The expected subtree number index in random polyphenylene and spiro chains

Yu Yang[a], Xiao-Jun Sun[b], Jia-Yi Cao[c], Hua Wang[d], Xiao-Dong Zhang[e,*]

[a]*School of Computer Science, Pingdingshan University, Pingdingshan 467000, China*
[b]*School of Electrical and Mechanical Engineering, Pingdingshan University, Pingdingshan 467000, China*
[c]*Navigation College, Dalian Maritime University, Dalian, 116026, China*
[d]*Department of Mathematical Sciences, Georgia Southern University, Statesboro, GA 30460, USA*
[e]*School of Mathematical Sciences, MOE-LSC, SHL-MAC, Shanghai Jiao Tong University, Shanghai, 200240, China*

Abstract

The subtree number index $STN(G)$ of a simple graph G is the number of nonempty subtrees of G. It is a structural and counting topological index that has received more and more attention in recent years. In this paper we first obtain exact formulas for the expected values of subtree number index of random polyphenylene and spiro chains, which are molecular graphs of a class of unbranched multispiro molecules and polycyclic aromatic hydrocarbons. Moreover, we establish a relation between the expected values of the subtree number indices of a random polyphenylene and its corresponding hexagonal squeeze. We also present the average values for subtree number indices with respect to the set of all polyphenylene and spiro chains with n hexagons.

Keywords: Subtree number index, Random polyphenylene chain, Random spiro chain, Expected value, Average value
MSC[2020] 05C80, 05C05

1. Introduction

The subtree number index $STN(G)$ of a graph G is a structure-based index, defined as the total number of non-empty subtrees of G. It is discovered to have applications in the design of reliable communication network [21], bioinformatics [11], and characterizing physicochemical and structural properties of molecular graphs [13, 26, 25]. In recent years there have been related works on enumerating subtrees [22, 15, 3, 2, 28], characterizing extremal graphs and values [16, 29, 10,

---

*Corresponding author

Email addresses: yangyu@pdsu.edu.cn (Yu Yang), pdschris@pdsu.edu.cn (Xiao-Jun Sun), jiayicao0219@163.com (Jia-Yi Cao), hwang@georgiasouthern.edu (Hua Wang), xiaodong@sjtu.edu.cn (Xiao-Dong Zhang)

30], analyzing relations with other topological indices such as the Wiener index [26, 25, 17, 19], average order and density of subtrees [18, 9, 6].

Polyphenylenes, spiro compounds and their derivatives are important polycyclic aromatic hydrocarbons in organic chemistry and have many applications in industry including organic synthesis, drug synthesis, heat exchangers, etc. For more details one may see [4, 5, 1, 12] and the references cited therein.

Regarding topological indices of random polyphenylene and spiro chains, Yang and Zhang [23] found the expected value of the Wiener index of a random polyphenylene chain. Huang, Kuang and Deng [7] obtained the expected values of the Kirchhoff index of random polyphenyl and spiro chains. Subsequently, Huang, Kuang and Deng [8] presented explicit formulas for the expected values of the Hosoya index and the Merrifield-Simmons index of a random polyphenylene chain. More recently, Liu [20] presented explicit formulas for the expected values of ABC and GA indices in random spiro chains and compare the expected values of these two indices. Zhang, Li, Li, and Zhang [27] established explicit analytical expressions for the expected values of the Schultz index, Gutman index, multiplicative degree-Kirchhoff index and additive degree-Kirchhoff index of a random polyphenylene chain.

As far as the subtree number is concerned, there is no mathematical or computational studies on these two random chains. In this paper, we fill in the gap by studying the subtree number index of the random polyphenylene and spiro chains.

The rest of the paper is organized as follows. Section 2 contains the necessary definitions and lemmas. In Section 3, we provide the expected value of the subtree number index of random polyphenylene and spiro chains, and a relation between the subtree number of these two random chains. Lastly, we briefly discuss the average value of the subtree number index of the polyphenylene and spiro chains in Section 4.

2. Preliminaries

We first introduce the technical notations and lemmas that will be used in the discussion. For more background information one may check [26, 22, 20, 27].

Let $G = (V(G), E(G); f, g)$ be a weighted graph on n vertices and m edges, with vertex-weight function $f : V(G) \to \Re$ and edge-weight function $g : E(G) \to \Re$ (where $\Re$ is a commutative ring with a unit element 1). Denote by $\mathcal{ST}(G)$ the set of all nonempty subtrees of G. Given vertex subset $V_S \subseteq V(G)$ and edge subset $E_S \subseteq E(G)$, denote by $\mathcal{ST}(G, V_S)$, $\mathcal{ST}(G, E_S)$ the set of subtrees containing V_S, E_S respectively.

For a given subtree $T \in \mathcal{ST}(G)$, its *weight* is defined as

$$\omega(T) = \prod_{v \in V(T),\ e \in E(T)} f(v)g(e).$$

2

And we define the *subtree generating function* of G by

$$F(G; f, g) = \sum_{T \in \mathcal{ST}(G)} \omega(T).$$

Similarly, the *subtree generating function of G containing V_S, E_S* are as follows:

$$F(G; f, g; V_S) = \sum_{T \in \mathcal{ST}(G, V_S)} \omega(T),$$

$$F(G; f, g; E_S) = \sum_{T \in \mathcal{ST}(G, E_S)} \omega(T).$$

Letting $\eta(\cdot)$ be the number of subtrees in set $\mathcal{ST}(\cdot)$, we have

$$\eta(G) = F(G; 1, 1)$$

and

$$\eta(G, V_S) = F(G; 1, 1; V_S), \quad \eta(G, E_S) = F(G; 1, 1; E_S).$$

Through introducing the subtree weight and using generating function, Yan and Yeh [22] developed algorithms for counting the subtrees under various constrains. We summarize their approach as follows.

