CROATICA CHEMICA ACTA CCACAA **73** (4) 1017–1026 (2000)

ISSN-0011-1643
CCA-2698 *Original Scientific Paper*

On the Hosoya Hyperindex and the Molecular Indices Based on a New Decomposition of the Hosoya *Z* Matrix

Vicente García Espeso,[a,*] *José Javier Molins Vara,*[a] *Beatriz Roy Lázaro,*[a] *Fèlix Riera Parcerisas,*[b] *and Dejan Plavšić* [c,*]

[a] *Institut Químic de Sarrià, Universitat Ramon Llull, E-08017 Barcelona, Spain*

[b] *Enginyeria La Salle, Universitat Ramon Llull, E-08022 Barcelona, Spain*

[c] *The Ruđer Bošković Institute, P. O. Box 180, HR-10002, Zagreb, Croatia*

Received July 19, 1999; revised July 28, 2000; accepted August 1, 2000

The decomposition of the Hosoya $\boldsymbol{Z}$ matrix into the sum of ${}_k\boldsymbol{Z}$ matrices, $k = 0, 1, 2, \ldots$, is proposed. The ${}_k\boldsymbol{Z}$ matrix is based on the independent sets of k edges of the spanning subgraphs generated in the construction of the $\boldsymbol{Z}$ matrix. The Hosoya hyperindex H and a set of structurally related molecular indices ${}_kZ$ defined as the sum of all off-diagonal entries in the upper triangle of the $\boldsymbol{Z}$ matrix and the corresponding ${}_k\boldsymbol{Z}$ matrices, respectively, are put forward and studied analytically.

Key words: Hosoya hyperindex, decomposition of Hosoya $\boldsymbol{Z}$ matrix, new molecular indices.

INTRODUCTION

The representation of a molecule (or molecular graph) by a single number (molecular index) brings about a certain loss of information concerning the molecular structure. Hence, to diminish the loss of information an intensive search for novel molecular indices that would improve the graph-theoretical characterization of molecular structure has been undertaken in several directions.[1] One direction of the search is the creation of novel graph

* Author to whom correspondence should be addressed.

matrices whose selected invariants could be used as molecular indices. In the framework of these efforts Randić[2] has recently put forward a novel graph matrix, the Hosoya $\boldsymbol{Z}$ matrix, the construction of which is related to the construction of the Hosoya Z index.[3] He derived two structurally related molecular indices, the path numbers ^1Z and ^2Z, from the $\boldsymbol{Z}$ matrix and reported a few empirical regularities for the $\boldsymbol{Z}$ matrix entries in case of n-alkanes.[2] Plavšić *et al.* introduced the $\boldsymbol{Z}_{ew}$ matrix being the generalization of the $\boldsymbol{Z}$ matrix.[4] They also put forward the generalized path numbers and studied them and their relationships to the Hosoya Z index analytically.[4–6] The testing of the path numbers on diverse sets of compounds revealed their usability as predictor variables in QSPR (quantitative structure-property relationship) modeling, in particular in the studies of isomeric series.[2,4,5]

In this article we will examine a possibility to construct new molecular indices from the $\boldsymbol{Z}$ matrix, in particular a set of novel structurally related molecular indices.

DEFINITIONS

Z *Index*

Hosoya noticed the chemical significance of independent sets of edges of a molecular graph[3] G (a set of edges of G is independent if no two edges of the set are adjacent in G).[7] He introduced the quantities $a(G, k)$ as the number of independent sets of k edges of G and put forward the graph-theoretical index Z.

The Z index, $Z = Z(G)$, of a connected undirected graph G is defined as the total number of independent sets of edges of G,[3]

$$Z = \sum_{k=0}^{[N/2]} a(G, k) , \tag{1}$$

where N is the number of vertices of G and the Gaussian brackets, [ ], represent the integer part of $N/2$. Since the empty set and all singleton (1-element) sets are independent, $a(G, 0) = 1$, and $a(G, 1)$ is equal to the number of edges of G. If G is disconnected and consists of components G_i ($i = 1, 2, \ldots, r$),

$$G = \bigcup_{i=1}^{r} G_i , \tag{2}$$

the contributions of G_i to $Z(G)$ are multiplied,[3,6]

