Research Article

Saylé Sigarreta*, Saylí Sigarreta, and Hugo Cruz-Suárez

Topological indices for random spider trees

https://doi.org/10.1515/mgmc-2022-0025
received April 18, 2022; accepted November 07, 2022

Abstract: In this study, we characterize the structure and some topological indices of a class of random spider trees (RSTs) such as degree-based Gini index, degree-based Hoover index, generalized Zagreb index, and other indices associated with these. We obtain the exact and asymptotic distributions of the number of leaves via probabilistic methods. Moreover, we relate this model to the class of RSTs that evolves in a preferential attachment manner.

Keywords: theoretical chemistry, random trees, spider trees, Gini index, Zagreb index, topological indices

1 Introduction

Initiated in 1736 by Euler and developed in the 19th century by the Englishmen A. Cayley and J.J. Silvester, graph theory has become a very powerful practical and theoretical tool (Abbas et al., 2021a,b; Afzal Siddiqui et al., 2021; Ahmad et al., 2022; Alatawi et al., 2021; Imran et al., 2021; Nadeem et al., 2021; Raza et al., 2021, 2022; Zuo et al., 2021). A graph G is determined by two sets (V, E), the set of nodes and edges. The edges and nodes are interpreted according to the problem to be modeled. Highlighting the trees as a very important and studied family of graphs, which from its origin has proven to have many applications in different areas. In mathematical chemistry, trees are used to characterize the molecular structure of chemical compounds; in this context, the nodes represent the molecules and the edges the chemical bonds (Kier and Hall, 1986). One relevant class of trees for chemical studies are the trees with a given number of pendants. A node is called pendant if it has degree 1. Ducoffe et al. (2018) proved that the trees with n pendants ($n \geq 3$) that maximize the modified first Zagreb connection index must be spider trees or double stars. On the other hand, Shiu (2008) reported that spider trees are used to study hexagonal systems that model benzenoid molecules and unbranched catacondensed benzenoid molecules.

The structural information of a graph can be represented in different ways: matrices, polynomials, topological indices, etc. The topological indices quantify the structural information contained in the graph and are independent of the numbering of the nodes and edges; hence they are called topological. The theoretical and practical interest of topological indices have experienced explosive growth from its introduction, resulting in countless papers published that are able to position them as a useful tool in multiple practical problems of computer science (Gutman et al., 2018), physic (Estrada, 2010), ecology (Pineda-Pineda et al., 2020), and chemistry (Kashif et al., 2021; Rao et al., 2021; Reždepović and Furtula, 2020; Shao et al., 2022). As a summary, the first research in this area appeared in the report by Wiener (1947) giving rise to the now well-known Wiener index to analyze and correlate the physicochemical properties of alkanes. In 1971, Haruo Hosoya continued the research on topological indices by introducing the Z index of Hosoya (Hosoya, 1971). On the other hand, the Zagreb index appeared for the first time in Gutman and Trinajstić (1972). Then, it was defined by Randić (1975), the Randić index, considered possibly the most studied and applied topological index at present, giving way to generalizations such as the "molecular connectivity indices," introduced by Kier et al. (1975).

In the development of applications, it has become natural to conclude that random graphs are an appropriate and useful tool to analyze phenomena that evolve over time, since many important characteristics are difficult to capture using deterministic models. In this sense, it is important to mention that some works perform studies of topological indices on random graphs. For a better treatment we refer interested readers to Aguilar-Sánchez et al. (2021), Kazemi (2021), Li et al. (2021a, 2021b), Martínez-Martínez et al. (2020), Pegu et al. (2021), and Zhang and Wang (2022). In particular, motivated by the substantial increase in interests in random tree models and considering the arguments put forward in the previous

*** Corresponding author: Saylé Sigarreta,** Benemérita Universidad Autónoma de Puebla, Ave. San Claudio y Río Verde, Col. San Manuel, CU, Puebla 72570, Puebla, Mexico,
e-mail: sayle.sigarretar@alumno.buap.mx
Saylí Sigarreta, Hugo Cruz-Suárez: Benemérita Universidad Autónoma de Puebla, Ave. San Claudio y Río Verde, Col. San Manuel, CU, Puebla 72570, Puebla, Mexico

paragraphs, in this article, we considered a class of spider trees that are incorporated with randomness, called random spider trees (RSTs), and we investigated several useful topological indices of this random class, including degree-based Gini index, degree-based Hoover index, generalized Zagreb index, and other indices associated with these. Specifically, a central limit theorem is developed for the asymptotic distribution of the number of leaves in an RST.

Notation: $\mathbb{R}$ denotes the set of real numbers. The expected value and the variance of a random variable X in $(\Omega, \mathscr{F}, P)$ are denoted as $E(X)$ and $V(X)$. On the other hand, $X \sim F$ means that the random variable X has a distribution function F and M_X denotes the moment generating function of the random variable X. $\text{Bin}(n-1, p)$ represents a random variable with binomial distribution with parameters $n - 1 \geq 0$ and p in [0,1]. $\text{Ber}(p)$ represents a random variable with Bernoulli distribution with parameter p in [0,1]. $N(\mu, \sigma^2)$ represents a random variable with normal distribution where μ in $\mathbb{R}$ is the mean and $\sigma^2 > 0$ is the variance of the random variable. $\chi^2(\lambda, k)$ represents a random variable with noncentral chi-squared distribution where $k > 0$ is the degrees of freedom and $\lambda > 0$ is the non-centrality parameter. For probabilistic convergence we use $\rightarrow P$ to denote convergence in probability and $\rightarrow D$ to denote convergence in distribution. Let $r > 0$, $\rightarrow L_r$ denotes convergence in r-mean. Given two real-valued functions $f(x)$ and $g(x) \neq 0$, we call $f(x) = o(g(x))$, if we have $f(x)/g(x) \to 0$ as $x \to \infty$. Given two real-valued functions $f(x)$ and $g(x)$, we call $f(x) = O(g(x))$, if there exists a positive real number N and a real number x_0 such that $|f(x)| \leq Ng(x)$ for all $x \geq x_0$.

2 RSTs

A spider tree is a connected tree with a centroid of degree of at least 3. All the remaining nodes are classified into two categories: internal nodes of degree 2 and leaves of degree 1, respectively. Thus, except for the centroid, all the nodes in a spider tree have degrees of at most 2. The class of RSTs considered in this study evolves in the following way: at time 1, an RST starts with a seed graph containing a centroid and three leaves. At each subsequent stage, the leaves and centroid will be able to recruit new nodes (at time n):

1) The centroid will be selected with probability p, $0 < p < 1$.
2) A leaf will be selected with probability $(1-p)/L_n$, where L_n denotes the number of leaves in the RST at time n, where

$$p + \sum_{1}^{L_n} \frac{1-p}{L_n} = p + \frac{(1-p)L_n}{L_n} = 1.$$

Note that only the centroid and leaves are qualified for recruiting new nodes. If the centroid is selected, a new leaf is attached to it; if a leaf is selected, a new leaf is attached to the (selected) leaf, and the recruiter is converted to an internal node. Finally, we have that at each stage the generated graph is a spider tree with $n + 3$ nodes.

