J Math Chem (2010) 48:395–400
DOI 10.1007/s10910-010-9681-7

ORIGINAL PAPER

One-two descriptor

Damir Vukičević · Marin Bralo · Ana Klarić · Antonija Markovina · Dina Spahija · Ana Tadić · Ana Žilić

Received: 28 January 2010 / Accepted: 12 April 2010 / Published online: 8 May 2010

Abstract In this paper, we introduce the one-two descriptor. The one-two descriptors is the sum of the vertex contributions such that each pendant vertex contributes 1, each vertex of degree two adjacent to pendant vertex contributes 2, and each vertex of degree higher than two also contributes 2. We test this descriptor on the benchmark data set of the octane isomers proposed by the International Academy of Mathematical Chemistry. We show that this descriptor is a good predictor of the heat capacity at P constant (CP) and of the total surface area (TSA). Linear model predictions for both these properties are better then predictions of any of the benchmark descriptors proposed by International Academy of Mathematical Chemistry. Linear model predictions of TSA are also better then predictions of any of Adriatic descriptors, while linear model predictions of CP are not as good as predictions by Adriatic descriptor

D. Vukičević (✉) · M. Bralo · A. Klarić · A. Markovina · D. Spahija · A. Tadić · A.Žilić
Faculty of Natural Sciences and Mathematics, University of Split, Nikole Tesle 12, 21000 Split, Croatia
e-mail: vukicevi@pmfst.hr
URL: www.pmfst.hr/~vukicevi

M. Bralo
e-mail: mbralo@gmail.com

A. Klarić
e-mail: anaklaric86@hotmail.com

A. Markovina
e-mail: antonija.markovina@gmail.com

D. Spahija
e-mail: dinaspahija@hotmail.com

A. Tadić
e-mail: anatadic11@hotmail.com

A. Žilić
e-mail: azilic@net.hr

called inverse sum lordeg index. We also analyze the mathematical properties of this descriptor and we find tight upper and lower bounds in the families of the trees with n vertices and the chemical trees with n vertices.

Keywords One-two descriptor · Molecular descriptor · Linear models · Extremal graphs

1 Introduction

The molecular descriptor is the final result of a logical and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment [1]. Molecular descriptors have been shown to be useful in modeling many physico-chemical properties in numerous QSAR and QSPR studies [2–4].

In this paper, we introduce one-two descriptor. It is defined as the sum of the vertex contributions is such a way that each pendant vertex contributes 1, each vertex of degree two adjacent to pendant vertex contributes 2, and each vertex of degree higher than two also contributes 2. We illustrate this definition for 3-ethyl-hexane by Fig. 1

One-two descriptor of graph G will be denoted by $OT(G)$. For instance, if G is 3-ethyl-hexane, then $OT(G) = 11$. We show that one-two descriptor is a good predictor of the heat capacity at P constant (CP) and of the total surface area (TSA) for octane isomers. Further, we analyze mathematical properties of this descriptor. Namely, we find tight upper and lower bounds in the families of the trees with n vertices and the chemical trees with n vertices.

2 QSAR results

International Academy of Mathematical Chemistry [5] proposed four benchmark sets [6] as sets for testing the molecular descriptors. Also, recently Adriatic descriptors [7–11] have been proposed and in many cases they have provided better results then benchmark descriptors [7,10]. In this paper, we show that one-two descriptor is a good predictor of TSA and CP. Linear model predictions of the best benchmark descriptor, the best Adriatic descriptor and the one-two descriptor are presented in Figs. 2 and 3.

It can be seen that linear model prediction of TSA of one-two descriptor are better than linear model predictions of the best benchmark descriptor and the best Adriatic descriptor. Linear model predictions of CP of one-two descriptor are slightly better

