**Transactions on Combinatorics**
ISSN (print): 2251-8657, ISSN (on-line): 2251-8665
Vol. 9 No. 3 (2020), pp. 161-169.

www.combinatorics.ir
www.ui.ac.ir

# HOSOYA INDEX OF TREE STRUCTURES

RAMIN KAZEMI* AND ALI BEHTOEI

Communicated by Behruz Tayfeh Rezaie

Abstract. The Hosoya index, also known as the $Z$ index, of a graph is the total number of matchings in it. In this paper, we study the Hosoya index of the tree structures. Our aim is to give some results on $Z$ in terms of Fibonacci numbers in such structures. Also, the asymptotic normality of this index is given.

## 1. Introduction

In chemical graph theory and in mathematical chemistry, a molecular graph or chemical graph is a representation of the structural formula of a chemical compound in terms of graph theory. A chemical graph is a labeled graph whose vertices correspond to the atoms of the compound and edges correspond to chemical bonds. Its vertices are labeled with the kinds of the corresponding atoms and edges are labeled with the types of bonds. For particular purposes any of the labelings may be ignored. A hydrogen-depleted molecular graph or hydrogen-suppressed molecular graph is the molecular graph with hydrogen vertices deleted. Molecular graphs can distinguish between structural isomers, compounds which have the same molecular formula but non-isomorphic graphs- such as isopentane and neopentane. On the other hand, the molecular graph normally does not contain any information about the three-dimensional arrangement of the bonds, and therefore cannot distinguish between conformational isomers (such as cis and trans 2-butene) or stereoisomers (such as D- and L-glyceraldehyde).

MSC(2010): Primary: 05C05; Secondary: 60F05.
Keywords: Hosoya index, tree structures, average, variance, asymptotic normality.
Received: 29 February 2020, Accepted: 10 June 2020.
*Corresponding author.

DOI: http://dx.doi.org/10.22108/toc.2020.121874.171

.

In some important cases (topological index calculation etc.) the following classical definition is sufficient: molecular graph is connected undirected graph one-to-one corresponded to structural formula of chemical compound so that vertices of the graph correspond to atoms of the molecule and edges of the graph correspond to chemical bonds between these atoms. One variant is to represent materials as infinite Euclidean graphs, in particular, crystals as periodic graphs.

In 1971 the Japanese chemist Haruo Hosoya introduced a molecular-graph based structure descriptor, which he named topological index and denoted by $Z$. The Hosoya index $H(G)$ of a graph $G$ is defined as the total number of independent edge subsets, that is, the total number of its matchings (see [8]). This topological index, also called the Hosoya $Z$ index [8], has found an extensive use in quantitative structure-property relationships (QSPRs) and quantitative structure-activity relationships (QSARs) (see [11]-[15] and references therein). In some cases such as in simple linear regressions based on topological indices the $Z$ index gives better QSPR and QSAR models than most indices [14]. It is also applicable to many different problems, not only in chemistry, but also in mathematics (e.g. combinatorial theory), chemoinformatics (e.g. coding and identification of molecules) and physics (e.g. dimer statistics).

Trees are defined as connected graphs without cycles, and their properties are basics of graph theory. For example, a connected graph is a tree, if and only if the number of edges equals the number of nodes minus 1. The path graph, star tree, caterpillar tree, lobster tree, d-ary tree and starlike tree are the most applicable types of tree structures in graph theory. The goal of this paper is to study of Hosoya index in such structures [6].

## 2. The main results

Let $F_n$ be the $n$-th Fibonacci number and hence, $F_0 = 1, F_1 = 1, F_2 = 2$. Note that in some of literatures it is assumed that $F_0 = 0$ and $F_1 = 1$ but for convenient, in this paper we assume that $F_0 = 1, F_1 = 1, F_2 = 2$ and so on.