Let $T = (V(T), E(T); f, g)$ be a weighted tree on $n \geq 2$ vertices, assume u is a leaf vertex and $p_u = (u, v)$ is a pendant edge of T, we define a weighted tree $T' = (V(T'), E(T'); f', g')$ from T with $V(T') = V(T) \backslash u$, $E(T') = E(T) \backslash p_u$,

$$f'(w) = \begin{cases} f(v)(1 + f(u)g(p_u)) & \text{if } w = v, \\ f(w) & \text{otherwise.} \end{cases} \tag{1}$$

for any $w \in V(T')$, and $g'(e) = g(e)$ for any $e \in E(T')$.

Lemma 1 ([22]). *Assume T and T' are weighted trees defined above, and u ($\neq v_i$) is an arbitrary vertex, then*

$$F(T; f, g; v_i) = F(T'; f', g'; v_i),$$

$$F(T; f, g) = F(T'; f', g') + f(u).$$

Assume v_0, v_l are two distinct vertices of weighted tree $T = (V(T), E(T); f, g)$, denote by $P_{v_0 v_l} = v_0 v_1 \cdots v_l$ the unique path of length l (≥ 1) connecting v_0 and v_l with $V(P_{v_0 v_l}) = \{v_i | i = 0, 1, \ldots, l\}$, $E(P_{v_0 v_l}) = \{(v_i, v_{i+1}) | i = 0, 1, \ldots, l-1\}$. Moreover, denote by T_{v_i} the weighted subtree that contains vertex v_i ($i = 0, 1, \ldots, l$) after removing all edges in $E(P_{v_0 v_l})$ from T.

Lemma 2 ([22]). *With the above notations, we have*

$$F(T; f, g; v_0, v_l) = \prod_{i=0}^{l} f^*(v_i) \prod_{e \in E(P_{v_0 v_l})} g(e), \tag{2}$$

where $f^*(v_i) = F(T_{v_i}; f, g; v_i)$ for any $v_i \in V(P_{v_0 v_l})$.

Let $U_{m,n} = (V(U_{m,n}), E(U_{m,n}); f_{m,n}, g_{m,n})$ be a weighted unicyclic graph of order $m \geq 2$ whose unique cycle has vertices $v_1, \ldots, v_n$ with $m \geq n \geq 2$. At each vertex v_i let the tree T_{v_i} rooted at v_i be the component containing v_i after removing the cycle. We can contract each tree T_{v_i} to the vertex v_i on the unique cycle with the contraction rule as in eq. (1) to make the computing more efficient.

For convenience, denote by $U_n = (V(U_n), E(U_n); f_n, g_n)$ the weighted unicyclic graph obtained from $U_{m,n} = (V(U_{m,n}), E(U_{m,n}); f_{m,n}, g_{m,n})$ by contracting each tree T_{v_i} to the vertex v_i on the unique cycle defined above, where $V(U_n) = \{v_i | i = 1, 2, \ldots, n\}$, $E(U_n) = \{(v_i, v_{i+1}) | i = 1, 2, \ldots, n\}$ (assume $v_{n+i} = v_i$ for all i), and f_n is the vertex weight function and g_n is the edge weight function, note that the weight of each vertex (resp. edge) may be different with each other. Through classifying the subtrees of U_n into n types: subtrees that do not contain the edge (v_1, v_n); subtrees that contain the edge set $\bigcup_{k=-1}^{j} (v_{n-k}, v_{n-k-1})$, but not edge (v_{n-j-1}, v_{n-j-2}), where $v_{n+1} = v_1$ and $j = -1, 0, \ldots, n-3$. From the definitions of subtree weight and subtree generating function, Lemmas 1 and 2, it is not difficult to obtain the following theorem.

Theorem 2.1. *Let $U_n = (V(U_n), E(U_n); f_n, g_n)$ be a weighted unicyclic graph, then*

$$
\begin{aligned}
F(U_n; f_n, g_n) =& f_n(v_1) \Big(\prod_{j=1}^{n-1} g_n(v_{n-j+2}, v_{n-j+1}) f_n(v_{n-j+1}) \sum_{s=j+1}^{n} \prod_{k=1}^{n-s} g_n(v_k, v_{k+1}) f_n(v_{k+1}) \Big) \\
&+ \sum_{j=1}^{n-1} f_n(v_j) \Big(\sum_{s=1}^{n-j+1} \prod_{k=j}^{n-s} g_n(v_k, v_{k+1}) f_n(v_{k+1}) \Big).
\end{aligned}
$$
(3)

Through similar analysis, for any fixed vertex $v_j \in V(U_n)$, the subtree generating function of U_n containing v_j follows immediately.

Theorem 2.2. *Given weighted unicyclic graph $U_n = (V(U_n), E(U_n); f_n, g_n)$ and a fixed vertex $v_j \in V(U_n)$, we have*

$$
F(U_n; f_n, g_n; v_j) = \sum_{q=1}^{n} \Big(f_n(v_j) \prod_{k=0}^{q-2} g_n(v_k^1, v_k^2) f_n(v_k^2) \big(1 + \sum_{s=0}^{n-q-1} \prod_{k=0}^{s} g_n(v_k^3, v_k^4) f_n(v_k^4) \big) \Big),
$$
(4)

where $v_k^1 = v_{(j+k)(mod\ n)}$, $v_k^2 = v_{(j+k+1)(mod\ n)}$, $v_k^3 = v_{(n+j-k)(mod\ n)}$ and $v_k^4 = v_{(n+j-k-1)(mod\ n)}$.

