limit by μ_{Δ}. 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 $(\mu_{\Delta}-\varepsilon)n < EE(T_n^{\Delta}) < (\mu_{\Delta}+\varepsilon)n$ a.e. Now, we can formulate the following theorem.

Theorem 3.2. *For any $\varepsilon > 0$, the Estrada index of a tree in $\mathcal{T}_n^{\Delta}$ enjoys*

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

where μ_{Δ} 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 $\mathcal{T}_n^{\Delta}$, 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^{Δ}, $X_n(T_n^{\Delta}) = \sum_i \binom{d_i}{2} = \frac{1}{2}D(T_n^{\Delta}) - n + 1$. Therefore, the value of D also enjoys $(u_D+o(1))n$ a.e. for some constant u_D. Then, combining with Theorem 3.2, we can see that, for trees in $\mathcal{T}_n^{\Delta}$, the correlation between EE and D is approximately 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. Cvetković, 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] X. Li, Y. Li, The asymptotic value of the Randić index for trees, *Adv. Appl. Math.*, doi:10.1016/j.aam.2010.10.008, in press.

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

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