The asymptotic value of the Randić index for trees*

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

Abstract

Let $\mathcal{T}_n$ denote the set of all unrooted and unlabeled trees with n vertices, and (i, j) a double-star. By assuming that every tree of $\mathcal{T}_n$ is equally likely, we show that the limiting distribution of the number of occurrences of the double-star (i, j) in $\mathcal{T}_n$ is normal. Based on this result, we obtain the asymptotic value of the Randić index for trees. Fajtlowicz conjectured that for any connected graph G the Randić index of G is at least its average distance. Using this asymptotic value, we show that this conjecture is true not only for almost all connected graphs but also for almost all trees.

Keywords: generating function, tree, double-star, normal distribution, asymptotic value, (general) Randić index, average distance.

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

1 Introduction

Let $\mathcal{T}_n$ denote the set of all unrooted and unlabeled trees T_n with n vertices. A *pattern* $\mathcal{M}$ is a given small tree. We say that $\mathcal{M}$ *occurs* in a tree T_n if $\mathcal{M}$ is a subtree of T_n such that except for the vertices of $\mathcal{M}$ with degree 1, the other vertices must have the same degrees with the corresponding vertices in T_n. Surely, we can also let the vertices with degree 1 match with each other. Set $t_n = |\mathcal{T}_n|$. We introduce two functions:

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

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

where the coefficients $t_{n,k}$ denote the number of trees in $\mathcal{T}_n$ that have k occurrences of the pattern $\mathcal{M}$. We assume that every tree of $\mathcal{T}_n$ is equally likely. Let X_n denote the number of occurrences of $\mathcal{M}$ in a tree of $\mathcal{T}_n$. Therefore, X_n is a random variable on $\mathcal{T}_n$ with probability

$$\Pr[X_n = k] = \frac{t_{n,k}}{t_n}. \quad (1)$$

*Supported by NSFC No.10831001, PCSIRT and the "973" program.

In [7], Kok showed that for any pattern $\mathcal{M}$ the limiting distribution of $(X_n - EX_n)/\sqrt{VarX_n}$ is a distribution with density of the form $(\alpha+\beta t^2)\,\mathrm{e}^{-\gamma t^2}$, and $E(X_n) = (\mu+o(1))n$ and $Var(X_n) = (\sigma+o(1))n$, where $\alpha, \beta, \gamma, \mu$ and σ are some constants. Clearly, if $\beta = 0$, it is a normal distribution. It was shown that if the pattern is a *star* or a *path*, the corresponding distribution is asymptotically normal. We refer the readers to [7, 12, 13] for more details.

Recall that a *path* is a graph with a sequence of vertices such that there is an edge between every two consecutive vertices. A *star* is a complete bipartite graph such that one part of the bipartition contains only one vertex, and we call this vertex the *center* of the star. A *double-star* is a graph which is formed from two stars by connecting their centers with an edge.

In this paper, we will show that if the pattern $\mathcal{M}$ is a *double-star*, the corresponding limiting distribution is also a normal distribution, and get an estimate for the number X_n of occurrences of a double-star for almost all trees. Based on this result, we then obtain the asymptotic value of the *Randić index* for almost all trees in $\mathcal{T}_n$.

The Randić index was introduced by Randić [11] in 1975, and later, Bollobás and Erdös [2] generalized it to the *general Randić index*. The definition will be given in Section 3, and for a detailed survey we refer the readers to [9]. There is a conjecture on the relation between the Randić index and the average distance of a connected graph, proposed by Fajtlowicz in [5], which is stated as follows:

Conjecture 1. *Let $R(G)$ and $D(G)$ denote, respectively, the Randić index and the average distance of a graph G. Then, for any connected graph G, $R(G) \geq D(G)$.*

We will show that the conjecture is true not only for almost all connected graphs but also for almost all trees.

In Section 2, we explore the limiting distribution of X_n corresponding to a double-star. In Section 3, we apply the results in Section 2 to the Randić index.

2 The distribution of X_n for a double-star

In this section, we concentrate on the limiting distribution of X_n for a double-star. Throughout this paper, we use (i, j) to denote the double-star with one vertex corresponding to a center of degree i and the other of degree j. Evidently, the number of occurrences of (i, j) in a tree is the number of edges in the tree such that one end of the edge is of degree i while the other is of degree j. Without loss of generality, we always assume $i \leq j$.

In what follows, we first introduce some terminology and notation, which will be used in the sequel. For those not defined here, we refer the readers to the book [6].

Analogous to trees, we also have generating functions for rooted trees and planted trees. Let $\mathcal{R}_n$ be the set of all rooted trees with n vertices, and $r_n = |\mathcal{R}_n|$. We have

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

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

and $r_{n,k}$ is the number of all rooted trees in $\mathcal{T}_n$ that have k occurrences of (i,j). A *planted tree* is formed from a rooted tree and a new vertex by connecting the vertex and the root of the rooted tree with a new edge. The new vertex is called the *plant*, and we never count it in the sequel. Let $\mathcal{P}_n$ denote the set of all planted trees with n vertices and $p_n = |\mathcal{P}_n|$. Then, we have generating functions:

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

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

where $p_{n,k}$ denotes the number of planted trees in $\mathcal{P}_n$ that have k occurrences of (i,j). By the definitions of planted trees and rooted trees, it is easy to see that

$$r(x,1) = r(x) = p(x,1) = p(x).$$

Furthermore, suppose that the radius of the convergence of $r(x)$ is x_0, Otter [10] showed that x_0 satisfies $r(x_0) = 1$ and the asymptotic expansion of $r(x)$ is

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

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

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

To show that the limiting distribution of the number of occurrences of the double-star (i,j) for all trees is normal, we first introduce a useful lemma, which was used to explore the distribution of the number of occurrences of a pattern for some other families of trees, such as planar trees, labelled trees, rooted trees, *et al.* We refer the readers to [3, 4] for more details.