**Lemma 2.1.** ([6, 16]) *If $G_1, G_2, \ldots, G_k$ are the connected components of a graph $G$, then $H(G) = \prod_{i=1}^{k} H(G_i)$.*

It is well known that among all $n$-vertex trees, the path $P_n$ has the maximum Hosoya index and the star $S_n$ has the minimum Hosoya index [6].

**Lemma 2.2.** [2] *We have $H(S_n) = n$ and $H(P_n) = F_n$ for each $n \in \mathbb{N}$.*

**Lemma 2.3.** *For each pair of positive integers $k, t$ with the condition $k \geq t$ we have*

$$F_{t-1}F_{k+1} - F_tF_k = (-1)^{t+1}F_{k-t}.$$

DOI: http://dx.doi.org/10.22108/toc.2020.121874.1713

*Proof.* By using the recurrence relation of the Fibonacci numbers we have

$$\begin{aligned} F_{t-1}F_{k+1} - F_tF_k &= F_{t-1}(F_k + F_{k-1}) - (F_{t-1} + F_{t-2})F_k \\ &= F_{t-1}F_{k-1} - F_{t-2}F_k \\ &= (F_{t-2} + F_{t-3})F_{k-1} - F_{t-2}(F_{k-1} + F_{k-2}) \\ &= F_{t-3}F_{k-1} - F_{t-2}F_{k-2}. \end{aligned}$$

Thus, by an inductive process we can see that

$$\begin{aligned} F_{t-1}F_{k+1} - F_tF_k &= F_{t-3}F_{k-1} - F_{t-2}F_{k-2} \\ &= F_{t-5}F_{k-3} - F_{t-4}F_{k-4} \\ &\vdots \\ &= \begin{cases} F_1F_{k-t+3} - F_2F_{k-t+2} & t \text{ is even} \\ F_0F_{k-t+2} - F_1F_{k-t+1} & t \text{ is odd} \end{cases} \\ &= \begin{cases} F_{k-t+3} - 2F_{k-t+2} & t \text{ is even} \\ F_{k-t+2} - F_{k-t+1} & t \text{ is odd} \end{cases} \\ &= \begin{cases} F_{k-t+1} - F_{k-t+2} & t \text{ is even} \\ F_{k-t} & t \text{ is odd} \end{cases} \\ &= \begin{cases} -F_{k-t} & t \text{ is even} \\ F_{k-t} & t \text{ is odd} \end{cases} \end{aligned}$$

□

**Theorem 2.4.** *For each pair of positive integers $k, t$ we have*

$$F_tF_k \leq F_2F_{t+k-2}.$$

*Proof.* Without lose of generality, assume that $k \geq t$. If $t = 1$, then

$$F_tF_k = F_k = F_{k-1} + F_{k-2} \leq 2F_{k-1} = F_2F_{k+t-2}.$$

DOI: http://dx.doi.org/10.22108/toc.2020.121874.1713

The case $t \in \{2,3\}$ is obvious and hence, assume that $t \geq 4$. By Lemma 2.3 we obtain

$$\begin{aligned}
F_tF_k &= F_{t-1}F_{k+1} + (-1)^t F_{k-t}\\
&= \big(F_{t-2}F_{k+2} + (-1)^{t-1}F_{(k-1)-(t-1)}\big) + (-1)^t F_{k-t}\\
&= F_{t-2}F_{k+2} + (-1)^{t-1}F_{k-t+2} + (-1)^t F_{k-t}\\
&= F_{t-3}F_{k+3} + (-1)^{t-2}F_{k-t+4} + (-1)^{t-1}F_{k-t+2} + (-1)^t F_{k-t}\\
&\vdots\\
&= F_2F_{k+t-2} + (-1)^3F_{k+t-6} + (-1)^4F_{k+t-8} + \cdots + (-1)^{t-1}F_{k-t+2} + (-1)^t F_{k-t}\\
&= F_2F_{k+t-2} + \big(-F_{k+t-6} + F_{k+t-8}\big) + \big(-F_{k+t-10} + F_{k+t-12}\big) + \cdots\\
&\leq F_2F_{k+t-2}.
\end{aligned}$$

□

From Theorem 2.4 and by an inductive process, the following result directly follows.