2.1 Leaves

In the following, L_n denotes the number of leaves in an RST at time n, with $n \geq 1$. For $n \geq 2$, by the construction of the model it follows that

$$P(L_n = L_{n-1} + 1 | L_{n-1}) = p.$$

Consequently, let $t \in \mathbb{R}$, then we obtain a recurrence relationship

$$M_{L_n}(t) = \sum_{i=3}^{n+2} e^{ti} P(L_n = i) = (1-p) \sum_{i=3}^{n+1} e^{ti} P(L_{n-1} = i)$$
$$+ p \sum_{i-1=3}^{n+1} e^{ti} P(L_{n-1} = i - 1) = \left(1 - p + pe^t\right) M_{L_{n-1}}(t).$$

Now, we solve the recurrence relationship with the initial value $L_1 = 3$, obtaining

$$M_{L_n}(t) = (1 - p + pe^t)^{n-1} e^{3t}, \quad t \in \mathbb{R}.$$

By the above result, $L_n - 3 = \text{Bin}(n-1, p)$, applying the central limit theorem (Gut, 2005) we get

$$\frac{L_n - 3 - (n-1)p}{\sqrt{(1-p)p(n-1)}} \xrightarrow{D} N(0, 1).$$

Now, note that for each $k \in \mathbb{R}$

$$\frac{L_n - 3 - (n-1)p}{\sqrt{(1-p)p(n+k)}} = \frac{L_n - 3 - (n-1)p}{\sqrt{(1-p)p(n-1)}} \frac{\sqrt{(1-p)p(n-1)}}{\sqrt{(n+k)(1-p)p}}.$$

In addition, $\frac{\sqrt{(1-p)p(n-1)}}{\sqrt{(1-p)p(n+k)}} \to 1, \quad$ when $n \to \infty$. Finally, the following proposition follows from combining these results.

Proposition 1. *For $n \geq 1$ and $0 < p < 1$, the following statements hold*:

1) $E(L_n) = 3 + (n-1)p$ *and* $V(L_n) = (n-1)(1-p)p$.
2) $M_{L_n}(t) = (1 - p + pe^t)^{n-1} e^{3t}$, $t \in \mathbb{R}$.
3) *For each* $k \in \mathbb{R}$, $\frac{L_n - 3 - (n-1)p}{\sqrt{p(1-p)(n+k)}} \xrightarrow{D} N(0, 1)$ *as* $n \to \infty$.

2.2 A class of RSTs that evolves in a preferential attachment manner

In a very recent article (Ren et al., 2022), the authors inspired by the seminal paper (Barabási and Albert, 1999) introduced a class of RSTs that evolves in a preferential attachment manner as follows. At time 1, an RST starts with a seed graph containing a centroid of degree 3 and 3 leaves. At each subsequent point, the probability of a qualified node recruiting a newcomer is proportional to its degree. If the centroid is selected, a new leaf is attached to it; if a leaf is selected, a new leaf is attached to the (selected) leaf, and the recruiter is converted to an internal node. Consequently, for $n \geq 2$,

$$P(I_{v,n}) = \frac{\deg_{v,n-1}}{\sum_{u \in Q_{n-1}} \deg_{u,n-1}}$$

where v is a qualified node at time n, $I_{v,n}$ indicates the event that node v is chosen as recruiter at time n, $\deg_{i,n-1}$ is the degree of a node i at time $n-1$ and Q_{n-1} denotes the set of qualified nodes at time $n-1$. Then, for $n \geq 2$, it follows that:

1) The probability that the centroid recruits a newcomer at time n is $\frac{L_{n-1}}{2L_{n-1}} = \frac{1}{2}$.
2) The probability that a leaf recruits a newcomer at time n is $\frac{1}{2L_{n-1}}$.

Therefore, we can conclude that the class of RSTs that evolves in a preferential attachment manner (preferential model) is the model presented in Section 2 with $p = \frac{1}{2}$.

3 Topological indices

The purpose of topological indices is to study the structural properties associated with a graph and its invariants using a certain numerical value. The idea of capturing the information in numerical form is to be able to compare the graphs according to the property to be studied. Let $G = (V, E)$, then many important topological indices $(\mathrm{TI}(G))$ can be defined as follows:

$$\mathrm{TI}(G) = \sum_{v \in V} h(\deg_v)^\alpha \tag{1}$$

where $\alpha \in \mathbb{R}$, $h : \{1,2,\ldots\} \to (0,\infty)$, and $\deg_v$ is the degree of a node v. In Section 3, we will study the indices that satisfy Eq. 1 in the model introduced in Section 2. At each stage, the generated tree has three types of nodes, centroid, leaves, and internal, for which their degrees are L_n, 1 and 2, respectively.

Proposition 2. *Let* TI_n *be the value of the topological index at stage* n. *For each* $n \geq 1$, *we have*

$$E(\mathrm{TI}_n) = E(h(L_n)^\alpha) + (h(1)^\alpha - h(2)^\alpha)E(L_n) + h(2)^\alpha(n+2)$$
$$V(TI_n) = V(h(L_n)^\alpha + (h(1)^\alpha - h(2)^\alpha)L_n).$$

Proof. Note that $I_n + L_n + 1 = n + 3$, where I_n is the number of internal nodes in the tree at stage n, it follows that:

$$\begin{aligned}\mathrm{TI}_n &= h(L_n)^\alpha + h(1)^\alpha L_n + h(2)^\alpha(n + 2 - L_n)\\ &= h(L_n)^\alpha + (h(1)^\alpha - h(2)^\alpha)L_n + h(2)^\alpha(n+2).\end{aligned} \tag{2}$$

By Eq. 2, we immediately get the mean and the variance of TI_n. □

Proposition 3. *Let* $n \geq 1, t \in \mathbb{R}$. *If* $h(L_n) = aL_n + b$ *with* $a, b \in \mathbb{R}$, *then* $M_{h(L_n)}(t) = (1 - p + pe^{at})^{n-1}e^{(3a+b)t}$.

3.1 Generalized Zagreb index

Zagreb index was introduced by chemists Gutman and Trinajstić (1972). Later, some of its general mathematical properties were pointed out and its relationship with other quantities of interest in chemical graph theory was shown (Gutman and Das, 2004). In fact, Zagreb index and its variants have been used in the studies of quantitative structure-property/activity relationships (QSPR/QSAR) (Devillers and Balaban, 1999; Khadikar et al., 2001; Sardana and Madan, 2002), while the overall Zagreb indices exhibited a potential applicability for deriving multilinear regression models. Nowadays, as an indicator of its importance, the ideas outlined in the initial paper are explored by numerous other scholars (An, 2022; Filipovski, 2021; Milovanović et al., 2021).