Lemma 1. *[3, 4] Let $\boldsymbol{F}(x,\boldsymbol{y},u) = (F_1(x,\boldsymbol{y},u),\ldots,F_N(x,\boldsymbol{y},u))^T$ be a vector function, in which every function $F_i(x,\boldsymbol{y},u)$ is analytic at $x=0$, $\boldsymbol{y} = (y_1,\ldots,y_N)^T = \boldsymbol{0}$, $u=0$, with Taylor coefficients that are all 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 $x = x_0$, $\boldsymbol{y} = \boldsymbol{y}_0$ 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 the convergence of $\boldsymbol{F}$, and $\boldsymbol{I}$ 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 (6) is strongly connected, then $y_i(x,u)$ is in the form of

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

where $x(u)$, $g_i(x,u)$ and $h_i(x,u)$ are analytic around $x = x_0$, $u = 1$. And, $y_i(x,u)$ is analytically continued around $x = x(u)$, $u = 1$ with $arg(x - x(u)) \neq 0$. Moreover, $x(u)$ with $x(1) = x_0$ and $\boldsymbol{y}(u) = \boldsymbol{y}(x(u), u)$ are the solution of the system

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

Here and in what follows, for any function f or vector function $\mathbf{F}$ we use f_x or $\mathbf{F_x}$ to denote its the partial derivative, where x or $\mathbf{x}$ is a variable of the function.

Remark 1. *The dependency graph $G_{\boldsymbol{F}}$ of $\boldsymbol{y} = \boldsymbol{F}(x, \boldsymbol{y}, u)$ is strongly connected, if there is no subsystem of equations that can be solved independently from others. If $G_{\boldsymbol{F}}$ is strongly connected, then $\boldsymbol{I} - \boldsymbol{F}_{\boldsymbol{y}}(x_0, \boldsymbol{y}_0, 1)$ has rank $N - 1$. We refer the readers to [3, 4] for more details.*

Now, we are going to establish functional equations for a pattern (i, j). For i and j, we distinguish the following three cases. Since only the tree with exactly two vertices contains the pattern $(1, 1)$, we do not need to consider the case for $i = j = 1$.

Case 1. $i \neq j > 1$.

We split $\mathcal{P}_n$ into three subsets according to the degree of the root: the root is of degree i, j and neither i nor j, and we respectively let $a_i(x,u)$, $a_j(x,u)$ and $a_0(x,u)$ be the generating functions (or a_i, a_j, a_0 for short). It is easy to see that

$$a_0(x,u) + a_i(x,u) + a_j(x,u) = p(x,u). \tag{8}$$

In what follows, there appears an expression of the form $Z(S_n, f(x,u))$ (or $f(x)$ for $f(x,u)$), 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 [6] for details.

Employing the classic Pólya Enumeration Theorem, we have $Z(S_{k-1}; p(x))$ as the counting series of the planted trees whose roots have degree k, and the coefficient of x^p in $Z(S_{k-1}; p(x))$ is the number of planted trees of order $p + 1$ (see [6] p.51–54). Therefore, $p(x)$ satisfies

$$p(x) = x \sum_{k \geq 0} Z(S_k; p(x)) = x \ \mathrm{e}^{\sum_{k \geq 1} \frac{1}{k} p(x^k)}. \tag{9}$$

By the same way, we can obtain the following functional equations

$$a_0(x,u) = x \ \mathrm{e}^{\sum_{k \geq 1} \frac{1}{k} p(x^k, u^k)} - xZ(S_{i-1}; p(x,u)) - xZ(S_{j-1}; p(x,u)), \tag{10}$$

$$a_i(x,u) = x \sum_{\ell_1 + \ell_2 = i-1} Z(S_{\ell_1}; a_0(x,u) + a_i(x,u)) \cdot Z(S_{\ell_2}; a_j(x,u)) u^{\ell_2}, \tag{11}$$

$$a_j(x,u) = x \sum_{m_1 + m_2 = j-1} Z(S_{m_1}; a_0(x,u) + a_j(x,u)) \cdot Z(S_{m_2}; a_i(x,u)) u^{m_2}. \tag{12}$$

For $a_0(x,u)$, since the degrees of the roots are neither i nor j, therefore there are two minor modifications in Equ.(10). For $a_i(x,u)$, if there exist ℓ_2 vertices of degree j adjacent to the root, we should count ℓ_2 occurrences of (i,j) in addition, and thus it is of $Z(S_{\ell_1}; a_0(x,u) + a_i(x,u)) \cdot Z(S_{\ell_2}; a_j(x,u)) u^{\ell_2}$. Analogously, we can get Equ.(12) for $a_j(x,u)$.