$$Z = Z(G) = \prod_{i=1}^{r} Z(G_i) . \tag{3}$$

Hosoya **Z** *Matrix*

The Hosoya $\boldsymbol{Z}$ matrix, $\boldsymbol{Z} = \boldsymbol{Z}(\mathrm{T})$, of a connected undirected acyclic graph (tree) T with N vertices is the symmetric matrix of order N, whose entry in the i-th row and j-th column, z_{ij}, is defined by the expression,[2]

$$z_{ij} = \begin{cases} Z(\mathrm{T} - p_{ij}) \text{ for } i \neq j \\ 0 \text{ for } i = j, \end{cases} \tag{4}$$

where p_{ij} stands for the path connecting vertices i and j in T and $Z(\mathrm{T} - p_{ij})$ is the Z index of the spanning subgraph $\mathrm{T} - p_{ij}$ of T. The spanning subgraph $\mathrm{T} - p_{ij}$ is obtained from T by erasing from it all edges of the path p_{ij}, but keeping all its vertices. Hence, T and $\mathrm{T} - p_{ij}$ have an equal number of vertices. One should notice that in case of the cycle-containing and edge-weighted graphs the expression for z_{ij}, Eq. (4), has to be modified.[4] The finding of the elements of the $\boldsymbol{Z}$ matrix is illustrated for 2-methylbutane in Figure 1.

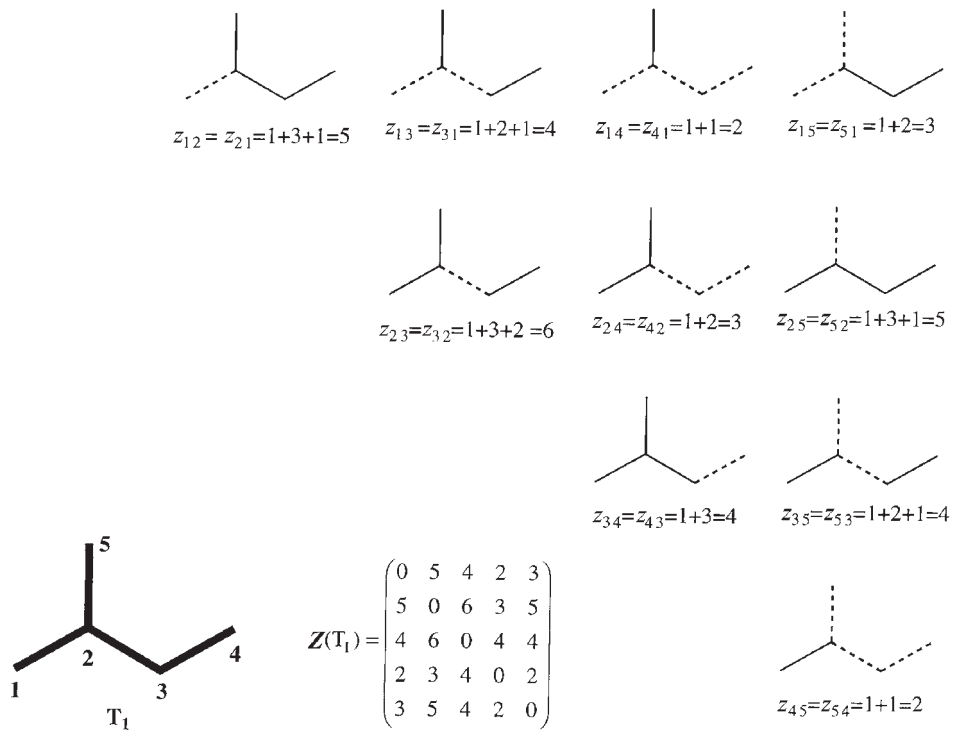

Figure 1. Calculation of the elements of the $\boldsymbol{Z}$ matrix of the labeled hydrogen-suppressed graph T_1 representing 2-methylbutane. The broken lines represent erased edges.