At time $n \geq 1$, taking $h(x) = x$ and $\alpha \in \mathbb{R}$ in Eq. 1, we obtain the generalized Zagreb index (Z_n^g). According to Eq. 2,

$$Z_n^g = L_n^\alpha + (1 - 2^\alpha)L_n + 2^\alpha(n+2). \tag{3}$$

Proposition 4. *Let* $\alpha \in \{1,2,\ldots\}$ *and* $t \in \mathbb{R}$, *we have*

$$\frac{\mathrm{d}^\alpha M_{L_n}}{\mathrm{d}t}(t) = \sum_{i=1}^{\alpha} C_{\alpha,i}p^i e^{it}\frac{\mathrm{d}^i M_{L_n}}{\mathrm{d}u}(u(t))$$

with $C_{\alpha,\alpha} = C_{\alpha,1} = 1$ and $C_{\alpha,i} = C_{\alpha-1,i-1} + iC_{\alpha-1,i}$ for $i \in \{2, 3, \ldots, \alpha - 1\}$.

Proof. We will get the proof via mathematical induction on α. First, observe that the point (2) in Proposition 1 may be simplified by defining a new variable $u(t) = 1 - p + pe^t$. Substituting $1 - p + pe^t$ by $u(t)$, we get $M_{L_n}(t) = M_{L_n}(u(t)) = u(t)^{n-1}\left(\frac{u(t)+p-1}{p}\right)^3$. Thus, for the base, that is $\alpha = 3$, we have

$$\frac{\mathrm{d}^3 M_{L_n}}{\mathrm{d}t}(t) = p^3 e^{3t}\frac{\mathrm{d}^3 M_{L_n}}{\mathrm{d}u}(u(t)) + 3p^2 e^{2t}\frac{\mathrm{d}^2 M_{L_n}}{\mathrm{d}u}(u(t)) + pe^t\frac{\mathrm{d}M_{L_n}}{\mathrm{d}u}(u(t))$$

with $C_{3,1} = C_{3,3} = 1$ and $C_{3,2} = C_{2,1} + 2C_{2,2} = 3$. We assume that the statement holds for all α, i.e.,

$$\frac{\mathrm{d}^\alpha M_{L_n}}{\mathrm{d}t}(t) = \sum_{i=1}^{\alpha} C_{\alpha,i} p^i e^{it}\frac{\mathrm{d}^i M_{L_n}}{\mathrm{d}u}(u(t)) \tag{4}$$

with $C_{\alpha,\alpha} = C_{\alpha,1} = 1$ and $C_{\alpha,i} = C_{\alpha-1,i-1} + iC_{\alpha-1,i}$ for $i \in \{2, 3, \ldots, \alpha - 1\}$. Note that for each $i \in \{1, 2, \ldots, \alpha\}$ we obtain that:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(C_{\alpha,i} p^i e^{it}\frac{\mathrm{d}^i M_{L_n}}{\mathrm{d}u}(u(t))\right) = iC_{\alpha,i} p^i e^{it}\frac{\mathrm{d}^i M_{L_n}}{\mathrm{d}u}(u(t)) + C_{\alpha,i} p^{i+1} e^{(i+1)t}\frac{\mathrm{d}^{i+1} M_{L_n}}{\mathrm{d}u}(u(t)). \tag{5}$$

By Eqs. 4 and 5, we have proved the following result:

$$\frac{\mathrm{d}^{\alpha+1} M_{L_n}}{\mathrm{d}t}(t) = \sum_{i=1}^{\alpha+1} C_{\alpha+1,i} p^i e^{it}\frac{\mathrm{d}^i M_{L_n}}{\mathrm{d}u}(u(t))$$

with $C_{\alpha+1,\alpha+1} = C_{\alpha+1,1} = 1$ and $C_{\alpha+1,i} = C_{\alpha,i-1} + iC_{\alpha,i}$ for $i \in \{2, 3, \ldots, \alpha\}$, which completes the proof. □

A special case of Proposition 4 has the following result, which is valid when $t = 0$ in Eq. 4.

Corollary 1. *For $\alpha \in \{1,2,\ldots\}$, it is verified that*

$$\frac{\mathrm{d}^\alpha M_{L_n}}{\mathrm{d}t}(0) = \sum_{i=1}^{\alpha} C_{\alpha,i} p^i \frac{\mathrm{d}^i M_{L_n}}{\mathrm{d}u}(u(0))$$

with $C_{\alpha,\alpha} = C_{\alpha,1} = 1$ and $C_{\alpha,i} = C_{\alpha-1,i-1} + iC_{\alpha-1,i}$ for $i \in \{2, 3, \ldots, \alpha - 1\}$.

Theorem 1. *For $n \geq 1$, $p \in (0,1)$ and $\alpha \in \{1,2,\ldots\}$, the following identity holds*:

$$\frac{\mathrm{d}^\alpha M_{L_n}}{\mathrm{d}t}(0) = p^\alpha n^\alpha + \frac{\alpha}{2}(\alpha(1-p) - p + 5)p^{\alpha-1}n^{\alpha-1} + O(n^{\alpha-2}).$$

Proof. First, observe that

$$M_{L_n}(u(t)) = \frac{1}{p^3}(u(t)^{n+2} + 3(p-1)u(t)^{n+1} + 3(p-1)^2 u(t)^n + (p-1)^3 u(t)^{n-1}),\ t \in \mathbb{R}.$$

Then, for each

$$i \in \{1,2,\ldots\},$$

we obtain

$$\frac{\mathrm{d}^i M_{L_n}}{\mathrm{d}u}(u(0)) = \sum_{k=-1}^{2}\frac{b_k}{p^3}(n+k)(n+k-1)\ldots(n+k-(i-1))u(0)^{n+k-i}$$

with $b_{-1} = (p-1)^3$, $b_0 = 3(p-1)^2$, $b_1 = 3(p-1)$ and $b_2 = 1$. Accordingly, it follows that:

$$\frac{\mathrm{d}^i M_{L_n}}{\mathrm{d}u}(1) = \frac{1}{p^3}\sum_{k=-1}^{2} b_k n^i + \frac{1}{2p^3}\sum_{k=-1}^{2} b_k i(2k+1-i)n^{i-1} + O(n^{i-2}) = n^i - \frac{i(ip+p-6)}{2p}n^{i-1} + O(n^{i-2}).$$

By Corollary 1, it is concluded that for each $\alpha \in \{1,2,\ldots\}$,

$$\frac{\mathrm{d}^\alpha M_{L_n}}{\mathrm{d}t}(0) = p^\alpha n^\alpha - \frac{\alpha}{2p}(\alpha p + p - 6)p^\alpha n^{\alpha-1} + C_{\alpha,\alpha-1}p^{\alpha-1}n^{\alpha-1} + O(n^{\alpha-2}).$$

As

$$C_{\alpha,\alpha-1} = C_{\alpha-1,\alpha-2} + (\alpha-1)C_{\alpha-1,\alpha-1} = C_{\alpha-1,\alpha-2} + \alpha - 1 = \frac{\alpha(\alpha-1)}{2}.$$