Then, for rooted trees, we have

$$\begin{aligned} r(x,u) &= x\,\mathrm{e}^{\sum_{k\ge 1}\frac{1}{k}p(x^k,u^k)} - xZ(S_i;p(x,u)) - xZ(S_j;p(x,u)) \\ &\quad + x\sum_{\ell_1+\ell_2=i} Z(S_{\ell_1};a_0(x,u)+a_i(x,u))\cdot Z(S_{\ell_2};a_j(x,u))u^{\ell_2} \\ &\quad + x\sum_{m_1+m_2=j} Z(S_{m_1};a_0(x,u)+a_j(x,u))\cdot Z(S_{m_2};a_i(x,u))u^{m_2}. \end{aligned}$$

In order to get the generating function for general trees, we need the following lemma, Lemma 2, which was used in [10] to get the famous equation

$$t(x) = r(x) - \frac{1}{2}p(x)^2 + \frac{1}{2}p(x^2). \tag{13}$$

We can also get a similar equation for $t(x,u)$ from Lemma 2.

First, let us recall some terminology. 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 of the trees 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*. Now we can state the lemma.

Lemma 2. *[10] For any tree, the number of rooted trees corresponding to this tree minus the number of nonsimilar edges (except the symmetric edge) is exactly* 1.

Note that, if we delete any edge of a set of similar edges in a tree, the yielded trees are the same two trees. Hence, different pairs of planted trees correspond to nonsimilar edges. We refer the readers to [10] for details. Then, analogous to (13), we have

$$t(x,u) = r(x,u) - \frac{1}{2}p(x,u)^2 + \frac{1}{2}p(x^2,u^2) + a_i(x,u)\cdot a_j(x,u)(1-u). \tag{14}$$

The last term serves to count the occurrences of (i,j) when joining two planted trees to form a tree, in which one has a root of degree i and the other of degree j. Moreover, from Equ.(13) it follows that

$$t(x_0,1) = (1+r(x_0^2))/2 = c_0.$$

Note that $x_0 < 1$, and thus x_0^2 is surely inside the region of the convergence of $r(x)$.

First, we shall use Lemma 1 to get the expression of (7) for a_0, a_i and a_j. However, we only need to verify that the system of functions Equs.(10), (11) and (12) satisfies Equ.(5), since the other conditions are easy to verify. We still denote this system of functions by $\mathbf{F}$. It is a function of vector $\mathbf{a}(x,u) = (a_0(x,u), a_i(x,u), a_j(x,u))^T$. Combining the fact that the partial derivative enjoys (see [13])

$$Z_{s_1}(S_n;s_1,\ldots,s_n) = Z(S_{n-1};s_1,\ldots,s_{n-1}),$$

with Equ.(2), we obtain that

$$\begin{aligned} \mathbf{F}_{a_0}(x_0,\mathbf{a}(x_0,1),1) &= \begin{pmatrix} 1 - x_0Z(S_{i-2};p(x_0,1)) - x_0Z(S_{j-2};p(x_0,1)) \\ x_0\sum_{\ell_1+\ell_2=i-1} Z(S_{\ell_1-1};a_0(x_0,1)+a_i(x_0,1))Z(S_{\ell_2};a_j(x_0,1)) \\ x_0\sum_{m_1+m_2=j-1} Z(S_{m_1-1};a_0(x_0,1)+a_j(x_0,1))Z(S_{m_2};a_i(x_0,1)) \end{pmatrix} \\ &= \begin{pmatrix} 1 - x_0Z(S_{i-2};p(x_0,1)) - x_0Z(S_{j-2};p(x_0,1)) \\ x_0Z(S_{i-2};p(x_0,1)) \\ x_0Z(S_{j-2};p(x_0,1)) \end{pmatrix}. \end{aligned}$$

Similarly, we can get that $\mathbf{F}_{a_i}(x_0, \mathbf{a}(x_0,1),1) = \mathbf{F}_{a_j}(x_0, \mathbf{a}(x_0,1),1) = \mathbf{F}_{a_0}(x_0, \mathbf{a}(x_0,1),1)$. Therefore, one can readily see that

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

So, for the system of generating functions a_0, a_i and a_j, all the conditions required by Lemma 1 are satisfied. Hence, we can suppose

$$\begin{aligned} a_0 &= g_0(u,x) - h_0(x,u)\sqrt{1 - \frac{x}{x(u)}}, \\ a_i &= g_i(u,x) - h_i(x,u)\sqrt{1 - \frac{x}{x(u)}}, \\ a_j &= g_j(u,x) - h_j(x,u)\sqrt{1 - \frac{x}{x(u)}}, \end{aligned}$$

such that all the corresponding functions satisfy the conditions of Lemma 1.

In what follows, we shall show that $t(x,u)$ is in the form of Equ.(15), and then use the following lemma, due to [7, 13], to get the final result.

Lemma 3. *[7, 13] Suppose that $t(x,u)$ has the form*

$$t(x,u) = \overline{g}(x,u) - \overline{h}(x,u)\left(1 - \frac{x}{x(u)}\right)^{3/2} \tag{15}$$

where $\overline{g}(x,u)$, $\overline{h}(x,u)$ and $x(u)$ are analytic functions around $x = x(1)$ and $u = 1$ that satisfy $x(1) > 0$ and $x_u(1) < 0$, and $\overline{h}(x(1),1) \neq 0$. Moreover, $t(x,u)$ is analytically continued around $x = x(u)$, $u = 1$ with $\arg(x - x(u)) \neq 0$. Suppose that X_n is defined as Equ.(1) corresponding to $t(x,u)$. Then, $E(X_n) = (\mu + o(1))n$ and $Var(X_n) = (\sigma + o(1))n$, where $\mu = -x_u(1)/x(1)$ and $\sigma = \mu^2 + \mu - x_{uu}(1)/x(1)$.