DECOMPOSITIONS OF HOSOYA $\boldsymbol{Z}$ MATRIX

First Decomposition

If one takes the lengths of the erased paths into account, then the Hosoya $\boldsymbol{Z}$ matrix of a tree T can be decomposed into the sum of the sparse Hosoya matrices,[2]

$$\boldsymbol{Z} = \sum_{n=1}^{d(\mathrm{T})} {}^{n}\boldsymbol{Z} \,, \tag{5}$$

where the summation goes over all path lengths in T, *i.e.*, from the length 1 (of an edge) up to the diameter[7] of T, d(T). The sparse Hosoya matrix ${}^{n}\boldsymbol{Z}$ is the symmetric matrix whose entry ${}^{n}z_{ij}$ is given by the expression,

$${}^{n}z_{ij} = \begin{cases} Z(\mathrm{T} - p_{ij}) \text{ if the length of } p_{ij} \text{ is } n \\ 0 \text{ otherwise}. \end{cases} \tag{6}$$

This decomposition of the $\boldsymbol{Z}$ matrix is illustrated for the $\boldsymbol{Z}(\mathrm{T}_1)$ matrix shown in Figure 1,

$$\begin{pmatrix} 0 & 5 & 4 & 2 & 3 \\ 5 & 0 & 6 & 3 & 5 \\ 4 & 6 & 0 & 4 & 4 \\ 2 & 3 & 4 & 0 & 2 \\ 3 & 5 & 4 & 2 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 5 & 0 & 0 & 0 \\ 5 & 0 & 6 & 0 & 5 \\ 0 & 6 & 0 & 4 & 0 \\ 0 & 0 & 4 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 4 & 0 & 3 \\ 0 & 0 & 0 & 3 & 0 \\ 4 & 0 & 0 & 0 & 4 \\ 0 & 3 & 0 & 0 & 0 \\ 3 & 0 & 4 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 2 & 0 \end{pmatrix} .$$

Second Decomposition

If one replaces every non-zero number appearing in the $\boldsymbol{Z}(\mathrm{T}_1)$ matrix by the corresponding addition shown in Figure 1,

$$\begin{pmatrix} 0 & 5 & 4 & 2 & 3 \\ 5 & 0 & 6 & 3 & 5 \\ 4 & 6 & 0 & 4 & 4 \\ 2 & 3 & 4 & 0 & 2 \\ 3 & 5 & 4 & 2 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1+3+1 & 1+2+1 & 1+1 & 1+2 \\ 1+3+1 & 0 & 1+3+2 & 1+2 & 1+3+1 \\ 1+2+1 & 1+3+2 & 0 & 1+3 & 1+2+1 \\ 1+1 & 1+2 & 1+3 & 0 & 1+1 \\ 1+2 & 1+3+1 & 1+2+1 & 1+1 & 0 \end{pmatrix} ,$$

then the $\boldsymbol{Z}(\mathrm{T}_1)$ matrix can be decomposed as follows:

$$\begin{pmatrix} 0 & 5 & 4 & 2 & 3 \\ 5 & 0 & 6 & 3 & 5 \\ 4 & 6 & 0 & 4 & 4 \\ 2 & 3 & 4 & 0 & 2 \\ 3 & 5 & 4 & 2 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 3 & 2 & 1 & 2 \\ 3 & 0 & 3 & 2 & 3 \\ 2 & 3 & 0 & 3 & 2 \\ 1 & 2 & 3 & 0 & 1 \\ 2 & 3 & 2 & 1 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 2 & 0 & 1 \\ 1 & 2 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{pmatrix} .$$

In general, the Hosoya $\boldsymbol{Z}$ matrix of a tree T with N vertices in respect of the cardinal numbers, k, of independent sets of edges of the spanning subgraphs $T - p$ can be decomposed into the sum of matrices,

$$\boldsymbol{Z} = \sum_{k=0}^{k=m} {}_k\boldsymbol{Z} , \tag{7}$$

where ${}_k\boldsymbol{Z}$ is the symmetric matrix of order N, whose entry ${}_kz_{ij}$ is defined as

$${}_kz_{ij} = \begin{cases} a(T-p_{ij}, k) \text{ if } i \neq j \\ 0 \text{ if } i = j . \end{cases} \tag{8}$$

The summation in Eq. (7) runs from the cardinal number 0 (of the empty set) up to the cardinal number m of the largest independent set of edges possible in $T - p_{rs}$ (r = 1, 2, …, $N - 1$ & $r < s \leq N$).