Let $P_{v_{r_i} v_{r_j}} = v_{r_i} \cdots v_{r_j}$ be a path of U_n, we define the weighted unicyclic graph $U_n^c = (V(U_n^c), E(U_n^c); f_n^c, g_n^c)$ by contracting the path $P_{v_{r_i} v_{r_j}} = v_{r_i} \cdots v_{r_j}$ to v_{r_i}, with $V(U_n^c) = \{v_{r_i}\} \cup \{V(U_n) \backslash V(P_{v_{r_i} v_{r_j}})\}$, $E(U_n^c) = E(U_n) \backslash E(P_{v_{r_i} v_{r_j}})$,

$$
f_n^c(v_{r_i}) = \prod_{v \in V(P_{v_{r_i} v_{r_j}})} f_n(v) \prod_{e \in E(P_{v_{r_i} v_{r_j}})} g_n(e),
$$

4

$f_n^c(v) = f_n(v)$ for $v \in V(U_n^c) \backslash v_{r_i}$, and $g_n^c(e) = g_n(e)$ for $e \in E(U_n^c)$. From the definitions of subtree weight and subtree generating function, with Theorem 2.2, we can obtain the subtree generating function of U_n containing path $P_{v_{r_i} v_{r_j}}$ as follows.

Theorem 2.3. *Assume U_n and U_n^c are weighted unicyclic graphs defined above, and $P_{v_{r_i} v_{r_j}} = v_{r_i} \cdots v_{r_j}$ a path of U_n, then*

$$F(U_n; f_n, g_n; P_{v_{r_i} v_{r_j}}) = F(U_n^c; f_n^c, g_n^c; v_{r_i}). \tag{5}$$

Assume v_i, v_j are two distinct vertices of U_n, the two paths connecting v_i and v_j are denoted by

$$P_{v_i v_j}^1 = v_i v_{(i+1)(\bmod\ n)} \cdots v_{(j-1)(\bmod\ n)} v_j$$

and

$$P_{v_i v_j}^2 = v_i v_{(i-1)(\bmod\ n)} \cdots v_{(j+1)(\bmod\ n)} v_j,$$

respectively.

We define the weighted unicyclic graph $U_n^{c_1} = (V(U_n^{c_1}), E(U_n^{c_1}); f_n^{c_1}, g_n^{c_1})$ from U_n by contracting the path $P_{v_i v_j}^1$ to v_i, with $V(U_n^{c_1}) = \{v_i\} \cup \{V(U_n) \backslash V(P_{v_i v_j}^1)\}$, $E(U_n^{c_1}) = E(U_n) \backslash E(P_{v_i v_j}^1)$,

$$f_n^{c_1}(v_i) = \prod_{v \in V(P_{v_i v_j}^1)} f_n(v) \prod_{e \in E(P_{v_i v_j}^1)} g_n(e),$$

$f_n^{c_1}(v) = f_n(v)$ for $v \in V(U_n^{c_1}) \backslash v_i$, and $g_n^{c_1}(e) = g_n(e)$ for $e \in E(U_n^{c_1})$.

Similarly, we define the weighted unicyclic graph $U_n^{c_2} = (V(U_n^{c_2}), E(U_n^{c_2}); f_n^{c_2}, g_n^{c_2})$ from U_n by contracting the path $P_{v_i v_j}^2$ to v_i, with $V(U_n^{c_2}) = \{v_i\} \cup \{V(U_n) \backslash V(P_{v_i v_j}^2)\}$, $E(U_n^{c_2}) = E(U_n) \backslash E(P_{v_i v_j}^2)$,

$$f_n^{c_2}(v_i) = \prod_{v \in V(P_{v_i v_j}^2)} f_n(v) \prod_{e \in E(P_{v_i v_j}^2)} g_n(e),$$

$f_n^{c_2}(v) = f_n(v)$ for $v \in V(U_n^{c_2}) \backslash v_i$, and $g_n^{c_2}(e) = g_n(e)$ for $e \in E(U_n^{c_2})$.

From the definitions of subtree weight and subtree generating function, with Theorem 2.3, we can obtain the subtree generating function of U_n containing any prescribed two distinct vertices of U_n as follows.

Corollary 2.4. *Assume U_n, $U_n^{c_1}$ and $U_n^{c_2}$ are weighted unicyclic graph defined above, and v_i, v_j are two prescribed distinct vertices of U_n, then*

$$F(U_n; f_n, g_n; v_i, v_j) = F(U_n^{c_1}; f_n^{c_1}, g_n^{c_1}; v_i) + F(U_n^{c_2}; f_n^{c_2}, g_n^{c_2}; v_i). \tag{6}$$

A polyphenylene chain RPC_n with n hexagons can be obtained by adjoining a polyphenylene chain RPC_{n-1} with $n-1$ hexagons and a new terminal hexagon H_n with a cut edge (see Fig. 1), for $n \geq 3$, the terminal hexagon can be attached

5

in three different ways, which results in the local arrangements we describe as RPC_{n+1}^1, RPC_{n+1}^2, RPC_{n+1}^3 (see Fig. 2).

A random polyphenylene chain $RPC(n, p_1, p_2)$ with n hexagons is a polyphenylene chain obtained by step-wise addition of terminal hexagons. At each step $(i = 3, 4, \ldots, n)$, a random selection is made from one of the three possible constructions:

(1) $RPC_{i-1} \to RPC_i^1$ with probability p_1,

(2) $RPC_{i-1} \to RPC_i^2$ with probability p_2,

(3) $RPC_{i-1} \to RPC_i^3$ with probability $1 - p_1 - p_2$,

Here the probabilities p_1 and p_2 are constants. Namely, the process described is a zeroth-order Markov Process.

(a) A polyphenylene chain RPC_n with n hexagons.