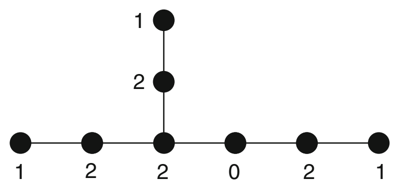

Fig. 1 Vertex contributions of 3-ethyl-hexane

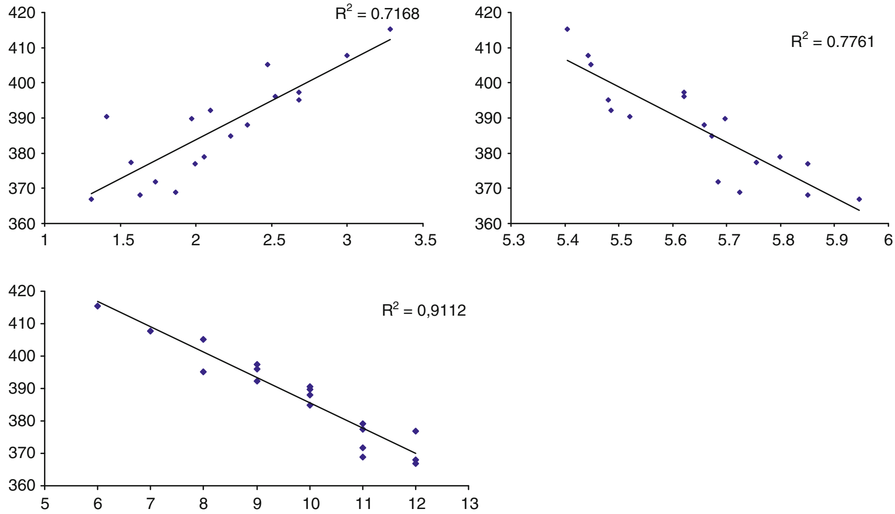

Fig. 2 Linear model predictions of the total surface area by the best benchmark descriptor (the second Mohar indeks TI2), by the best Adriatic descriptor (inverse sum lordeg indeks), and by the one-two descriptor

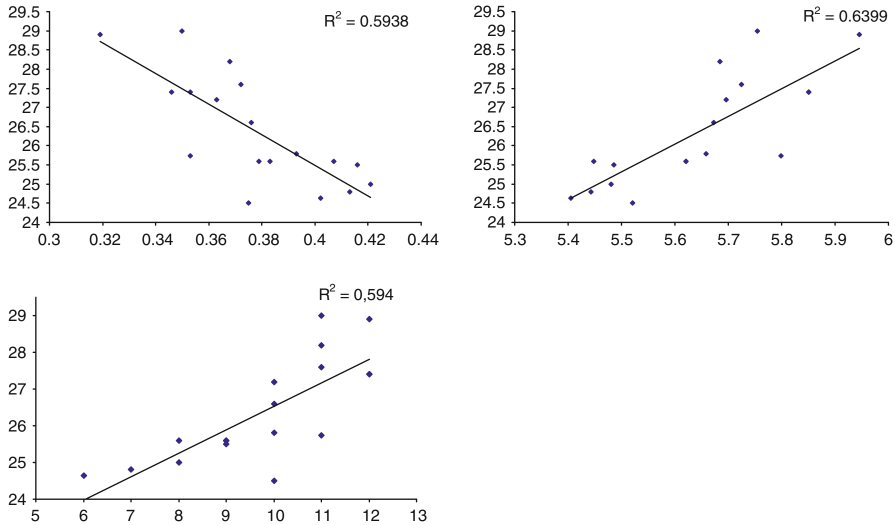

Fig. 3 Linear model predictions of the heat capacity at P constant by the best benchmark descriptor (average connectivity indeks χ_2), by the best Adriatic descriptor (inverse sum lordeg indeks), and by the one-two descriptor

then the best benchmark descriptor (we can say that they are of comparable quality), but they are not as good as linear model predictions by Adriatic descriptor called inverse sum lordeg index.

3 Mathematical properties

Before proving Theorem 1, let us introduce some notation. By n_i we denote the number of vertices of degree i and by $d(u)$ we denote the degree of vertex u. Let x be any real number. By $\lfloor x \rfloor$ we denote the greatest integer not greater than x. In the proof of the Theorem, we shall use the following well known Lemma:

Lemma 1 *Let G be a tree with at least 2 vertices. Then it holds:*

$$n_1(G) = \sum_{i \geq 3} (i-2)\, n_i(G) + 2.$$

Now, we can prove:

Theorem 1 *Let G be a tree with n vertices. It holds*

$$\left.\begin{array}{l} 0,\ n=1 \\ 1,\ n=2 \\ 4,\ n=3 \\ 5,\ n=4 \\ 6,\ n\geq 5 \end{array}\right\} \leq OT(G) \leq \begin{cases} 0, & n=1 \\ \left\lfloor \frac{5n-2}{3} \right\rfloor, & n \geq 2 \end{cases}$$

These bounds are tight and the bounds can be obtained for chemical graphs.