NEW MOLECULAR INDICES DERIVED FROM HOSOYA $\boldsymbol{Z}$ MATRIX

Molecular Indices Based on Second Decomposition of Hosoya **Z** *Matrix*

The first decomposition of the $\boldsymbol{Z}$ matrix inspired Randić[2] to introduce the path numbers. The path number ${}^nZ$ of a tree T, ${}^nZ = {}^nZ(T)$, is defined by the expression,

$${}^nZ = \sum_{i<j} {}^nz_{ij} , \tag{9}$$

where the summation goes over all off-diagonal entries ${}^nz_{ij}$ in the upper triangle of the ${}^n\boldsymbol{Z}$ matrix of T. The path numbers are important because they give one an opportunity to form a basis of structurally related molecular indices. Such a basis may offer a simpler interpretation of the regression equations, in comparison with a basis of *ad hoc* indices, in topology-property/activity studies.[8]

The question arises whether it is possible to create another set of structurally related molecular indices from the $\boldsymbol{Z}$ matrix. The answer is affirma-

tive since the second decomposition of the $\boldsymbol{Z}$ matrix offers us directly a possibility to construct such a set. We propose a series of new molecular indices ${}_kZ$, $k = 0, 1, \ldots, m$, based on the ${}_k\boldsymbol{Z}$ matrices. The ${}_kZ$ index of a tree T, ${}_kZ = {}_kZ(\mathrm{T})$, is defined by the formula,

$$ {}_kZ = \sum_{i<j} {}_kz_{ij} \ , \tag{10} $$

where the summation goes over all off-diagonal entries ${}_kz_{ij}$ in the upper triangle of the ${}_k\boldsymbol{Z}$ matrix of T.

We have derived analytical expressions for the first few ${}_kZ'$s of a tree enabling us to calculate the ${}_kZ'$s directly from structural characteristics (like number of vertices and paths, valencies of vertices, *etc.*) of the tree. The derivations of these explicit formulae shall be stated.

From the definition of the ${}_0\boldsymbol{Z}$ matrix, Eq. (8), follows that $\forall\ {}_0z_{ij}$ $(i \neq j)$ equals 1 and hence the ${}_0Z$ index of a tree with N vertices can be expressed by means of N only,

$$ {}_0Z = \binom{N}{2} . \tag{11} $$

The molecular index ${}_1Z$, Eq. (10) (k=1), of a tree T with N vertices equals the total number of edges of the spanning subgraphs $\mathrm{T} - p_{rs}$ (r = 1, 2, …, $N - 1$ & $r < s \leq N$) of T, and consequently ${}_1Z$ can be expressed by means of N and the number of paths of length n, ${}^np$, in T,

$$ {}_1Z = \sum_{n=1}^{n \leq N-2} (N - 1 - n)\, {}^np . \tag{12} $$

The number of paths of a given length n in T can be calculated by means of the analytical expression,[9]

$$ {}^np = \sum_{r=1}^{n-2+g} (-1)^{r+h} (n - r - 1) \sum_{\substack{d(i,j)=r \\ i<j}} v_i v_j + (-1)^{n+1} \left(\frac{n-1}{2} \sum_i v_i^2 + N - 1 \right), \tag{13} $$

where v_i is the valency of the vertex i and $d(i, j)$ stands for the distance between vertices i and j in T. The parameters g and h take the following values

$$ g = \begin{cases} 1 \text{ for } n = 1 \\ 0 \text{ for } n \geq 2 \end{cases} \tag{14} $$

and

$$h = \begin{cases} 1 & \text{for } n = \text{odd} \\ 0 & \text{for } n = \text{even}. \end{cases} \tag{15}$$

It is worthwhile to mention that ${}^{n}p$ ($n \geq 3$) can also be calculated recursively. [9]