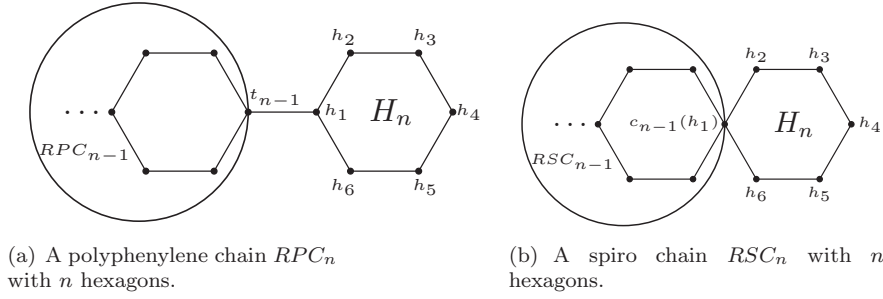(b) A spiro chain RSC_n with n hexagons.

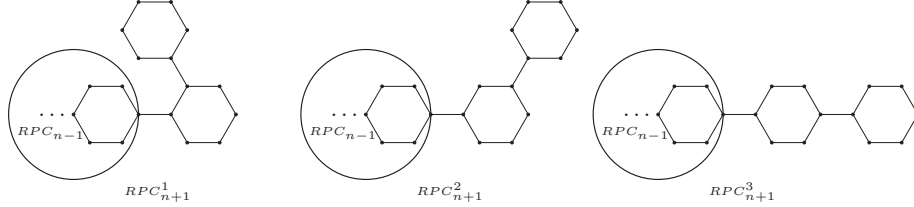Figure 1. A polyphenylene chain RPC_n and a spiro chain RSC_n.

Figure 2. Three types of local arrangements in polyphenylene chains.
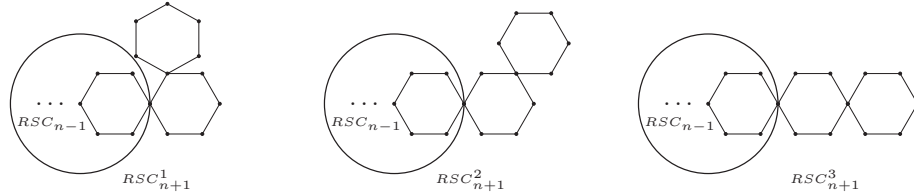

Figure 3. Three types of local arrangements in spiro chains.

Similarly, a spiro chain RSC_n with n hexagons can be obtained by adjoining a new terminal hexagon H_n to a spiro chain RSC_{n-1} with $n - 1$ hexagons (see

Fig. 1), for $n \geq 3$, the terminal hexagon can be attached in three different ways, which results in the local arrangements we describe as RSC_{n+1}^1, RSC_{n+1}^2, RSC_{n+1}^3 (see Fig. 3).

And a random spiro chain $RSC(n, p_1, p_2)$ with n hexagons is a spiro chain obtained by stepwise addition of terminal hexagons. At each step $(i = 3, 4, \ldots, n)$, a random selection is made from one of the three possible constructions:

(1) $RSC_{i-1} \to RSC_i^1$ with probability p_1,
(2) $RSC_{i-1} \to RSC_i^2$ with probability p_2,
(3) $RSC_{i-1} \to RSC_i^3$ with probability $1 - p_1 - p_2$,

Here the probabilities p_1 and p_2 are constants.

Specially, the random polyphenylene chain $RPC(n, 1, 0)$, $RPC(n, 0, 1)$ and $RPC(n, 0, 0)$ are the polyphenylene ortho-chain $\overline{O}_n$, meta-chain $\overline{M}_n$, para-chain $\overline{P}_n$, by setting $(p_1, p_2) = (1, 0), (0, 1), (0, 0)$, respectively, see Fig. 2. And similarly, the random spiro chain $RSC(n, 1, 0)$, $RSC(n, 0, 1)$, $RSC(n, 0, 0)$ are the spiro ortho-chain O_n, meta-chain M_n and para-chain P_n, respectively, see Fig. 3.

3. The expected value of the subtree number index of random polyphenylene and spiro chains

3.1. Random polyphenylene chain

Firstly, we study the subtree number index of the random polyphenylene chain.

Theorem 3.1. *For $n \geq 1$, the expected value of the subtree number index of random polyphenylene $RPC(n, p_1, p_2)$ is*

$$E(\text{STN}(RPC(n, p_1, p_2))) = \frac{441}{(11 + 4p_1 + p_2)^2}(12 + 4p_1 + p_2)^n + \frac{144p_1 + 36p_2 - 45}{11 + 4p_1 + p_2}n$$
$$- \frac{441}{(11 + 4p_1 + p_2)^2}.$$

Proof. It is easy to know that $E(STN(RPC(1, p_1, p_2))) = 36$, for $n \geq 2$, we categorize the subtrees of the random polyphenylene chain RPC_n into two cases:

(i) not containing edge (t_{n-1}, h_1),
(ii) containing edge (t_{n-1}, h_1).

It is easy to see that the subtree number of case (i) is

$$STN(RPC_{n-1}) + 36. \tag{7}$$

The subtree set of case (ii), denoted by $RPC_n(t_{n-1}, h_1)$ can be described as:

$$RPC_n(t_{n-1}, h_1) = \{T_l + (t_{n-1}, h_1) + T_r | T_l \in RPC_{n-1}(t_{n-1}), T_r \in H_n(h_1)\}, \tag{8}$$

where $T_l + (t_{n-1}, h_1) + T_r$ is the subtree obtained by connecting subtree $T_l \in RPC_{n-1}(t_{n-1})$ (subtree of RPC_{n-1} containing vertex t_{n-1}) and subtree $T_r \in H_n(h_1)$ (subtree of H_n containing vertex h_1) with edge (t_{n-1}, h_1), see Fig. 1(a).