By the Equs.(10) through (12) and the expression of $r(x,u)$, the generating function $t(x,u)$ can be rewritten as

$$\begin{aligned} t(x,u) = {} & p(x,u) + x \cdot Z(S_{i-1};p) + x \cdot Z(S_{j-1};p) - x \cdot Z(S_i;p) - x \cdot Z(S_j;p) \\ & - x \sum_{\ell_1+\ell_2=i-1} Z(S_{\ell_1}; a_0 + a_i) \cdot Z(S_{\ell_2}; a_j) u^{\ell_2} - x \sum_{m_1+m_2=j-1} Z(S_{\ell_1}; a_0 + a_j) \cdot Z(S_{m_2}; a_i) u^{m_2} \\ & + x \sum_{\ell_1+\ell_2=i} Z(S_{\ell_1}; a_0 + a_i) \cdot Z(S_{\ell_2}; a_j) u^{\ell_2} + x \sum_{m_1+m_2=j} Z(S_{\ell_1}; a_0 + a_j) \cdot Z(S_{m_2}; a_i) u^{m_2} \\ & - \frac{1}{2} p(x,u)^2 + \frac{1}{2} p(x^2,u^2) + a_i a_j \cdot (1-u). \end{aligned}$$

By means of Taylor's theorem we get

$$Z(S_k; g - h\sqrt{1 - x/x(u)}) = \sum_{r=0}^{k} Z_{s_1}^{(r)}(S_k; g) h^r (1 - x/x(u))^{r/2} \frac{(-1)^r}{r!},$$

where $Z_{s_1}^{(r)}$ denotes the rth derivative of the first variable s_1 of the cycle index of S_k. Consequently, we have

$$
\begin{aligned}
t(x,u) = G(x,u) + \sqrt{1-\frac{x}{x(u)}}\Big\{ &-(h_0+h_i+h_j) + x\cdot Z(S_{i-2}; g_0+g_i+g_j)(h_0+h_i+h_j)(-1) \\
&+ x\cdot Z(S_{j-2}; g_0+g_i+g_j)(h_0+h_i+h_j)(-1) \\
&- x\cdot Z(S_{i-1}; g_0+g_i+g_j)(h_0+h_i+h_j)(-1) - x\cdot Z(S_{j-1}; g_0+g_i+g_j)(h_0+h_i+h_j)(-1) \\
&+ x \sum_{\ell_1+\ell_2=i-1} [Z(S_{\ell_1-1}; g_0+g_i)\cdot Z(S_{\ell_2}; g_j)(h_0+h_i)u^{\ell_2} + Z(S_{\ell_1}; g_0+g_i)\cdot Z(S_{\ell_2-1}; g_j)h_j u^{\ell_2}] \\
&+ x \sum_{m_1+m_2=j-1} [Z(S_{m_1-1}; g_0+g_j)\cdot Z(S_{m_2}; g_i)(h_0+h_j)u^{m_2} + Z(S_{m_1}; g_0+g_j)\cdot Z(S_{m_2-1}; g_i)h_i u^{m_2}] \\
&- x \sum_{\ell_1+\ell_2=i} [Z(S_{\ell_1-1}; g_0+g_i)\cdot Z(S_{\ell_2}; g_j)(h_0+h_i)u^{\ell_2} + Z(S_{\ell_1}; g_0+g_i)\cdot Z(S_{\ell_2-1}; g_j)h_j u^{\ell_2}] \\
&- x \sum_{m_1+m_2=j} [Z(S_{m_1-1}; g_0+g_j)\cdot Z(S_{m_2}; g_i)(h_0+h_j)u^{m_2} + Z(S_{m_1}; g_0+g_j)\cdot Z(S_{m_2-1}; g_i)h_i u^{m_2}] \\
&+ (g_0+g_i+g_j)(h_0+h_i+h_j) + (u-1)(g_ih_j+g_jh_i) - H(x,u)\left(1-\frac{x}{x(u)}\right)\Big\} \\
:= G(x,u) + h(x,u)&\sqrt{1-\frac{x}{x(u)}} - H(x,u)\left(1-\frac{x}{x(u)}\right)^{3/2},
\end{aligned}
$$

where we use $G(x,u)$, $h(x,u)$ and $H(x,u)$ to denote some functions analytic around $x = x(1)$ and $u = 1$. Note that $\frac{1}{2}p(x^2,u^2)$ is contained in $G(x,u)$. Then, we try to show $h(x,u) \equiv 0$ around $x = x(1)$ and $u = 1$.