Then,

$$\frac{\mathrm{d}^\alpha M_{L_n}}{\mathrm{d}t}(0) = p^\alpha n^\alpha + \frac{\alpha}{2}(\alpha(1-p) - p + 5)p^{\alpha-1}n^{\alpha-1} + O(n^{\alpha-2}).$$

□

In consequence, by Theorem 1, the first two moments of Z_n^g for $\alpha \in \{3,4,\ldots\}$ are given by

$$E(Z_n^g) = p^\alpha n^\alpha + \frac{\alpha}{2}(\alpha(1-p) - p + 5)p^{\alpha-1}n^{\alpha-1} + O(n^{\alpha-2})$$

and

$$E((Z_n^g)^2) = p^{2\alpha}n^{2\alpha} + \alpha(2\alpha(1-p) - p + 5)p^{2\alpha-1}n^{2\alpha-1} + O(n^{2\alpha-2}).$$

Then,

$$V(Z_n^g) = \alpha^2(1-p)p^{2\alpha-1}n^{2\alpha-1} + O(n^{2\alpha-2}).$$

Theorem 2. *For any $\alpha \in \{3,4,\ldots\}$, it is verified that $\frac{Z_n^g}{n^\alpha} \xrightarrow{P} p^\alpha$ when n goes to infinity.*

Proof. Let

$$X_n = \frac{1}{n^\alpha}\left(Z_n^g - \frac{\alpha}{2}(\alpha(1-p) - p + 5)p^{\alpha-1}n^{\alpha-1} - O(n^{\alpha-2})\right).$$

By Chebyshev's inequality (Gut, 2005), we get

$$P(|\, X_n - p^\alpha \,| \geq \epsilon) \leq \frac{1}{n^{2\alpha}\epsilon^2}(\alpha^2(1-p)p^{2\alpha-1}n^{2\alpha-1} + O(n^{2\alpha-2}))$$

for any $\varepsilon > 0$. If $n \to \infty$, then

$$\frac{1}{n^{2\alpha}\epsilon^2}(\alpha^2(1-p)p^{2\alpha-1}n^{2\alpha-1} + O(n^{2\alpha-2})) \to 0$$

so $X_n \xrightarrow{p} p^\alpha$. On the other hand, $X_n = \frac{Z_n^g}{n^\alpha} - x_n$ with

$$x_n = \frac{1}{n^\alpha}\left(\frac{\alpha}{2}(\alpha(1-p) - p + 5)p^{\alpha-1}n^{\alpha-1}O(n^{\alpha-2})\right). \qquad \square$$

Then, $x_n \to 0$ as $n \to \infty$. Therefore, the result follows.

Corollary 2. *For any $\alpha \in \{3,4,\ldots\}$ and $r > 0$, it is verified that $\frac{Z_n^g}{n^\alpha} \xrightarrow{L_r} p^\alpha$ when n goes to infinity.*

Proof. Using Theorem 2, we have $\frac{Z_n^g}{n^\alpha} \xrightarrow{p} p^\alpha$ when n goes to infinity. On the other hand, $\frac{Z_n^g}{n^\alpha} \geq 0$ for all $n \geq 1$. For each $r > 0$, there exists $N \in \{1,2,\ldots\}$ such that $N > r$. By Theorem 4.2 from Chapter 5 of Gut (2005), we obtain $\left\{\left(\frac{Z_n^g}{n^\alpha}\right)^r, n \geq 1\right\}$ which is uniformly integrable. Then, applying Theorem 5.4 from Chapter 5 of Gut (2005) with $X_n = \frac{Z_n^g}{n^\alpha}$, we obtain the convergence in r-mean and the proof is completed. $\square$

3.2 Zagreb index

At time $n \geq 1$, taking $h(x) = x$ and $\alpha = 2$ in Eq. 1, we obtain the Zagreb index (Z_n). According to Eq. 3, we have

$$Z_n = L_n^2 - 3L_n + 4(n+2). \tag{6}$$

We can obtain the moments of Z_n by Eq. 6. Clearly, for $n \geq 1$,

$$E(Z_n) = n^2p^2 + (-3p^2 + 4p + 4)n + 2p^2 - 4p + 8.$$

$$\begin{aligned} V(Z_n) = {} & (-4p^4 + 4p^3)n^3 + (22p^4 - 40p^3 + 18p^2)n^2 \\ & + (-38p^4 + 92p^3 - 70p^2 + 16p)n + 20p^4 - 56p^3 \\ & + 52p^2 - 16p. \end{aligned}$$

Proposition 5. *For $p \in (0,1)$, when n goes to infinity, it is verified that*

1) *For all $k \in \mathbb{R}$, $\frac{Z_n - n^2p^2}{2\sqrt{p^3(1-p)(n+k)^3}} \xrightarrow{D} N(0,1)$.*
2) *For all $r > 0$, $\frac{Z_n}{n^2} \xrightarrow{L_r} p^2$.*

Proof. (1) By Eq. 6, $Z_n = \left(L_n - \frac{3}{2}\right)^2 + \frac{16n+23}{4}$ for each $n \geq 1$. Note that point (3) in Proposition 1 implies that $\frac{L_n - \frac{3}{2}}{\sqrt{p(1-p)(n+k)}}$ is equivalent to a normal random variable with mean $\frac{\frac{3}{2} + (n-1)p}{\sqrt{p(1-p)(n+k)}}$ and variance 1 in distribution. Therefore, $\left(\frac{L_n - \frac{3}{2}}{\sqrt{p(1-p)(n+k)}}\right)^2 \sim \chi^2\left(\frac{\left(\frac{3}{2} + (n-1)p\right)^2}{p(1-p)(n+k)}, 1\right)$, indicating that

$$\frac{Z_n}{p(1-p)(n+k)} - \frac{16n+23}{4p(1-p)(n+k)} \sim \chi^2\left(\frac{\left(\frac{3}{2} + (n-1)p\right)^2}{p(1-p)(n+k)}, 1\right).$$

By the well-known normal approximation of non-central chi-squared distribution (Severo and Zelen, 1960), it is obtained that

$$\frac{\frac{Z_n}{p(1-p)(n+k)} - \frac{16n+23}{4p(1-p)(n+k)} - \left(1 + \frac{\left(\frac{3}{2} + (n-1)p\right)^2}{p(1-p)(n+k)}\right)}{\sqrt{2\left(1 + \frac{2\left(\frac{3}{2} + (n-1)p\right)^2}{p(1-p)(n+k)}\right)}} \xrightarrow{D} N(0,1)$$

as $n \to \infty$. In particular,

$$\begin{aligned} \frac{\frac{Z_n}{p(1-p)(n+k)} - \frac{16n+23}{4p(1-p)(n+k)} - 1 - \frac{\left(\frac{3}{2} + (n-1)p\right)^2}{p(1-p)(n+k)}}{\sqrt{2\left(1 + \frac{2\left(\frac{3}{2} + (n-1)p\right)^2}{p(1-p)(n+k)}\right)}} &= \frac{Z_n - \frac{16n+23}{4} - p(1-p)(n+k) - \left(\frac{3}{2} + (n-1)p\right)^2}{\sqrt{p(1-p)(n+k)\sqrt{2\left(p(1-p)(n+k) + 2\left(\frac{3}{2} + (n-1)p\right)^2\right)}}} \\ &= \frac{Z_n - \frac{16n+23}{4} - p(1-p)(n+k) - \left(\frac{3}{2} + (n-1)p\right)^2}{\sqrt{4p^3(1-p)(n+k)^3 + o(n^3)}} \\ &= a_n \frac{Z_n - n^2p^2}{\sqrt{4p^3(1-p)(n+k)^3}} + b_n \end{aligned}$$

where

$$a_n = \frac{\sqrt{4p^3(1-p)(n+k)^3}}{\sqrt{4p^3(1-p)(n+k)^3 + o(n^3)}}$$

$$b_n = \frac{o\left(n^{\frac{3}{2}}\right)}{\sqrt{4p^3(1-p)(n+k)^3 + o(n^3)}}.$$