Since the subtree number of H_n containing vertex h_1 is 21, by eq. (8), we have the subtree number of case (ii) as

$$21STN(RPC_{n-1}(t_{n-1})). \tag{9}$$

Thus by eqs. (7) and (9), we have

$$STN(RPC_n) = STN(RPC_{n-1}) + 21STN(RPC_{n-1}(t_{n-1})) + 36. \tag{10}$$

Namely,

$$STN(RPC_{n+1}) = STN(RPC_n) + 21STN(RPC_n(t_n)) + 36. \tag{11}$$

For a random polyphenylene chain $RPC(n, p_1, p_2)$, the subtree number of $RPC(n, p_1, p_2)$ containing vertex t_n is a random variable, and its expected value is denoted by

$$T_n = E(STN(RPC(n, p_1, p_2; t_n))). \tag{12}$$

By the expectation operator and eqs. (11) and (12), we can obtain a recursive relation for the expected value of the subtree number index of a random polyphenylene chain $RPC(n, p_1, p_2)$

$$E(STN(RPC(n+1, p_1, p_2))) = E(STN(RPC(n, p_1, p_2))) + 21T_n + 36. \tag{13}$$

Now, we consider computing T_n. Take the weighted random polyphenylene chain $RPC(n-1, p_1, p_2)$ as a single vertex "t_{n-1}" whose weight is $STN(RPC(n-1, p_1, p_2; t_{n-1}))$, see Fig. 1(a), then, with Theorem 2.2, we have

(i) If $RPC_n \to RPC_{n+1}^1$ with probability p_1,

$$STN(RPC(n, p_1, p_2; t_n)) = 21 + 16STN(RPC(n-1, p_1, p_2; t_{n-1}))$$

with probability p_1.

(ii) If $RPC_n \to RPC_{n+1}^2$ with probability p_2,

$$STN(RPC(n, p_1, p_2; t_n)) = 21 + 13STN(RPC(n-1, p_1, p_2; t_{n-1}))$$

with probability p_2.

(iii) If $RPC_n \to RPC_{n+1}^3$ with probability $1 - p_1 - p_2$,

$$STN(RPC(n, p_1, p_2; t_n)) = 21 + 12STN(RPC(n-1, p_1, p_2; t_{n-1}))$$

with probability $1 - p_1 - p_2$.

From (i)-(iii) above, we immediately obtain

$$
\begin{aligned}
T_n =& p_1[21 + 16STN(RPC(n-1, p_1, p_2; t_{n-1}))] + p_2[21 + 13STN(RPC(n-1, p_1, p_2; t_{n-1}))] \\
& + (1 - p_1 - p_2)[21 + 12STN(RPC(n-1, p_1, p_2; t_{n-1}))] \\
=& (12 + 4p_1 + p_2)STN(RPC(n-1, p_1, p_2; t_{n-1})) + 21.
\end{aligned}
\tag{14}
$$

By applying the expectation operator to the above eq. (14), we obtain

$$T_n = (12 + 4p_1 + p_2)T_{n-1} + 21. \tag{15}$$

Since $T_1 = 21$, using the above recurrence relation, we have

$$T_n = \frac{21}{11 + 4p_1 + p_2}(12 + 4p_1 + p_2)^n - \frac{21}{11 + 4p_1 + p_2}. \tag{16}$$

From eq. (13), we have

$$
\begin{aligned}
E(STN(RPC(n+1, p_1, p_2))) =& 21\left[\frac{21}{11 + 4p_1 + p_2}(12 + 4p_1 + p_2)^n - \frac{21}{11 + 4p_1 + p_2}\right] \\
&+ E(STN(RPC(n, p_1, p_2))) + 36 \\
=& \frac{441}{11 + 4p_1 + p_2}(12 + 4p_1 + p_2)^n - \frac{441}{11 + 4p_1 + p_2} \\
&+ E(STN(RPC(n, p_1, p_2))) + 36.
\end{aligned}
\tag{17}
$$

Using the above recurrence relation, we have

$$
\begin{aligned}
E(STN(RPC(n, p_1, p_2))) =& \frac{441(12 + 4p_1 + p_2)}{(11 + 4p_1 + p_2)^2}\left((12 + 4p_1 + p_2)^{n-1} - 1\right) \\
&+ (36 - \frac{441}{11 + 4p_1 + p_2})(n - 1) + 36.
\end{aligned}
\tag{18}
$$

The theorem thus follows.

Specially, by taking $(p_1, p_2) = (1, 0), (0, 1)$ or $(0, 0)$, respectively, and Theorem 3.1, we have the following.

Corollary 3.2. *The subtree number indices of the polyphenylene ortho-chain $\overline{O}_n$, meta-chain $\overline{M}_n$ and para-chain $\overline{P}_n$ are*

$$
\begin{aligned}
STN(\overline{O}_n) &= \frac{49(16^n - 1)}{25} + \frac{33n}{5}, \\
STN(\overline{M}_n) &= \frac{49(13^n - 1)}{16} - \frac{3n}{4}, \\
STN(\overline{P}_n) &= \frac{441(12^n - 1)}{121} - \frac{45n}{11}.
\end{aligned}
$$

The results of Corollary 3.2 agree with the subtree numbers of $\overline{O}_n$, $\overline{M}_n$ and $\overline{P}_n$ presented in [26].