Recall that $x(u)$, $\mathbf{a}(x(u),u)$ is the solution of the system of functions

$$\mathbf{a} = \mathbf{F}(x,\mathbf{a},u), 0 = \det(\mathbf{I} - \mathbf{F_a}(x,\mathbf{a},u)),$$

that is,

$$g_0 = x\, \mathrm{e}^{\sum_{k\ge1}\frac{1}{k}p(x^k,u^k)} - xZ(S_{i-1}; g_0+g_i+g_j) - xZ(S_{j-1}; g_0+g_i+g_j), \tag{16}$$

$$g_i = x \sum_{\ell_1+\ell_2=i-1} Z(S_{\ell_1}; g_0+g_i)\cdot Z(S_{\ell_2}; g_j)u^{\ell_2}, \tag{17}$$

$$g_j = x \sum_{m_1+m_2=j-1} Z(S_{m_1}; g_0+g_j)\cdot Z(S_{m_2}; g_i)u^{m_2}, \tag{18}$$

and

$$\begin{vmatrix} A-1 & A & A \\ B_1 & B_1-1 & B_2 \\ C_1 & C_2 & C_1-1 \end{vmatrix} = 0, \tag{19}$$

where

$$\begin{aligned}
A &:= x\, \mathrm{e}^{\sum_{k\geq 1}\frac{1}{k}p(x^k,u^k)} - xZ(S_{i-2}; g_0+g_i+g_j) - xZ(S_{j-2}; g_0+g_i+g_j),\\
B_1 &:= x \sum_{\ell_1+\ell_2=i-1} Z(S_{\ell_1-1}; g_0+g_i)\cdot Z(S_{\ell_2}; g_j)u^{\ell_2},\\
B_2 &:= x \sum_{\ell_1+\ell_2=i-1} Z(S_{\ell_1}; g_0+g_i)\cdot Z(S_{\ell_2-1}; g_j)u^{\ell_2},\\
C_1 &:= x \sum_{m_1+m_2=j-1} Z(S_{m_1-1}; g_0+g_j)\cdot Z(S_{m_2}; g_i)u^{m_2},\\
C_2 &:= x \sum_{m_1+m_2=j-1} Z(S_{m_1}; g_0+g_j)\cdot Z(S_{m_2-1}; g_i)u^{m_2}.
\end{aligned}$$

Therefore, for some constants α, β and γ, not all zeros, we have

$$\alpha\begin{pmatrix}A-1\\ A\\ A\end{pmatrix} + \beta\begin{pmatrix}B_1\\ B_1-1\\ B_2\end{pmatrix} + \gamma\begin{pmatrix}C_1\\ C_2\\ C_1-1\end{pmatrix} = \mathbf{0}.$$

Let $u=1$ and $x=x_0$. Recalling the expression of $\mathbf{F}_{\mathbf{a}}(x_0,\mathbf{a}(x_0,1),1)$, we have $\alpha=\beta=\gamma=1$. Thus, it follows that

$$B_1+C_1 = 1-A = B_1+C_2 = B_2+C_1. \tag{20}$$

In conjunction with Equs.(16) through (18) as well as (20), the expression of $h(x,u)$ can be simplified as

$$\begin{aligned}
h(x,u) &= -(h_0+h_i+h_j) - x\cdot Z(S_{i-2}; g_0+g_i+g_j)(h_0+h_i+h_j)\\
&- x\cdot Z(S_{j-2}; g_0+g_i+g_j)(h_0+h_i+h_j)\\
&+ x\cdot Z(S_{i-1}; g_0+g_i+g_j)(h_0+h_i+h_j) + x\cdot Z(S_{j-1}; g_0+g_i+g_j)(h_0+h_i+h_j)\\
&+ (h_0+h_i)B_1 + h_jB_2 + (h_0+h_j)C_1 + h_iC_2 - (h_0+h_i)g_i - uh_jg_i - (h_0+h_j)g_j - uh_ig_j\\
&+ (g_0+g_i+g_j)(h_0+h_i+h_j) + (u-1)(g_ih_j+g_jh_i).
\end{aligned}$$

Furthermore, it follows that

$$\begin{aligned}
h(x,u) &= (h_0+h_i+h_j)[-1 - x\cdot Z(S_{i-2}; g_0+g_i+g_j) - x\cdot Z(S_{j-2}; g_0+g_i+g_j)\\
&+ x\cdot Z(S_{i-1}; g_0+g_i+g_j) + x\cdot Z(S_{j-1}; g_0+g_i+g_j) + (1-A) + g_0]\\
&= (h_0+h_i+h_j)[-x\cdot Z(S_{i-2}; g_0+g_i+g_j) - x\cdot Z(S_{j-2}; g_0+g_i+g_j)\\
&+ x\cdot Z(S_{i-1}; g_0+g_i+g_j) + x\cdot Z(S_{j-1}; g_0+g_i+g_j) - g_0 - x\cdot Z(S_{i-1}; g_0+g_i+g_j)\\
&- x\cdot Z(S_{j-1}; g_0+g_i+g_j) + x\cdot Z(S_{i-2}; g_0+g_i+g_j) + x\cdot Z(S_{j-2}; g_0+g_i+g_j) + g_0]\\
&\equiv 0.
\end{aligned}$$

Therefore, $t(x,u)$ is in the form of

$$t(x,u) = G(x,u) - H(x,u)\left(1-\frac{x}{x(u)}\right)^{3/2},$$

and G and H are analytic around $x=x(1)$, $u=1$. Recalling that $x=x(1)=x_0$ and $u=1$, therefore $t(x,u)$ has the expression of Equ.(3), that is, $H(x,u)\neq 0$ around $x=x(1)$ and $u=1$.