It is verified that $a_n \to 1$ and $b_n \to 1$ as $n \to \infty$, then

$$\frac{Z_n - n^2p^2}{2\sqrt{p^3(1-p)(n+k)^3}} \xrightarrow{D} N(0,1). \qquad \square$$

(2) The proof can be verified similar to that of Corollary 2.

We conduct a numerical experiment to verify point (1) of Proposition 5 with $k = 0$, developed in this section. Given a fixed $p\epsilon(0,1)$, we independently generate 5,000 replications of RSTs after $n = 10{,}000$ evolutionary steps. For each simulated RST, its Zagreb index is computed, then the sample data are formed by 5,000 Zagreb indices from independent simulated graphs. The histogram of the sample data with a normal approximation curve is given in Figure 1 for $p = 0.3$, 0.5, and 0.7, respectively. We further confirm the conclusion via the Shapiro–Wilk normality test, which yields that the p-value equals 0.070, 0.365, and 0.469 for $p = 0.3$, 0.5, and 0.7, respectively.

3.2.1 Gordon-Scantlebury index

Defining S_n as the Gordon-Scantlebury index at time $n \geq 1$, which verifies that $Z_n = 2(S_n + E_n)$ (Nikolić et al., 2003) where E_n is the number of edges at time n. The tree generated by the model at time n has $n + 3$ nodes and $n + 2$ edges, thus $S_n = \frac{Z_n}{2} - n - 2$. For $n \geq 1$, we get the following proposition:

Proposition 6. *For $p \in (0,1)$, it is verified that*

1) $E(S_n) = \frac{n^2p^2}{2} + \left(-\frac{3}{2}p^2 + 2p + 1\right)n + p^2 - 2p + 2.$

2) $V(S_n) = (-p^4 + p^3)n^3 + \left(\frac{11}{2}p^4 - 10p^3 + \frac{9}{2}p^2\right)n^2 + \left(-\frac{19}{2}p^4 + 23p^3 - \frac{35}{2}p^2 + 4p\right)n + 5p^4 - 14p^3 + 13p^2 - 4p.$

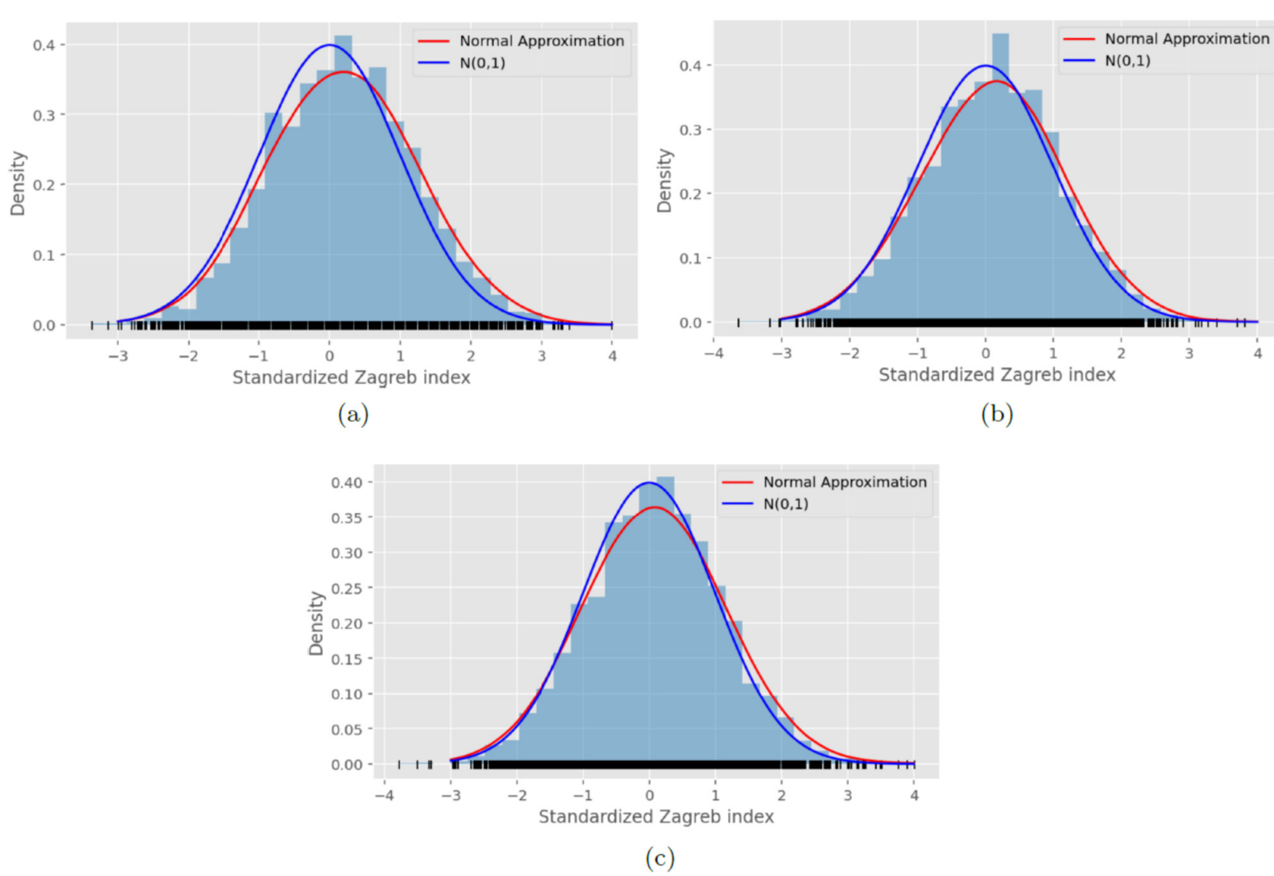

Figure 1: Histogram of the standardized Zagreb indices of 5,000 independently generated RSTs with $n = 10{,}000$ taking $p = 0.3$, 0.5 and 0.7 in (a–c) respectively; the thick red curve is the estimated density of the sample.