**Corollary 2.5.** *Let $d, m, n_1, n_2, \ldots, n_d$ be positive integers such that $n_1 + n_2 + \cdots + n_d = m$. Then we have $\prod_{j=1}^{d} F_{n_j} \leq (F_2)^{d-1} F_{m-2d+2}$.*

**Theorem 2.6.** *Let $T$ be a tree of order $n+1$, $x$ be a leaf in $T$ such that $ux \in E(T)$ and $d = \deg_{T-x}(u)$. Then we have $H(T) = H(T-x) + c(n,d)$ in which*

$$n - d \leq c(n,d) \leq (F_2)^{d-1} F_{n-2d+1}.$$

*Forthermore, these bounds are sharp.*

*Proof.* It is straightforward to see that $T-x$ is a tree of order $n$ and each matching (set of independent edges) in $T-x$ is a matching in $T$ that does not contain the edge $ux$, and vice versa. Also, each matching in $T$ which contains the edge $xu$ corresponds to a matching in the forest $T-u$. Let $T_1, T_2, \ldots, T_d$ be the connected components of $T-u$ other that the component consisting of the vertex $x$. Thus, for each $j \in \{1,2,\ldots,d\}$ the graph $T_j$ is a tree and let $n_j$ be its order. Hence, we have $n_1 + n_2 + \cdots + n_d = n-1$. By Lemma 2.2 and for each $j \in \{1,2,\ldots,d\}$ we have

$$n_j = H(S_{n_j}) \leq H(T_j) \leq H(P_{n_j}) = F_{n_j}.$$

DOI: http://dx.doi.org/10.22108/toc.2020.121874.1713

Now Lemma 2.1 implies that $H(T-u)=\prod_{j=1}^{d}H(T_j)$ and hence,

$$\begin{aligned} n_1n_2\cdots n_d &= \prod_{j=1}^{d}H(S_{n_j}) \\ &\leq H(T-u) \\ &\leq \prod_{j=1}^{d}H(P_{n_j}) \\ &= \prod_{j=1}^{d}F_{n_j} \end{aligned}$$

Since $n_1+n_2+\cdots+n_d=n-1$, from Theorem 2.5 and also by a simple application of calculus, we obtain

$$\begin{aligned} 1\times\cdots\times 1\times(n-1-(d-1)) &\leq n_1n_2\cdots n_d \\ &\leq H(T-u) \\ &\leq \prod_{j=1}^{d}F_{n_j} \\ &\leq (F_2)^{d-1}F_{n-2d+1}. \end{aligned}$$

To see that the lower bound is sharp, it is enough to consider the star graph $S_{n+1}$ in which its central vertex is $u$ and $ux\in(S_{n+1})$. Then, we see that

$$\begin{aligned} H(S_{n+1}) &= n+1 \\ &= n+(n-(n-1)) \\ &= H(S_{n+1}-x)+\big(n-\deg_{S_{n+1}-x}(u)\big). \end{aligned}$$

To see that the upper bound is sharp, consider the path graph $P_{n+1}$ in which $x$ is a leaf adjacent to the vertex $u$. In this case and we have $d=\deg_{P_{n+1}-x}(u)=1$ and

$$\begin{aligned} H(P_{n+1}) &= F_{n+1}=F_n+F_{n-1} \\ &= H(P_{n+1}-x)+2^{1-1}F_{n-2+1}. \end{aligned}$$

This completes the proof. □

We explain the following evolution process for random trees of order $n$, which turns out to be appropriate when studying the Hosoya index of trees. Every tree of order $n$ can be obtained uniquely by attaching $n$-th node to one of the $n-1$ nodes in a tree of order $n-1$. It is one of particular interests in applications to assume the random tree model and to speak about a random tree with $n$ nodes, which means that all trees of order $n$ are considered to appear equally likely. Equivalently, one may describe random trees via the following tree evolution process, which generates random trees of arbitrary order $n$. At step 1 the process starts with the root. At step $i$ the $i$-th node is attached

DOI: http://dx.doi.org/10.22108/toc.2020.121874.1713

to a previous node $v$ of the already grown tree $T$ of order $i-1$ with probability $p_i(v) = \frac{1}{i-1}$. For applicability of our own results and specially for some connections with the chemical relevance, see [12].