The ${}_2Z$ index, Eq. (10) (k = 2), of a tree T with N vertices can be expressed in the form

$${}_2Z = \sum_{p \in P(\mathrm{T})} a\,(\mathrm{T} - p, 2), \tag{16}$$

where a(T – p, 2) is the number of independent sets of 2 edges of the spanning subgraph T – p of T, P(T) denotes the set of all paths in T whose lengths are ≥1, and the summation runs over all elements of P(T). It is easy to see that a(T – p, 2) can be expressed as

$$a(\mathrm{T} - p, 2) = a(\mathrm{T}, 2) - \sum_{e \in p} \Delta_e + \binom{n-1}{2}, \tag{17}$$

where Δ_e is the number of independent sets of 2 edges of T that contain the edge e, and n is the length of the path p. The summation in the second term on the right-hand side of Eq. (17) goes over all edges making up the path p. Combining Eqs. (16) and (17) and by noting that the following expressions hold[5,10]

$$a(\mathrm{T}, 2) = \binom{N}{2} - \frac{1}{2}\sum_i v_i^2 \tag{18}$$

and

$$\sum_{p \in P(\mathrm{T})} \sum_{e \in p} \Delta_e = \sum_{e_{ij}} n_i(e_{ij})\, n_j(e_{ij})\, \Delta_{e_{ij}}, \tag{19}$$

where $n_i(e_{ij})$ and $n_j(e_{ij}) = N - n_i(e_{ij})$ denote the numbers of vertices of the two components of T – e_{ij}, one obtains

$${}_2Z = \binom{N}{2}\left[\binom{N}{2} - \frac{1}{2}\sum_i v_i^2\right] - \sum_{e_{ij}} n_i(e_{ij})\, n_j(e_{ij})\, \Delta_{e_{ij}} + \sum_{n=3}^{n \leq N-1} \binom{n-1}{2}\, {}^{n}p\ . \tag{20}$$

Since the following relations hold:

$$\Delta_{e_{ij}} = N - (v_i + v_j)\,, \tag{21}$$

$$\sum_i v_i^2 = \sum_{e_{ij}} (v_i + v_j) \,, \tag{22}$$

where v_i and v_j are the valences of vertices i and j in T joined by the edge e_{ij}, and

$$\sum_{e_{ij}} n_i(e_{ij}) n_j(e_{ij}) = \sum_{n=1}^{n \le N-1} n^n p \,, \tag{23}$$

the sought after explicit formula for calculation of ${}_2Z$ of T reads as follows

$${}_2Z = \binom{N}{2}^2 + \sum_{n=1}^{n \le N-1} \left[\binom{n-1}{2} - Nn \right]^n p + \sum_{e_{ij}} \left[n_i(e_{ij}) n_j(e_{ij}) - \frac{1}{2}\binom{N}{2} \right] (v_i + v_j) \,. \tag{24}$$

Paths

A special case of a tree is a path graph (tree with minimum number of terminals, 2). The explicit expressions for the calculation of ${}_1Z$ and ${}_2Z$, Eqs. (12) and (24), respectively, in case of the path graph with N vertices, P_N, take rather simple forms:

$${}_1Z(P_N) = \frac{N(N-1)(N-2)}{3} \tag{25}$$

and

$${}_2Z(P_N) = \frac{(N-1)(N-2)(N-3)(3N-8)}{24} \,, \tag{26}$$

respectively.

Hosoya Hyperindex H

From the definitions of the $\boldsymbol{Z}$ matrix, Eq. (4), the path numbers ${}^nZ$, Eq. (9), and the molecular indices ${}_kZ$, Eq. (10), follows

$$\sum_{i<j} z_{ij} = \sum_{n=1}^{d(T)} \sum_{i<j} {}^n z_{ij} = \sum_{k=0}^{m} \sum_{i<j} {}_k z_{ij} \,. \tag{27}$$

The quantity represented by Eq. (27) is also an invariant of a tree graph. We name it the Hosoya hyperindex H.

It is possible to derive explicit analytical expressions for the calculation of the Hosoya hyperindex but only for some special classes of compounds

(graphs). We have derived an explicit formula for the calculation of the Hosoya hyperindex of P_N , $H(P_N)$. Its derivation shall be stated.