3) *For all* $k \in \mathbb{R}$, $\frac{S_n - \frac{n^2p^2}{2}}{\sqrt{p^3(1-p)(n+k)^3}} \xrightarrow{D} N(0,1)$ *when* n *goes to infinity.*
4) *For all* $r > 0$, $\frac{S_n}{n^2} \xrightarrow{L_r} \frac{p^2}{2}$ *when* n *goes to infinity.*

3.2.2 Platt index

Let P_n denote the Platt index at time $n \geq 1$, which verifies that $P_n = 2S_n$ (Nikolić et al., 2003). Thus, we obtain the following proposition:

Proposition 7. *For* $p \in (0,1)$, *it is verified that*
1) $E(P_n) = n^2p^2 + (-3p^2 + 4p + 2)n + 2p^2 - 4p + 4$.
2) $V(P_n) = V(Z_n)$.
3) *For all* $k \in \mathbb{R}$, $\frac{P_n - n^2p^2}{2\sqrt{p^3(1-p)(n+k)^3}} \xrightarrow{D} N(0,1)$ *when* n *goes to infinity.*
4) *For all* $r > 0$, $\frac{P_n}{n^2} \xrightarrow{L_r} p^2$ *when* n *goes to infinity.*

3.2.3 Forgotten index

At time $n \geq 1$, taking $h(x) = x$ and $\alpha = 3$ in Eq. 1, we obtain the forgotten index (F_n). According to Eq. 3, we have $F_n = L_n^3 - 7L_n + 8(n+2)$. Our next task is to calculate the first moment of F_n, and consequently to get the variance of F_n:

$$E(F_n) = n^3p^3 + (12p^2 - 6p^3)n^2 + (11p^3 - 36p^2 + 30p + 8)n - 6p^3 + 24p^2 - 30p + 22.$$

$$\begin{aligned} V(F_n) = {} & 9p^5(1-p)n^5 - 9p^4(1-p)(13p - 18)n^4 \\ & + 3p^3(1-p)(197p^2 - 490p + 302)n^3 \\ & - 9p^2(1-p)(159p^3 - 530p^2 + 572p - 192)n^2 \\ & - 6p(1-p)^2(272p^3 - 803p^2 + 714p - 150)n \\ & + 36p(1-p)^2(19p^3 - 64p^2 + 71p - 25). \end{aligned}$$

According to Corollary 2, we obtain the following result:

Corollary 3. *For all* $r > 0$, *it is verified that* $\frac{F_n}{n^3} \xrightarrow{L_r} p^3$ *as* $n \to \infty$.

3.3 Degree-based Gini index

Recently, a degree-based Gini index for general graphs was proposed by Domicolo and Mahmoud (2020). This index is a topological measure on a graph capturing the proximity to regular graphs. Ren et al. (2022) considered the degree-based Gini index introduced by Domicolo and Mahmoud (2020), with slight modifications. In this section, we will study the degree-based Gini index defined by Ren et al. (2022). By definition, the degree-based Gini index of a graph within the class of RSTs at time $n \geq 1$ is given by

$$G_n = \frac{\sum_{i<j \in V_n} |\deg_i - \deg_j|}{(n+3)^2 E(\deg_{v^*})}$$

where v^* is an arbitrary node of a randomly selected graph from the class of RSTs and V_n denotes the node set at time n. We take $E(G_n)$ as the degree-based Gini index of the class. Due to the characteristics of the model, we get

$$\begin{aligned} \sum_{i,j \in V_n} |\deg_i - \deg_j| = {} & |\, L_n - 1 \,|\, L_n + |L_n - 2|(n + 2 - L_n) \\ & + L_n(n + 2 - L_n) = -L_n^2 + (2n+5)L_n \\ & - 2n - 4 \end{aligned}$$

since $L_n \geq 3$ for all $n \geq 1$. Finally,

$$E(d_{v^*}) = \frac{2(n+2)}{n+3}.$$

Thus,

$$G_n = \frac{-L_n^2 + (2n+5)L_n - 2n - 4}{2(n+3)(n+2)}. \tag{7}$$

It follows that

$$E(G_n) = \frac{(2p - p^2)n^2 + (3p^2 - 4p + 4)n - 2p^2 + 2p + 2}{2(n+3)(n+2)}.$$

Next we get an asymptotic property of the degree-based Gini index of the class of RSTs at time n.

Proposition 8. *As* $n \to \infty$, *we have* $E(G_n) \to \frac{p(2-p)}{2}$.

We see from Eq. 7 *that*

$$G_n = \frac{-\left(L_n - \frac{2n+5}{2}\right)^2 + \left(\frac{2n+5}{2}\right)^2 - 2n - 4}{2(n+3)(n+2)}.$$

Therefore,

$$V(G_n) = \frac{4p(1-p)^3n^3 + 2p(11p - 6)(1-p)^2n^2 + 2p(1-p)(19p^2 - 23p + 6)n - 4p(1-p)(5p^2 - 5p + 1)}{4(n+3)^2(n+2)^2}.$$

Corollary 4. *It is verified that* $G_n \xrightarrow{P} \frac{p(2-p)}{2}$, *as* $n \to \infty$.

Proof. By Chebyshev's inequality (Gut, 2005), we have

$$P(|G_n - E(G_n)| \geq \epsilon) \leq \frac{V(G_n)}{\epsilon^2}$$

for any $\epsilon > 0$. If $n \to \infty$ then $\frac{V(G_n)}{\epsilon^2} \to 0$, so $G_n - E(G_n) \xrightarrow{P} 0$. Therefore, Proposition 8 completes the proof. □

Corollary 5. *For all* $r > 0$, *we have* $G_n \xrightarrow{L_r} \frac{p(2-p)}{2}$, *when* n *goes to infinity*.

Proof. Argued in a similar manner to Corollary 2 by Corollary 4, the result follows. □

Remark 1

a) In view of Proposition 8 and Corollary 4, we define $f(p) = \frac{p(2-p)}{2}$, $p \in (0,1)$. Note that f is strictly increasing, consequently, if we want a more regular class we must choose smaller values of p, since Domicolo and Mahmoud (2020) showed that a smaller value of degree-based Gini index suggests more regularity of a graph or a class of graphs, which makes sense in this case since the center would have a lower degree. Specifically, for $p \in \left(0, \frac{1}{2}\right)$, we have that the class of RSTs that evolves in a preferential attachment manner is relatively less regular than the class of RSTs studied in this work (Figure 2).
b) Domicolo and Mahmoud (2020) compared the regularity of two classes of binary trees. The authors showed that the class of uniform binary trees has degree-based Gini index $\frac{3}{16}$ (Domicolo and Mahmoud, 2020; Section 6.1), whereas that of the class of binary search trees is $\frac{2}{9}$ (Domicolo and Mahmoud, 2020; Section 6.2). Note that for $p \leq 1 - \frac{\sqrt{10}}{4} \approx 0.209$, we have $f(p) \leq \frac{3}{16} < \frac{2}{9}$, then the class of uniform binary and binary search trees are relatively less regular than the class of RSTs studied in this work for $p \in \left(0, 1 - \frac{\sqrt{10}}{4}\right)$ (Figure 2).
c) Zhang and Wang (2022) concluded that the degree-based Gini index for the class of random caterpillars is $\frac{1}{2}$. Since $f(p) < \frac{1}{2}$ for $p \in (0,1)$, we conclude that the class of RSTs is more regular than the class of random caterpillars of Zhang and Wang (2022) (Figure 2).