3.2. Random spiro chain

Theorem 3.3. *For $n \geq 1$, the expected value of the subtree number index of a random spiro chain $RSC(n, p_1, p_2)$ is*

$$E(\text{STN}(RSC(n, p_1, p_2))) = \frac{400}{(11 + 4p_1 + p_2)^2}(12 + 4p_1 + p_2)^n + \frac{140p_1 + 35p_2 - 15}{11 + 4p_1 + p_2}n$$
$$- \frac{400}{(11 + 4p_1 + p_2)^2} + 1.$$

Proof. It is not difficult to obtain that $E(STN(RSC(1, p_1, p_2))) = 36$, for $n \geq 2$, assume that the terminal hexagon is spanned by vertices h_2, h_3, h_4, h_5, h_6, and the vertex h_1(i.e. c_{n-1}) (see Fig. 1(b)). We categorize the subtrees of the random spiro chain RSC_n into four cases:

(i) contain neither (c_{n-1}, h_2) nor (c_{n-1}, h_6),
(ii) contain (c_{n-1}, h_2) but not (c_{n-1}, h_6),
(iii) contain (c_{n-1}, h_6) but not (c_{n-1}, h_2),
(iv) contain both (c_{n-1}, h_2) and (c_{n-1}, h_6).
by Lemma 1, the subtree number of case (i) is

$$STN(RSC_{n-1}) + 15. \tag{19}$$

Taking the the random spiro chain RSC_{n-1} as a single vertex "c_{n-1}" with weight $STN(RSC_{n-1}(c_{n-1}))$ (namely, subtree number of RSC_{n-1} containing vertex c_{n-1}) (see Fig. 1(b)), then, by Lemma 2, we know that both the subtree number of case (ii) and case (iii) are

$$5STN(RSC_{n-1}(c_{n-1})). \tag{20}$$

and further with Theorem 2.3, we can obtain that the subtree number of case (iii) is

$$10STN(RSC_{n-1}(c_{n-1})). \tag{21}$$

thus, we have

$$STN(RSC_n) = STN(RSC_{n-1}) + 20STN(RSC_{n-1}(c_{n-1})) + 15. \tag{22}$$

Namely,

$$STN(RSC_{n+1}) = STN(RSC_n) + 20STN(RSC_n(c_n)) + 15. \tag{23}$$

For a random spiro chain $RSC(n, p_1, p_2)$, the subtree number of $RSC(n, p_1, p_2)$ containing vertex c_n is a random variable, and its expected value is denoted by

$$C_n = E(STN(RSC(n, p_1, p_2; c_n))). \tag{24}$$

10

By the expectation operator and eqs. (23) and (24), we can obtain a recursive relation for the expected value of the subtree number index of a random spiro chain $RSC(n, p_1, p_2)$

$$E(STN(RSC(n+1, p_1, p_2))) = E(STN(RSC(n, p_1, p_2))) + 20C_n + 15. \quad (25)$$

Now, we consider C_n. Again, take the the random spiro chain RSC_{n-1} as a single vertex "c_{n-1}" with weight $STN(RSC_{n-1}(c_{n-1}))$ (see Fig. 1(b)), then, with Theorem 2.2,

(i) If $RSC_n \rightarrow RSC_{n+1}^1$ with probability p_1,

$$STN(RSC(n, p_1, p_2; c_n)) = 5 + 16STN(RSC(n-1, p_1, p_2; c_{n-1}))$$

with probability p_1.

(ii) If $RSC_n \rightarrow RSC_{n+1}^2$ with probability p_2,

$$STN(RSC(n, p_1, p_2; c_n)) = 8 + 13STN(RSC(n-1, p_1, p_2; c_{n-1}))$$

with probability p_2.

(iii) If $RSC_n \rightarrow RSC_{n+1}^3$ with probability $1 - p_1 - p_2$,

$$STN(RSC(n, p_1, p_2; c_n)) = 9 + 12STN(RSC(n-1, p_1, p_2; c_{n-1}))$$

with probability $1 - p_1 - p_2$.

From (i)-(iii) above, we immediately obtain

$$
\begin{aligned}
C_n =& p_1[5 + 16STN(RSC(n-1, p_1, p_2; c_{n-1}))] + p_2[8 + 13STN(RSC(n-1, p_1, p_2; c_{n-1}))] \\
&+ (1 - p_1 - p_2)[9 + 12STN(RSC(n-1, p_1, p_2; c_{n-1}))] \\
=& (12 + 4p_1 + p_2)STN(RSC(n-1, p_1, p_2; c_{n-1}) + 9 - (4p_1 + p_2).
\end{aligned}
$$

$$(26)$$

By applying the expectation operator to the above eq. (26), we obtain

$$C_n = (12 + 4p_1 + p_2)C_{n-1} + 9 - (4p_1 + p_2). \quad (27)$$

Since $C_1 = 21$, using the above recurrence relation, we have

$$C_n = \frac{20}{11 + 4p_1 + p_2}(12 + 4p_1 + p_2)^n + 1 - \frac{20}{11 + 4p_1 + p_2}. \quad (28)$$

It is easy to see that $E(STN(RSC(1, p_1, p_2))) = 36$, from eq. (25), we have

$$
\begin{aligned}
E(STN(RSC(n+1, p_1, p_2))) =& 20[\frac{20}{11 + 4p_1 + p_2}(12 + 4p_1 + p_2)^n + 1 - \frac{20}{11 + 4p_1 + p_2}] \\
&+ E(STN(RSC(n, p_1, p_2))) + 15 \\
=& \frac{400}{11 + 4p_1 + p_2}(12 + 4p_1 + p_2)^n - \frac{400}{11 + 4p_1 + p_2} \\
&+ E(STN(RSC(n, p_1, p_2))) + 35.
\end{aligned}
$$

$$(29)$$

11

Using the above recurrence relation, we have

$$
\begin{aligned}
E(STN(RSC(n, p_1, p_2))) =& \frac{400(12 + 4p_1 + p_2)}{(11 + 4p_1 + p_2)^2}\big((12 + 4p_1 + p_2)^{n-1} - 1\big) \\
&+ (35 - \frac{400}{11 + 4p_1 + p_2})(n - 1) + 36.
\end{aligned}
\tag{30}
$$

The theorem holds immediately.