Moreover, suppose that $\mathbf{v}^T$ is a vector satisfying $\mathbf{v}^T(\mathbf{I}-\mathbf{F}_{\mathbf{y}}(x_0,\mathbf{y}_0,1))=0$. It has been shown that for the system of functions, we have

$$\mu=\frac{-x_u(1)}{x(1)}=\frac{1}{x_0}\frac{\mathbf{v}^T\mathbf{F}_u(x_0,\mathbf{y}_0,1)}{\mathbf{v}^T\mathbf{F}_x(x_0,\mathbf{y}_0,1)}. \tag{21}$$

From Remark 1, we know that $\mathbf{v}$ is unique up to a nonzero factor. We refer the readers to [3] for more details. It is easy to find that $\mu>0$ and $x_u(1)<0$ for $\mathbf{a}(x,u)$. And, since $t(x,u)$ has the form of Equ.(14), one can readily see that $t(x,u)$ is analytically continued around $x=x(u)$, $u=1$ with $arg(x-x(u))\neq 0$. Thus, all the conditions in Lemma 3 hold. Consequently, we get that the number X_n of occurrences of the pattern (i,j) $(i\neq j>1)$ is asymptotically normally distributed.

In what follows, we will do further calculation on the mean value. From the expression of $\mathbf{F}_{\mathbf{a}}(x_0,\mathbf{a}(x_0,1),1)$, it is not difficult to obtain that $\mathbf{v}^T=(1,1,1)$ is a basic solution. We will compute $\mathbf{v}^T\mathbf{F}_x(x_0,\mathbf{a}(x_0,1),1)$ and $\mathbf{v}^T\mathbf{F}_u(x_0,\mathbf{a}(x_0,1),1)$ to estimate μ, which would be more brief than just to do with $\mathbf{F}_x(x_0,\mathbf{a}(x_0,1),1)$ and $\mathbf{F}_u(x_0,\mathbf{a}(x_0,1),1)$. Then, we have

$$\mathbf{v}^T\mathbf{F}_x(x_0,\mathbf{a}(x_0,1),1)=\frac{1}{x_0}+\sum_{k=2}p_x(x_0^k,1)x_0^{k-1}, \tag{22}$$

$$\begin{aligned}\mathbf{v}^T\mathbf{F}_u(x_0,\mathbf{a}(x_0,1),1)=&\sum_{k=2}p_u(x_0^k,1)\\&+x_0\sum_{\ell_1+\ell_2=i-1}Z(S_{\ell_1};a_0(x_0,1)+a_i(x_0,1))\cdot Z(S_{\ell_2};a_j(x_0,1))\cdot\ell_2\\&+x_0\sum_{m_1+m_2=j-1}Z(S_{m_1};a_0(x_0,1)+a_j(x_0,1))\cdot Z(S_{m_2};a_i(x_0,1))\cdot m_2.\end{aligned} \tag{23}$$

In view of $p(x,1)=p(x)=r(x)$, combining it with Equs.(2) and (9), it follows that

$$\frac{1}{x_0}+\sum_{k=2}p_x(x_0^k,1)x_0^{k-1}=\frac{p_x(x,1)(1-p(x,1))}{p(x,1)}\Big|_{x=x_0}=b^2/2,$$

and thus

$$\mathbf{v}^T\mathbf{F}_x(x_0,\mathbf{a}(x_0,1),1)=\frac{b^2}{2}.$$

However, we failed to do any further simplification for Equ.(23). For convenience, denote the value of $\mathbf{v}^T\mathbf{F}_u(x_0,\mathbf{a}(x_0,1),1)$ by $w(i,j)$. One can use a computer to get an approximate value of it. Thus,

$$\mu=\frac{2}{x_0b^2}w(i,j).$$

Case 2. $i=1$, $j>1$.

We proceed to obtain the result in a same way as in Case 1. We still use the same notation. But notice that when we split up $\mathcal{P}_n$ according to the degrees of the roots, there exists only one planted tree with root of degree 1, i.e., the tree with only two nodes. Thus, we have

$$x+a_0(x,u)+a_j(x,u)=p(x,u),$$

and the system of functions is as follows

$$a_0(x,u)=x\,\mathrm{e}^{\sum_{k\geq 1}\frac{1}{k}p(x^k,u^k)}-x-xZ(S_{j-1};p(x,u)), \tag{24}$$

$$a_j(x,u)=x\sum_{m_1+m_2=j-1}Z(S_{m_1};p(x,u)-x)x^{m_2}u^{m_2}. \tag{25}$$

The same as previous, we can establish the generating functions for rooted trees

$$r(x,u) = x\ \mathrm{e}^{\sum_{k\geq 1}\frac{1}{k}p(x^k,u^k)} - xa_j(x,u)(1-u)$$
$$- x\sum_{m_1+m_2=j} Z(S_{m_1}; p(x,u)-x)\cdot Z(S_{m_2}; x)(1-u^{m_2}),$$

and for general trees

$$t(x,u) = r(x,u) - \frac{1}{2}p(x,u)^2 + \frac{1}{2}p(x^2,u^2) + xa_j(x,u)(1-u). \tag{26}$$

It is not difficult to verify that Equs.(24) and (25) satisfy the conditions of Lemma 1. Analogous to the Case 1, we can get that X_n is also asymptotically normally distributed. Moreover, we obtain that $\mathbf{v}^T = (1,1)$,

$$\mathbf{v}^T\mathbf{F}_x(x_0, \mathbf{a}(x_0,1),1)$$
$$= \left\{x\ \mathrm{e}^{\sum_{k\geq 1}\frac{1}{k}p(x^k,u^k)}\left(1+\sum_{k\geq 2}p_x(x^k,u^k)x^{k-1}\right) + \ \mathrm{e}^{\sum_{k\geq 1}\frac{1}{k}p(x^k,u^k)} - 1\right\}\Big|_{(x=x_0,u=1)}$$
$$= \frac{b^2}{2},$$

and

$$\mathbf{v}^T\mathbf{F}_u(x_0,\mathbf{a}(x_0,1),1) = \sum_{k\geq 2}p_u(x_0^k,1) + x_0\sum_{\ell_1+\ell_2=j-1}Z(S_{\ell_1}; p(x_0,1)-x_0)x_0^{\ell_2}\cdot \ell_2.$$

Again, for convenience, we denote $\mathbf{v}^T\mathbf{F}_u(x_0,\mathbf{a}(x_0,1),1)$ by $w(1,j)$. Then, it follows that

$$\mu = \frac{2}{x_0 b^2}w(1,j).$$

Case 3. $i = j > 1$.