Let $H_n$ be the Hosoya index of a random tree of order $n$ and $\mathcal{F}_n$ be the sigma-field generated by the first $n$ stages of these trees [1, 13]. Let $U_n$ be a randomly chosen node belonging to a tree of order $n$.

**Theorem 2.7.** *For $n \geq 1$, the average value of the Hosoya index is given by*

$$\mathbf{E}(H_n) = \sum_{j=1}^{n-1} \frac{1}{j} \sum_{i=1}^{j} c(i, d_i).$$

*Proof.* From Theorem 2.6,

$$H_n = H_{n-1} + c(n-1, d_{U_{n-1}}). \tag{2.1}$$

Then

$$\begin{aligned} \mathbf{E}(H_n|\mathcal{F}_{n-1}) &= H_{n-1} + \mathbf{E}(c(n-1, d_{U_{n-1}})|\mathcal{F}_{n-1}) \\ &= H_{n-1} + \frac{1}{n-1}\sum_{i=1}^{n-1} c(i, d_i). \end{aligned} \tag{2.2}$$

The recurrence equation (2.2) leads to

$$\mathbf{E}(H_n) = \sum_{j=1}^{n-1} \frac{1}{j} \sum_{i=1}^{j} c(i, d_i),$$

and proof is completed. □

**Corollary 2.8.** *Since the sum of all degrees is equal to twice the number of edges:*

$$\frac{1}{4}(n-6)(n-1) + 2\mathcal{H}_{n-1} \leq \mathbf{E}(H_n) \leq \sum_{j=1}^{n-1} \frac{1}{j} \sum_{i=1}^{j} (F_2)^{d_i - 1} F_{n-2d_i+1},$$

*where $\mathcal{H}_n = \sum_{j=1}^{n} \frac{1}{j}$ is the $n$th harmonic number.*

**Theorem 2.9.** *For $n \geq 1$,*

$$\mathbf{V}ar(H_n) = n + \mathcal{O}(\log n).$$

*Proof.* Let $Z_1 = 0$ and for $n \geq 2$,

$$Z_n := H_n - H_{n-1} - \frac{1}{n-1}\sum_{i=1}^{n-1} c(i, d_i).$$

Then $\mathbf{E}(Z_n|\mathcal{F}_{n-1}) = 0$. From (2.1),

$$\mathbf{E}((H_n - H_{n-1})^2|\mathcal{F}_{n-1}) = \mathbf{E}(c(n-1, d_{U_{n-1}})^2) = \frac{1}{n-1}\sum_{i=1}^{n-1} c^2(i, d_i) \tag{2.3}$$

DOI: http://dx.doi.org/10.22108/toc.2020.121874.1713

and also

$$\begin{aligned}\mathbf{E}((H_n - H_{n-1})^2|\mathcal{F}_{n-1}) &= \mathbf{E}\left(\left(Z_n + \frac{1}{n-1}\sum_{i=1}^{n-1} c(i,d_i)\right)^2 |\mathcal{F}_{n-1}\right) \\ &= \mathbf{E}(Z_n^2|\mathcal{F}_{n-1}) + \left(\frac{1}{n-1}\sum_{i=1}^{n-1} c(i,d_i)\right)^2. \quad (2.4)\end{aligned}$$

Now, from (2.3) and (2.4),

$$\mathbf{E}(Z_n^2) = \frac{1}{n-1}\sum_{i=1}^{n-1} c^2(i,d_i) - \left(\frac{1}{n-1}\sum_{i=1}^{n-1} c(i,d_i)\right)^2, \ n \geq 2. \quad (2.5)$$