Let $(p_1, p_2) = (1, 0), (0, 1),$ or $(0, 0)$, we can obtain the subtree number index of the spiro ortho-chain O_n, the meta-chain M_n and the para-chain P_n, respectively, with Theorem 3.3, we have

Corollary 3.4. *The subtree number indices of the spiro ortho-chain O_n, the meta-chain M_n and the para-chain P_n are*

$$
\begin{aligned}
STN(O_n) &= \frac{256(16^{n-1} - 1)}{9} + \frac{25(n-1)}{3} + 36, \\
STN(M_n) &= \frac{325(13^{n-1} - 1)}{9} + \frac{5(n-1)}{3} + 36, \\
STN(P_n) &= \frac{4800(12^{n-1} - 1)}{121} - \frac{15(n-1)}{11} + 36.
\end{aligned}
\tag{31}
$$

Again, the results of Corollary 3.4 agree with the subtree numbers of O_n, M_n and P_n presented in [26].

3.3. A relation between $E(\mathrm{STN}(RPC))$ and $E(\mathrm{STN}(RSC))$

It is easy to see that every spiro chain could be obtained by squeezing off the cut edges of a polyphenylene chain. Pavlović and Gutman [14], Deng [4] provided a formula of the relation between the Wiener indices of a polyphenylene chain and its squeeze independently. In 2015, Yang et al. [26] presented a formula of the relation between the subtree number index of these two chains.

When random structures are concerned, Yang and Zhang [23] presented an exact formula for the expected value of the Wiener index of a random polyphenylene chain $RPC(n, p_1, p_2)$ with the same probabilities p_1 and p_2, Regarding the random polyphenylene chain $RPC(n, p_1, p_2)$ and spiro chain $RSC(n, p_1, p_2)$, Huang, Kuang and Deng [7] presented a relation between the expected values of the Kirchhoff indices of these two chains.

In what follows, we present a relation between the expected values of the subtree number indices of the random polyphenylene chain $RPC(n, p_1, p_2)$ and the random spiro chain $RSC(n, p_1, p_2)$ from Theorems 3.1 and 3.3.

Theorem 3.5. *For a random polyphenylene chain $RPC(n, p_1, p_2)$ and a random spiro chain $RSC(n, p_1, p_2)$ with n hexagons, the expected values of their subtree number indices are related as*

$$
400E(\mathrm{STN}(RPC(n, p_1, p_2))) = 441E(\mathrm{STN}(RSC(n, p_1, p_2))) - 1035n - 441.
\tag{32}
$$

Proof. From Eq. (18) and (30), we have

$$\frac{E(STN(RPC(n,p_1,p_2)))-(36-\frac{441}{11+4p_1+p_2})(n-1)-36}{E(STN(RSC(n,p_1,p_2)))-(35-\frac{400}{11+4p_1+p_2})(n-1)-36}=\frac{441}{400}, \qquad (33)$$

or equivalently,

$$400E(STN(RPC(n,p_1,p_2)))=441E(STN(RSC(n,p_1,p_2)))-1035n-441.$$

$\square$

By Theorem 3.5, the expected value of the subtree number index of the random spiro chain is less than the random polyphenylene chain. In fact, for $n \to \infty$,

$$E(STN(RSC(n,p_1,p_2))) \approx \frac{400}{441}E(STN(RPC(n,p_1,p_2))).$$

From Theorems 3.1 and 3.3, we also point out that

$$E(STN(RPC(n,p_1,p_2))) \approx \frac{441}{(11+4p_1+p_2)^2}(12+4p_1+p_2)^n$$

and

$$E(STN(RSC(n,p_1,p_2))) \approx \frac{400}{(11+4p_1+p_2)^2}(12+4p_1+p_2)^n.$$

Namely, the values of $E(STN(RPC(n,p_1,p_2)))$ and $E(STN(RSC(n,p_1,p_2)))$ are asymptotic to exponential function in n as $n \longrightarrow \infty$.

4. Average value of the subtree number index

Let $\overline{G}_n$ be the set of all polyphenylene chains with n hexagons. The average value of the subtree number indices with respect to $\overline{G}_n$ is

$$STN_{avr}(\overline{G}_n) = \frac{1}{|\overline{G}_n|}\sum_{G\in\overline{G}_n}STN(G).$$

In order to obtain the average value of the subtree number indices with respect to $\overline{G}_n$, we only need to take $p_1 = p_2 = \frac{1}{3}$ in the random polyphenylene chain $RPC(n,p_1,p_2)$, i.e., the average value of the subtree number indices with respect to $\overline{G}_n$ is just the expected value of the subtree number index of the random polyphenylene chain $RPC(n,p_1,p_2)$ for $p_1 = p_2 = \frac{1}{3}$. From Theorem 3.1, we have

Theorem 4.1. *The average value of the subtree number indices with respect to $\overline{G}_n$ is*

$$\text{STN}_{avr}(\overline{G}_n) = \frac{3969}{1444}(\frac{41}{3})^n + \frac{45}{38}n - \frac{3969}{1444}.$$

13

Similarly, Let G_n be the set of all spiro chains with n hexagons. The average value of the subtree number indices with respect to G_n is

$$STN_{avr}(G_n) = \frac{1}{|G_n|} \sum_{G \in G_n} STN(G),$$

and the average value of the subtree number indices with respect to G_n is just the expected value of the subtree number index of the random spiro chain $RSC(n, p_1, p_2)$ for $p_1 = p_2 = \frac{1}{3}$. From Theorem 3.3, we have

Theorem 4.2. *The average value of the subtree number indices with respect to $\overline{G}_n$ is*

$$\text{STN}_{avr}(G_n) = \frac{900}{361}\left(\frac{41}{3}\right)^n + \frac{130}{38}n - \frac{539}{361}.$$

5. Concluding remarks

In this paper we obtain exact formulas for the expected values of subtree number index of the random polyphenylene and spiro chains, and then establish a relation between the expected values of the subtree number indices of a random polyphenylene and its corresponding random hexagonal squeeze, we also briefly study the average values for subtree number indices with respect to the set of all polyphenylene and spiro chains with n hexagons.