Since the procedure is the same as previous, we omit the details of the proof. However, we still use the same notations here without any conflicts.

$$a_0(x,u) + a_j(x,u) = p(x,u),$$
$$a_0(x,u) = x\ \mathrm{e}^{\sum_{k\geq 1}\frac{1}{k}p(x^k,u^k)} - xZ(S_{j-1}; p(x,u)),$$
$$a_j(x,u) = x\sum_{m_1+m_2=j-1}Z(S_{m_1}; a_0(x,u))\cdot Z(S_{m_2}; a_j(x,u))u^{m_2}.$$

For general trees, we have

$$t(x,u) = r(x,u) - \frac{1}{2}p(x,u)^2 + \frac{1}{2}p(x^2,u^2)$$
$$+ \frac{1}{2}a_j(x,u)\cdot a_j(x,u)(1-u) - \frac{1}{2}a_j(x^2,u^2)(1-u).$$

Analogously, we can also get that X_n is asymptotically normally distributed for this case.

Further, we have $\mathbf{v}^T = (1,1)$, $\mathbf{v}^T\mathbf{F}_x(x_0,\mathbf{a}(x_0,1),1) = b^2/2$ and

$$\mathbf{v}^T\mathbf{F}_u(x_0,\mathbf{a}(x_0,1),1)$$
$$= \sum_{k\geq 2}p_u(x_0^k,1) + x_0\sum_{m_1+m_2=j-1}Z(S_{m_1}; a_0(x_0,1))\cdot Z(S_{m_2}; a_j(x_0,1))\cdot m_2.$$

Then, we obtain that

$$\mu = \frac{2}{x_0 b^2} w(j,j),$$

where $w(j,j)$ denotes the value of $\mathbf{v}^T \mathbf{F}_u(x_0, \mathbf{a}(x_0,1),1)$.

As a conclusion, we can establish the following theorem now.

Theorem 4. *Suppose that X_n is the random variable corresponding to the occurrences of pattern (i,j). The probability of X_n is defined as Equ.(1) for the generating function $t(x,u)$ of trees. Then, the distribution of X_n is asymptotically normal with mean*

$$E(X_n) = \left(\frac{2}{x_0 b^2} \cdot w(i,j) + o(1)\right) n$$

and variance $Var(X_n) = (\sigma(i,j) + o(1))n$, where $w(i,j)$ and $\sigma(i,j)$ are some constants.

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 $\Pr(Q)$ in $\mathcal{G}_n$ converges to 1 as n tends to infinity. Occasionally, we will say *almost all* instead of almost every.

From the above theorem and employing Chebyshev inequality, it is easy to see that

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

Thus, for almost all trees in $\mathcal{T}_n$, X_n equals $\left(\frac{2}{x_0 b^2} \cdot w(i,j) + o(1)\right) n$. Consequently, one can get the following corollary.

Corollary 5. *For almost all trees, the number of occurrences of the pattern (i,j) is $\left(\frac{2}{x_0 b^2} \cdot w(i,j) + o(1)\right) n$.*

3 An application

In this section, we will use Corollary 5 to investigate the values of the *Randić index* and the *general Randić index*, and show that Conjecture 1 is true for almost all trees.

Let $G = (V,E)$ be a graph with vertex set V and edge set E. The Randić index is defined as

$$R(G) = \sum_{uv \in E} \frac{1}{\sqrt{d_u d_v}},$$

where d_u, d_v are the degrees of the vertices $u, v \in V$, respectively.

We know that the number of occurrences of the pattern (i,j) is the number of edges with one end of degree i and the other of degree j in the tree. Still, we assume $i \leq j$. Then, the number of this kind of edges in almost all trees of $\mathcal{T}_n$ is $\left(\frac{2}{x_0 b^2} \cdot w(i,j) + o(1)\right) n$. Moreover, every tree in T_n has $n-1$ edges. So, for any integer K, $\sum_{i \leq j \leq K} \frac{2}{x_0 b^2} \cdot w(i,j) \leq 1$, it follows that $\sum_{i \leq j} \frac{2}{x_0 b^2} \cdot w(i,j)$ is convergent. Consequently, $\sum_{i \leq j} \frac{2}{x_0 b^2 \sqrt{i \cdot j}} \cdot w(i,j)$ also converges to some constant λ. Although the exact value of λ can not be given, one can employ a computer to get that $0.1 < \lambda < 1/2$ (the upper bound can be seen from the fact that path attains the maximal value of the Randić index). Then, for any $\varepsilon > 0$, there exists an integer K_0 such that for any $K \geq K_0$,

$$\sum_{i \leq j, j \geq K} \frac{2}{x_0 b^2} \cdot w(i,j) < \varepsilon,$$

that is, for almost all trees, the number of edges with one end of degree larger than K is less than εn. Hence, the Randić index enjoys

$$\Big(\sum_{i\le j\le K}\frac{2}{x_0b^2\sqrt{i\cdot j}}w(i,j)+o(1)\Big)n<R(T_n)<\Big(\sum_{i\le j\le K}\frac{2}{x_0b^2\sqrt{i\cdot j}}w(i,j)+o(1)\Big)n+\varepsilon n \text{ a.e.}$$

Immediately, one can get the following result.