If one labels the vertices of P_N in the increasing order starting with 1 at a terminal vertex, then the entry z_{ij} ($i < j$) of $\boldsymbol{Z}(P_N)$ (see Eqs. (2)–(4) and recall that $Z(K_1 = P_1) = 1$) is given by the relation

$$z_{ij} = Z(P_i)Z(P_{N+1-j}) . \tag{28}$$

Combining Eqs. (27) and (28) one obtains

$$H(P_N) = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} Z(P_i)Z(P_{N+1-j}) = \sum_{r=1}^{N-1} \sum_{s=1}^{N-r} Z(P_r)Z(P_s) , \tag{29}$$

where the first double summation runs over rows and the second double summation goes over columns of the off-diagonal entries in the upper triangle of the $\boldsymbol{Z}(P_N)$ matrix. By noting that

$$\sum_{s=1}^{N-r} Z(P_s) = Z(P_{N+2-r}) - 2 , \tag{30}$$

Eq. (29) can be written in the form

$$H(P_N) = \sum_{r=1}^{N-1} Z(P_r)(Z(P_{N+2-r}) - 2) . \tag{31}$$

Since the N-th Fibonacci number, F_N, equals $Z(P_N)$, (the initial conditions for the Fibonacci recursion, $F_N = F_{N-1} + F_{N-2}$, are $F_0 = F_1 = 1$), $Z(P_r)$ and $Z(P_{N+2-r})$ can be expressed by the Binét formula,[11] and the sought after explicit analytical expression for H of P_N reads as follows

$$H(P_N) = \sum_{r=1}^{N-1} \left\{ \left[\left(\frac{1+\sqrt{5}}{2} \right)^{r+1} - \left(\frac{1-\sqrt{5}}{2} \right)^{r+1} \right] \left[\left(\frac{1+\sqrt{5}}{2} \right)^{N+3-r} - \left(\frac{1-\sqrt{5}}{2} \right)^{N+3-r} - 2 \right] \right\} . \tag{32}$$

This formula enables one to calculate the Hosoya hyperindex directly from P_N, namely from the knowledge of the number of vertices in P_N only.

Acknowledgment. – This research was supported in part by the Ministry of Science and Technology of the Republic of Croatia.

REFERENCES

1. M. Randić and N. Trinajstić, *J. Mol. Struct.* **300** (1993) 551–571.
2. M. Randić, *Croat. Chem. Acta* **67** (1994) 415–429.
3. H. Hosoya, *Bull. Chem. Soc. Japn.* **44** (1971) 2332–2339.

4. D. Plavšić, M. Šoškić, Z. Đaković, I. Gutman, and A. Graovac, *J. Chem. Inf. Comput. Sci.* **37** (1997) 529–534.
5. D. Plavšić, M. Šoškić, I. Landeka, I. Gutman, and A. Graovac, *J. Chem. Inf. Comput. Sci.* **36** (1996) 1118–1122.
6. I. Gutman, D. Plavšić, M. Šoškić, I. Landeka, and A. Graovac, *Croat. Chem. Acta* **70** (1997) 941–954.
7. G. Chartrand and O. R. Oellermann, *Applied and Algorithmic Graph Theory*, McGraw-Hill, New York, 1993.
8. M. Randić, *J. Chem. Inf. Comput. Sci.* **32** (1992) 686–692.
9. D. Plavšić, M. Šoškić, I. Landeka, and N. Trinajstić, *J. Chem. Inf. Comput. Sci.* **36** (1996) 1123–1126.
10. D. Plavšić, *Chem. Phys. Lett.* **304** (1999) 111–116.
11. I. Gutman and O. Polansky, *Mathematical Concepts in Organic Chemistry*, Springer-Verlag, Berlin, 1986, p. 131.

SAŽETAK

O Hosoyinu hiperindeksu i molekulskim indeksima temeljenim na novom rastavu Hosoyine matrice *Z*

Vicente García Espeso, José Javier Molins Vara, Beatriz Roy Lázaro, Fèlix Riera Parcerisas i Dejan Plavšić

Predložen je rastav Hosoyine matrice $\boldsymbol{Z}$ u sumu matrica ${}_k\boldsymbol{Z}$ (k = 0, 1, 2, …). Matrica ${}_k\boldsymbol{Z}$ temelji se na skupovima k nezavisnih bridova u razapinjućim podgrafovima koji se razmatraju pri konstrukciji matrice $\boldsymbol{Z}$. Predloženi su i analitički proučeni Hosoyin hiperindeks H i niz strukturno povezanih molekulskih indeksa ${}_k\boldsymbol{Z}$ definiranih kao suma elemenata iznad glavne dijagonale matrice $\boldsymbol{Z}$ odnosno odgovarajućih matrica ${}_k\boldsymbol{Z}$.