3.4 Degree-based Hoover index

Zhang and Wang (2022) proposed a degree-based Hoover index for graphs analogous to the degree-based Gini index introduced by Domicolo and Mahmoud (2020) as a competing measure for assessing graph regularity. In our context, at time $n \geq 1$, the degree-based Hoover index of a graph within the class of RSTs (H_n) is defined as follows:

$$H_n = \frac{\sum_{i \in V_n} |(n+3)\deg_i - 2(n+2)|}{4(n+2)(n+3)}$$

where V_n denotes the node set at time n. In a similar way, we take $E(H_n)$ as the degree-based Hoover index of the class. The same analysis applied in Section 3.2 is used in this section and we obtain the following results.

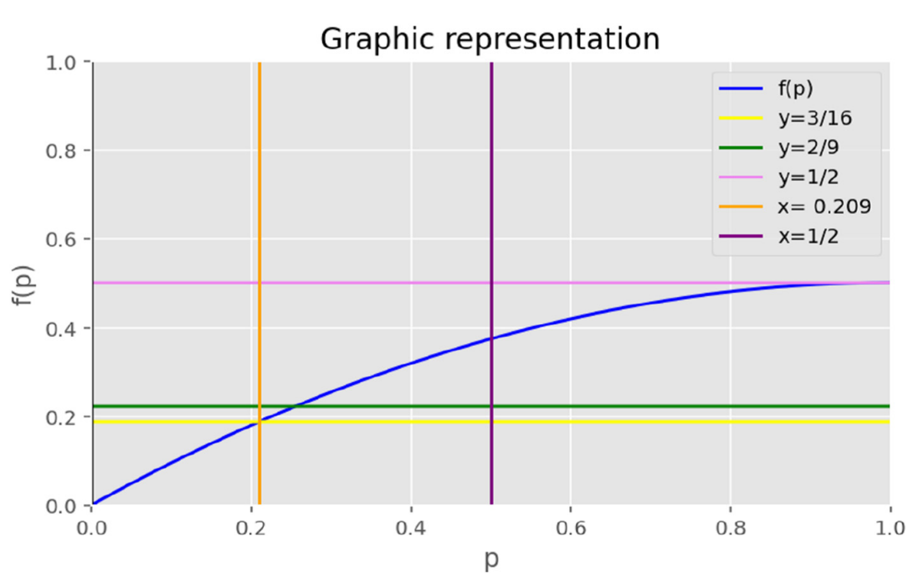

Figure 2: Graphic representation of functions in Remark 1.

Proposition 9. *For $n \geq 1$, we have*

$$H_n = \frac{(n+1)L_n}{2(n+3)(n+2)},\ E(H_n) = \frac{pn^2 + 3n + 3 - p}{2(n+3)(n+2)},$$
$$V(H_n) = \frac{p(1-p)(n^2-1)(n+1)}{4(n+3)^2(n+2)^2}.$$

Finally, for all $r > 0$, $H_n \xrightarrow{L_r} \frac{p}{2}$, as $n \to \infty$.

A direct consequence of Proposition 9 is the following corollary.

Corollary 6. *In the preferential model for all $r > 0$, it is verified that $H_n \xrightarrow{L_r} \frac{1}{4}$ and $E(H_n) \to \frac{1}{4}$ when n goes to infinity.*

Remark 2

a) In view of Proposition 9 we define $f_1(p) = \frac{p}{2}$, $p \in (0,1)$, then f_1 is strictly increasing. Since it is shown by Zhang and Wang (2022) that a value closer to 0 suggests that the graphs in the class tend to be more regular, by an argument similar to Remark 1a we get the same behavior as the degree-based Gini index studied in Section 3.2.
b) Moreover, it is observed that $f_1(p) < f(p)$ for $p \in (0,1)$. Then, the degree-based Hoover index of the class of RSTs presented in Section 2 is less than the degree-based Gini index of the same class when n goes to infinity.
c) Zhang and Wang (2022) (Section 3) concluded that the degree-based Hoover index for the class of random caterpillars is $\frac{1}{2}$ as $n \to \infty$. Since $f_1(p) < \frac{1}{2}$ for $p \in (0,1)$, we conclude that the class of RSTs is more regular than the class of random caterpillars of Zhang and Wang (2022), in these cases via the degree-based Hoover index.

4 Conclusion

We investigated a class of RSTs, the random variable of prime interest is the number of leaves as time proceeds and we calculated the moment generating function of the leaves and showed that the number of leaves follow a Gaussian law asymptotically. Next we investigated several useful topological indices for this class, including degree-based Gini index, degree-based Hoover index, generalized Zagreb index, and other indices associated with these. Moreover, Proposition 3 and Theorem 1 showed by Ren et al. (2022) are deduced from Proposition 1 of this work taking $p = 1/2$. In similar way, the results displayed in Sections 3.2.1 and 3.2.2 of Ren et al. (2022) are obtained as a special case of the results demonstrated in Sections 3.1 and 3.2, respectively. In particular, we conclude that the class of RSTs that evolves in a preferential attachment manner is relatively less regular than the class of RSTs studied in this work in some cases.

Acknowledgments: All three authors would like to thank three anonymous referees for their valuable comments that definitely helped improve the quality of this article.

Funding information: Authors state no funding involved.

Author contributions: Saylé Sigarreta: conceived the presented idea, developed the theory, and wrote the article; Hugo Cruz-Suárez: conceived the presented idea and verified the analytical methods; Saylí Sigarreta: performed the computations and wrote the article.

Conflict of interest: Authors state no conflict of interest.

Data availability statement: Authors declare that this article has been developed without any data.

References

Abbas G., Ibrahim M., Ahmad A., Azeem M., Elahi K., M-polynomials and associated topological indices of sodalite materials. Math. Probl. Eng., 2021a, 2021, 1–10.

Abbas G., Ibrahim M., Ahmad A., Azeem M., Elahi K., M-polynomials of tetra-cyanobenzene transition metal structure. Polycycl. Aromat. Comp., 2021b, 2021, 1–11.

Afzal Siddiqui H.M., Nadeem M.F., Azeem M., Arshad M.A., Haider A., Malik M.A., Topological properties of supramolecular chain of different complexes of N-Salicylidene-L-Valine. Polycycl. Aromat. Comp., 2021, 42, 1–14.

Aguilar-Sánchez R., Méndez-Bermúdez J.A., Rodríguez J.M., Sigarreta J.M., Normalized sombor indices as complexity measures of random networks. Entropy, 2021, 23(8), 976.