For future works, we plan to study the expected values of subtree number index of the random hexagonal chains, phenylene chains and other regular chemical structures such as cata-condensed hexagonal systems. Meanwhile, It is also interesting to study the expected values of the recently proposed multi-distance granularity structural α-subtree index [24] of the random polyphenylene and spiro chains, as well as other regular chemical structures.

References

[1] X.L. Chen, B. Zhao, P.Y. Zhao, Six-membered ring spiro chains with extremal Merrifield-Simmons index and Hosoya index, MATCH Commun. Math. Comput. Chem. 62 (2009) 657.

[2] A.J. Chin, G. Gordon, K.J. MacPhee, C. Vincent, Subtrees of graphs, J. Graph Theory. 89 (2018) 413-438.

[3] É. Czabarka, L.A. Székely, S. Wagner, On the number of nonisomorphic subtrees of a tree, J. Graph Theory. 87 (2018) 89-95.

[4] H.Y. Deng, Wiener indices of spiro and polyphenyl hexagonal chains, Math. Comput. Model. 55 (2012) 634-644.

[5] T. Došlić, F. Måløy, Chain hexagonal cacti: Matchings and independent sets, Discrete Math. 310 (2010) 1676-1690.

[6] J. Haslegrave, Extremal results on average subtree density of series-reduced trees, J. Comb. Theory, Ser. B. 107 (2014) 26-41.

[7] G.H. Huang, M.J. Kuang, H.Y. Deng, The expected values of Kirchhoff indices in the random polyphenyl and spiro chains, Ars Math. Contemp. 9 (2013) 207-217.

[8] G.H. Huang, M.J. Kuang, H.Y. Deng, The expected values of Hosoya index and Merrifield–Simmons index in a random polyphenylene chain, J. Comb. Optim. 32 (2016) 550-562.

[9] R.E. Jamison, On the average number of nodes in a subtree of a tree, J. Comb. Theory, Ser. B. 35 (1983) 207-223.

[10] R. Kirk, H. Wang, Largest number of subtrees of trees with a given maximum degree, SIAM J. Discret. Math. 22 (2008) 985-995.

[11] B. Knudsen, Optimal multiple parsimony alignment with affine gap cost using a phylogenetic tree, Algorithms Bioinformatics, Lect. Notes Comput. Sci. 2812 (2003) 433-446.

[12] X.Y. Li, G.P. Wang, H. Bian, R.W. Hu, The Hosoya polynomial decomposition for polyphenyl chains, MATCH Commun. Math. Comput. Chem. 67 (2012) 357.

[13] R.E. Merrifield, H.E. Simmons, Topological Methods in Chemistry, Wiley, New York, 1989.

[14] L. Pavlović, I. Gutman, Wiener numbers of phenylenes: an exact result, J. Chem. Inf. Comput. Sci. 37 (1997) 355-358.

[15] L.A. Székely, H. Wang, On subtrees of trees, Adv. Appl. Math. 34 (2005) 138-155.

[16] L.A. Székely, H. Wang, Binary trees with the largest number of subtrees, Discrete Appl. Math. 155 (2007) 374-385.

[17] L.A. Székely, H. Wang, Extremal values of ratios: Distance problems vs. subtree problems in trees II, Discrete Math. 322 (2014) 36-47.

[18] A. Vince, H. Wang, The average order of a subtree of a tree, J. Comb. Theory, Ser. B. 100 (2010) 161-170.

[19] S.G. Wagner, Correlation of graph-theoretical indices, SIAM J. Discret. Math. 21 (2007) 33-46.

[20] S.L. Wei, X.L. Ke, G.L. Hao, Comparing the excepted values of atom-bond connectivity and geometric-arithmetic indices in random spiro chains, J. Inequalities Appl. 2018 (2018) 45.

[21] Y.Z. Xiao, H.X. Zhao, Z. Liu, Y.P. Mao, Trees with large numbers of subtrees, Int. J. Comput. Math. 94 (2017) 372-385.

[22] W.G. Yan, Y.N. Yeh, Enumeration of subtrees of trees, Theor. Comput. Sci. 369 (2006) 256-268.

[23] W.L. Yang, F.J. Zhang, Wiener index in random polyphenyl chains, MATCH Commun. Math. Comput. Chem. 68 (2012) 371-376.

[24] Y. Yang, A.W. Fan, H. Wang, H.L. Lv, X.D. Zhang, Multi-distance granularity structural -subtree index of generalized Bethe trees, Appl. Math. Comput. 359 (2019) 107-120.

[25] Y. Yang, H.B. Liu, H. Wang, A.S. Deng, C. Magnant, On Algorithms for Enumerating Subtrees of Hexagonal and Phenylene Chains, Comput. J. 60 (2017) 690-710.

[26] Y. Yang, H.B. Liu, H. Wang, H.S. Fu, Subtrees of spiro and polyphenyl hexagonal chains, Appl. Math. Comput. 268 (2015) 547-560.

[27] L.L. Zhang, Q.S. Li, S.C. Li, M.J. Zhang, The expected values for the Schultz index, Gutman index, multiplicative degree-Kirchhoff index and additive degree-Kirchhoff index of a random polyphenylene chain, Discrete Appl. Math. 282 (2020) 243-256.

[28] X.-M. Zhang, H. Wang, X.-D. Zhang, On the eccentric subtree number in trees, Discrete Appl. Math. (2019). https://doi.org/10.1016/j.dam.2019.08.026.

[29] X.-M. Zhang, X.-D. Zhang, The Minimal Number of Subtrees with a Given Degree Sequence, Graphs Comb. 31 (2015) 309-318.

[30] X.-M. Zhang, X.-D. Zhang, D. Gray, H. Wang, The number of subtrees of trees with given degree sequence, J. Graph Theory. 73 (2013) 280-295.