By definition of $Z_n$, we have

$$\begin{aligned}\mathbf{V}ar(H_n) &= \mathbf{E}(H_n - \mathbf{E}(H_n))^2 \\ &= \sum_{i=1}^{n} \mathbf{E}(Z_i^2) \\ &= n + \mathcal{O}(\log n),\end{aligned}$$

since for any $b \leq i \neq j \leq n$, $\mathbf{E}(Z_i Z_j) = 0$ [1]. □

**Corollary 2.10.** *By Theorem 2.9, the process* $\{H_n - \mathbf{E}(H_n), \mathcal{F}_n\}_{n\geq 1}$ *is a martingale* [1], *since*

$$\mathbf{E}(H_n - \mathbf{E}(H_n)|\mathcal{F}_{n-1}) = \mathbf{E}(H_n|\mathcal{F}_{n-1}) - \mathbf{E}(H_n).$$

We use the notations $\xrightarrow{D}$ and $\xrightarrow{P}$ to denote convergence in distribution and in probability, respectively. The standard random $N(\mu, \sigma^2)$ appear in the last theorem for the normal distributed with mean $\mu$ and variance $\sigma^2$, respectively. These random variable appear in the results as limiting random variable.

**Theorem 2.11.** *As* $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n} \frac{H_i - \mathbf{E}(H_i)}{i} \xrightarrow{P} 0.$$

*Proof.* By Chebyshev's inequality, for any $\varepsilon > 0$ [1],

$$\begin{aligned} P\Big(\Big|\frac{1}{n}\sum_{i=1}^{n}\frac{H_i-\mathbf{E}(H_i)}{i}\Big| > \varepsilon\Big) &\leq \frac{1}{n^2\varepsilon}\mathbf{E}\left(\sum_{i=1}^{n}\frac{H_i-\mathbf{E}(H_i)}{i}\right)^2 \\ &\leq \frac{1}{n^2\varepsilon}\mathbf{E}\Big(\sum_{i=1}^{n}\frac{1}{i}\sum_{j=1}^{i}Z_j\Big)^2 \\ &= \frac{1}{n^2\varepsilon}\mathbf{E}\Big(\sum_{i=1}^{n}(\mathcal{H}_n-\mathcal{H}_{i-1})Z_i\Big)^2 \\ &= \frac{1}{n^2\varepsilon}\sum_{i=1}^{n}(\mathcal{H}_n-\mathcal{H}_{i-1})^2\mathbf{E}(Z_i^2) \\ &\leq \frac{1}{n^2\varepsilon}(\mathcal{H}_n-\min_i\mathcal{H}_{i-1})^2\sum_{i=1}^{n}\mathbf{E}(Z_i^2) \\ &= \frac{1}{n\varepsilon}\log(n)+\mathcal{O}\left(\log^2(n)\right). \end{aligned}$$

□

**Theorem 2.12.** *As $n \to \infty$,*

$$Z_n^* = \frac{H_n - \mathbf{E}(H_n)}{\sqrt{n}} \xrightarrow{D} N(0,1).$$

*Proof.* Suppose

$$X_{n,i} = \frac{Z_i}{\sqrt{n}},\ \ i = 1, 2, \ldots, n,\ \ n \geq 1.$$

By [7, Corollary 3.1], it is sufficient to show that for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P\Big(\Big|\sum_{i=1}^{n}\mathbf{E}(X_{n,i}^2|\mathcal{F}_{i-1})\Big| > \varepsilon\Big) = 0 \tag{2.6}$$

and

$$\lim_{n\to\infty} P\Big(\Big|\sum_{i=1}^{n}\mathbf{E}(X_{n,i}^2 I(|X_{n,i}| > \varepsilon)|\mathcal{F}_{i-1}) > \varepsilon\Big) = 0. \tag{2.7}$$