Ahmad A.-N., Ahmad A., Azeem M., Computation of edge- and vertex-degree-based topological indices for tetrahedral sheets of clay minerals. Main Group Met. Chem., 2022, 45(1), 26–34.

Alatawi M.S., Ahmad A., Koam A.N., Husain S., Azeem M., Edge weight-based entropy of magnesium iodide graph. J. Math., 2021, 2021, 1–7.

An M., The first Zagreb index, reciprocal degree distance and Hamiltonian-connectedness of graphs. Inform. Process. Lett., 2022, 176, 106247.

Barabási A.-L., Albert R., Emergence of scaling in random networks. Science, 1999, 286(5439), 509–512.

Devillers J., Balaban A., Topological Indices and Related Descriptors in QSAR and QSPR. Gordon and Breach Science Publishers, Singapore, 1999.

Domicolo C., Mahmoud H., Degree-based Gini index for graphs. Probab. Eng. Inform. Sc., 2020, 34(2), 157–171.

Ducoffe G., Marinescu-Ghemeci R., Obreja C., Popa A., Tache R. M., Extremal graphs with respect to the modified first Zagreb connection index. Proceedings of the 16th Cologne-Twente Workshop on Graphs and Combinatorial Optimization, CNAM Paris, France, June 18-20, 2018, 65–68.

Estrada E., Quantifying network heterogeneity. Phys. Rev. E, 2010, 82(6), 066102.

Filipovski S., New bounds for the first Zagreb index. MATCH-Commun. Math. Co., 2021, 85(2), 303–312.

Gut A., Probability: A graduate course. Springer, New York, 2005.

Gutman I., Das K.C., The first Zagreb index 30 years after. MATCH-Commun. Math. Co., 2004, 50(1), 83–92.

Gutman I., Furtula B., Katani V., Randić index and information. AKCE Int. J. Graphs Comb., 2018, 15(3), 307–312.

Gutman I., Trinajstić N., Graph theory and molecular orbitals. Total φ-electron energy of alternant hydrocarbons. Chem. Phys. Lett., 1972, 17(4), 535–538.

Hosoya H., Topological index. A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons. B. Chem. Soc. Jpn., 1971, 44(9), 2332–2339.

Imran M., Ahmad A., Ahmad Y., Azeem M., Edge weight based entropy measure of different shapes of carbon nanotubes. IEEE Access, 2021, 9, 139712–139724.

Kashif A., Aftab S., Javaid M., Awais H. M., M-polynomial-based topological indices of metal-organic networks. Main Group Met. Chem., 2021, 44(1), 129–140.

Kazemi R., Gordon-scantlebury and platt indices of random plane-oriented recursive trees. Math. Interdiscip. Res., 2021, 6(1), 1–10.

Khadikar P.V., Karmarkar S., Agrawal V.K., A novel PI index and its applications to QSPR/QSAR studies. J. Chem. Inf. Comp. Sci., 2001, 41(4), 934–949.

Kier L.B., Hall L.H., Molecular connectivity in structure-activity analysis. Research Studies, 1986.

Kier L.B., Hall L.H., Murray W.J., Randi M., Molecular connectivity I: Relationship to nonspecific local anesthesia. J. Pharm. Sci., 1975, 64(12), 1971–1974.

Li S., Shi L., Gao W., Topological indices computing on random chain structures. Int. J. Quantum Chem., 2021a, 121(8), 26589.

Li S., Shi L., Gao W., Two modified Zagreb indices for random structures. Main Group Met. Chem., 2021b, 44(1), 150–156.

Martínez-Martínez C., Méndez-Bermúdez J., Rodríguez J.M., Sigarreta J.M., Computational and analytical studies of the Randić index in Erdős–Rényi models. Appl. Math. Comput., 2020, 377, 125137.

Milovanović I., Matejić M., Milovanović E., Khoeilar R., A note on the first Zagreb index and coindex of graphs. Communications in Combinatorics and Optimization, 2021, 6(1), 41–51.

Nadeem M. F., Azeem M., Farman I., Comparative study of topological indices for capped and uncapped carbon nanotubes. Polycycl. Aromat. Comp., 2021, 42, 1–18.

Nikolić S., Kovačević G., Miličević A., Trinajstić N., The Zagreb indices 30 years after. Croat. Chem. Acta, 2003, 76(2), 113–124.

Pegu A., Deka B., Gogoi I.J., Bharali A., Two generalized topological indices of some graph structures. J. Math. Comput. Sci., 2021, 11(5), 5549–5564.

Pineda-Pineda J.J., Martínez-Martínez C.T., Méndez-Bermúdez J.A., Muñoz-Rojas J., Sigarreta J.M., Application of bipartite networks to the study of water quality. Sustainability, 2020, 12(12), 5143.

Randić M., Characterization of molecular branching. J. Am. Chem. Soc., 1975, 97(23), 6609–6615.

Rao Y., Kanwal A., Abbas R., Noureen S., Fahad A., Qureshi M.I., Some degree-based topological indices of caboxy-terminated dendritic macromolecule. Main Group Met. Chem., 2021, 44(1), 165–172.

Raza H., Nadeem M.F., Ahmad A., Asim M.A., Azeem M., Comparative study of valency based topological indices for tetrahedral sheets of clay minerals. Curr. Org. Synth., 2021, 18(7), 711–718.

Raza H., Waheed M., Jamil M.K., Azeem M., Structures devised by the generalizations of two graph operations and their topological descriptors. Main Group Met. Chem., 2022, 45(1), 44–56.

Reždepović I., Furtula B., Predictive potential of eigenvalue-based topological molecular descriptors. J. Comput. Aid. Mol. Des., 2020, 34, 975–982.

Ren Y., Zhang P., Dey D.K., Investigating several fundamental properties of random lobster trees and random spider trees. Methodol. Comput. Appl., 2022, 24(1), 431–447.

Sardana S., Madan A., Application of graph theory: Relationship of antimycobacterial activity of quinolone derivatives with eccentric connectivity index and Zagreb group parameters. MATCH-Commun. Math. Co., 2002, 45, 35–53.

Severo N.C., Zelen M., Normal approximation to the chi-square and non-central F probability functions. Biometrika, 1960, 47(3/4), 411–416.

Shao Z., Jahanbani A., Sheikholeslami S.M., Multiplicative topological indices of molecular structure in anticancer drugs. Polycycl. Aromat. Comp., 2022, 42(2), 475–488.

Shiu W.C., Extremal Hosoya index and Merrifield–Simmons index of hexagonal spiders. Discrete Appl. Math., 2008, 156(15), 2978–2985.

Wiener H., Structural determination of paraffin boiling points. J. Am. Chem. Soc., 1947, 69(1), 17–20.

Zhang P., Wang X., Several topological indices of random caterpillars. Methodol. Comput. Appl., 2022, 24(3), 1773–1789.

Zuo X., Nadeem M.F., Siddiqui M.K., Azeem M., Edge weight based entropy of different topologies of carbon nanotubes. IEEE Access, 2021, 9, 102019–102029.