Theorem 6. *For any $\varepsilon>0$, the Randić index of almost all trees enjoys*

$$(\lambda-\varepsilon)n<R(T_n)<(\lambda+\varepsilon)n. \tag{27}$$

Bollobás and Erdös [2] generalized the Randić index as

$$R_\alpha(G)=\sum_{uv\in E}(d_ud_v)^\alpha,$$

which is called the *general Randić index*, where α is a real number. Clearly, if $\alpha=-\frac{1}{2}$, then $R_{-\frac{1}{2}}(G)=R(G)$. We refer the readers to a survey [9] for more details on this index. Here, we suppose $\alpha<0$. Following the sketch of obtaining Equ.(27), we can analogously get an estimate of $R_\alpha(T_n)$. Then, one can get the following corollary.

Corollary 7. *Suppose $\alpha<0$. Then, for any $\varepsilon>0$ we have*

$$(\lambda_\alpha-\varepsilon)n\le R_\alpha(T_n)\le(\lambda_\alpha+\varepsilon)n \text{ a.e.},$$

where λ_α is some constant depending on α.

In what follows, we will consider Conjecture 1. Let $d(u,v)$ be the distance between vertices $u,v\in V$. The *average distance* of a graph G is defined as the average value of the distances between all pairs of vertices of G, i.e.,

$$D(G)=\frac{\sum_{u,v\in V}d(u,v)}{\binom{n}{2}}.$$

We will show that Conjecture 1 is true for almost all trees. To this end, we first introduce the concept of the *Wiener index* of a graph G, which is defined as

$$W(G)=\sum_{u,v\in V}d(u,v).$$

Clearly, $W(G)=\binom{n}{2}D(G)$. Regarding $W(T_n)$ as a random variable on $\mathcal{T}_n$, Wagner [14] established the following result.

Lemma 8. *The Wiener index $W(T_n)$ enjoys*

$$E(W(T_n))=(\omega+o(1))n^{5/2}$$

and

$$Var(W(T_n))=(\delta+o(1))n^5,$$

where ω and δ are some constants.

Employing Chebyshev inequality, from Lemma 8 we have

$$\Pr[|W(T_n) - E(W(T_n))| \geq n^{11/4}] \leq \frac{Var(W(T_n))}{n^{11/2}} \to 0, \text{ as } n \to \infty.$$

Since $E(W(T_n)) = O(n^{5/2})$, therefore for almost all trees the Wiener index $W(T_n)$ is $O(n^{11/4})$. Consequently, we can get that the average distance satisfies

$$D(T_n) = O(n^{3/4}) \text{ a.e.}$$

Combining this with Equ.(27), the following result is an immediate consequence.

Theorem 9. *For almost all trees in $\mathcal{T}_n$, $R(T_n) > D(T_n)$.*

Remark 2. *Recall the classic Erdös–Rényi model $\mathcal{G}_{n,p}$ of random graphs [1], which consists of all graphs $G_{n,p}$ with vertex set $[n] = \{1, 2, \ldots, n\}$ in which the edges are chosen independently with probability $0 < p < 1$. We can easily get the same result for the Erdös–Rényi model of random graphs. In fact, suppose that p is a constant. Recall that for almost all graphs the degree of a vertex is $(p + o(1))n$ (see [8]). Thus, for almost all graphs,*

$$R(G_{n,p}) = \frac{1}{2} \cdot \frac{1}{\sqrt{(p + o(1))^2 n^2}} \cdot (p + o(1))n \cdot n = \big(\frac{1}{2} + o(1)\big)n.$$

Moreover, it is well known that the diameter is not more than 2 for almost all graphs. Consequently, $D(G_{n,p}) \leq 2$ a.e.. Hence,

$$R(G_{n,p}) > D(G_{n,p}) \text{ a.e.}$$

Acknowledgement: The authors would like to thank Dr. Gerard Kok for helps. They are also very grateful to the editor and the referees for helpful comments and suggestions.

References

[1] B. Bollobás, Random Graphs, 2nd edition, Cambridge University Press, Cambridge, 2001.

[2] B. Bollobás, P. Erdős, Graphs of extremal weights, Ars Combin. 50 (1998) 225–233.

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

[4] M. Drmota, Systems of functional equations, Random Structures Algorithms 10 (1997) 103-124.

[5] S. Fajtlowicz, On conjectures of Graffiti, Discrete Math. 72 (1988) 113–118.

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

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

[8] G. Ivchenko, On the asymptotic behaviour of degrees of vertices in a random graph, Theory Probab. Appl. 18 (1973) 188–155.

[9] X. Li, Y. Shi, A survey on the Randić index, MATCH Commun. Math. Comput. Chem. 59 (2008) 127–156.

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

[11] M. Randić, On characterization of molecular branching, J. Amer. Chem. Soc. 97 (1975) 6609–6615.

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

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

[14] S.G. Wagner, On the average Wiener index of degree-restricted trees, Australas. J. Combin. 37 (2007) 187–204.