First,

$$\begin{aligned} \sum_{i=1}^{n}\mathbf{E}(X_{n,i}^2|\mathcal{F}_{i-1}) &= \frac{1}{n}\sum_{j=1}^{n-1}\left(\frac{H_j-\mathbf{E}(H_j)}{j}\right) \\ &+ \frac{1}{n}\sum_{j=1}^{n-1}\left(\frac{\mathbf{E}(H_j)}{j} - \left(\frac{1}{j}\sum_{i=1}^{j}c(i,d_i)\right)^2\right). \end{aligned}$$

Thus (2.6) follows by Theorem 2.11. Let $d_v$ be the out-degree of node $v$ in a tree $T_n$ of order $n$. With the same method and this fact that [13]

$$\lim_{n\to\infty} P\left(\frac{1}{\sqrt{n}}\max_{v\in V(T_n)} d_v > \varepsilon\right) = 0,$$

we can prove (2.7) and proof is completed. □

DOI: http://dx.doi.org/10.22108/toc.2020.121874.1713

## 3. **Conclusion**

In this paper, we studied the Hosoya index of the trees. Some results on this index in terms of Fibonacci numbers in such structures are given. The lower and upper bound for average value and variance of this index are obtained. Also, the asymptotic normality of this index is given.

## References


[1] P. Billingsley, *Probability and Measure*, John Wiley and Sons, New York, 1885.

[2] H. Deng, The largest Hosoya index of $(n, n+1)$-graphs, $C$omput. Math. App., **56** (2008) 2499–2506.

[3] J. Devillers and A. T. Balaban, *Topological Indices and Related Descriptors in QSAR and QSPR*, Gordon and Breach, Amsterdam, 1999.

[4] M. V. Diudea, *QSPR/QSAR Studies by Molecular Descriptors,* Nova Science Publishers, Huntington, N. Y., 2001.

[5] Y. D. Gao and H. Hosoya, Topological index and thermodynamic properties. IV. Size dependency of the structure activity correlation of alkanes, *Bull. Chem. Soc. Jpn.,* **61** (1988) 3093–3102.

[6] I. Gutman and O. E. Polansky, *Mathematical Concepts in Organic Chemistry*, Springer, Berlin, 1986.

[7] P. Hall and C. C. Heyde, *Martingale Limit Theory and Its Application*, Academic Press, New York, 1980.

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

[9] H. Hosoya, Topological index as a sorting device for coding chemical structures, *J. Chem. Docum.*, **12** (1972) 181–183.

[10] H. Hosoya and K. Hosoi, Topological index as applied to $\pi$-electron systems. III, Mathematical relations among various bond orders. *J. Chem. Phys.*, **64** (1976) 1065–1073.

[11] H. Hosoya, Chemical meaning of octane number analyzed by topological indices, *Croat. Chem. Acta*, **75** (2002) 433–445.

[12] R. Kazemi, The eccentric connectivity index of bucket recursive trees, *Iranian J. Math. Chem.*, **5** (2014) 77–83.

[13] R. Kazemi, The second Zagreb index of molecular graphs with tree structure, *MATCH Commun. Math. Comput. Chem.* **72** (2014) 753–760.

[14] M. Randić, Comparative regression analysis: regressions based on a single descriptor, *Croat. Chem. Acta*, **66** (1993) 289–312.

[15] R. Todeschini and V. Consonni, *Handbook of Molecular Descriptors,* Wiley-VCH, Weinheim, 2000.

[16] S. G. Wagner, Extremal trees with respect to Hosoya index and Merrifield-Simmons index, *MATCH Commu. Math. Comput. Chem.*, **57** (2007) 221–233.



**Ramin Kazemi**
Department of Statistics, Imam Khomeini International University, Qazvin, Iran
Email: r.kazemi@sci.ikiu.ac.ir

**Ali Behtoei**
Department of Pure Mathematics, Imam Khomeini International University, Qazvin, Iran
Email: a.behtoei@sci.ikiu.ac.ir


DOI: http://dx.doi.org/10.22108/toc.2020.121874.1713