Proof Let us prove the lower bound for $OT(G)$. It can be easily checked for $n \leq 4$. Hence, let us assume that $n \geq 5$. It is well known that each tree with at least two vertices has at least two leaves. If G is a star, then $OT(G) = 2 + (n-1)\cdot 1 = n+1 \geq 6$. If G is not a star, than there are at least two vertices adjacent to leaves hence $OT(G) \geq 2+2+1+1 = 6$. Examples of the extremal graphs obtaining the lower bounds are paths for $n \geq 5$.

Now, let us prove the upper bound. If $n \leq 4$, the claim is obvious. Hence, suppose that $n \geq 5$. Let G be a graph with maximal upper bound. First let us prove that G does not contain any vertices of contribution 0 to OT index. Supposed to the contrary that there is a vertex u adjacent to vertices v_1 and v_2 such that $d(v_1), d(v_2) \geq 2$. Let $G' = G - uv_2 + v_1v_2$. It can be easily seen that contributions of all the vertices except u and v_2 to OT index remained the same, the contribution of v_2 did not decrease and the contribution of u increased from 0 to 1. Hence, $OT(G') > OT(G)$, which is contradiction. Hence, indeed there are no vertices of contribution 0. Therefore,

$n_2(G) \le n_1(G)$. Let us denote $x = n_1(G) - n_2(G)$, $x \ge 0$. Hence,

$$\begin{aligned}
\frac{OT(G)}{n(G)-4} &= \frac{n_1(G) + 2\cdot\sum_{i\ge 2} n_i(G)}{n(G)-4} = \frac{3n_1(G) - 2x + 2\cdot\sum_{i\ge 3} n_i(G)}{n(G)-4} \\
&= \frac{3\left(\sum_{i\ge 3}(i-2)\cdot n_i(G) + 2\right) + 2\cdot\sum_{i\ge 3} n_i(G) - 2x}{n(G)-4} \\
&= \frac{\sum_{i\ge 3}(3i-4)\cdot n_i(G) - 2x}{n(G)-4} + \frac{6}{n(G)-4} \\
&= \frac{\sum_{i\ge 3}(3i-4)\cdot n_i(G) - 2x}{n_1(G) + n_2(G) + \sum_{i\ge 3} n_i(G) - 4} + \frac{6}{n(G)-4} \\
&= \frac{\sum_{i\ge 3}(3i-4)\cdot n_i(G) - 2x}{2n_1(G) - x + \sum_{i\ge 3} n_i(G) - 4} + \frac{6}{n(G)-4} \\
&= \frac{\sum_{i\ge 3}(3i-4)\cdot n_i(G) - 2x}{2\left(\sum_{i\ge 3}(i-2)\cdot n_i(G) + 2\right) - x + \sum_{i\ge 3} n_i(G) - 4} + \frac{6}{n(G)-4} \\
&= \frac{\sum_{i\ge 3}(3i-4)\cdot n_i(G) - 2x}{\sum_{i\ge 3}(2i-3)\, n_i(G) - x} + \frac{6}{n(G)-4} \\
&\le \frac{\max\limits_{i\ge 3}\left\{\frac{3i-4}{2i-3}\right\}\cdot\sum_{i\ge 3}(2i-3)\cdot n_i(G) - 2x}{\sum_{i\ge 3}(2i-3)\, n_i(G) - x} + \frac{6}{n(G)-4} \\
&= \frac{\frac{5}{3}\cdot\sum_{i\ge 3}(2i-3)\cdot n_i(G) - 2x}{\sum_{i\ge 3}(2i-3)\, n_i(G) - x} + \frac{6}{n(G)-4} \\
&= \frac{5}{3}\,\frac{\sum_{i\ge 3}(2i-3)\cdot n_i(G) - x}{\sum_{i\ge 3}(2i-3)\, n_i(G) - x} - \frac{\frac{1}{3}x}{\sum_{i\ge 3}(2i-3)\, n_i(G) - x} \\
&\quad + \frac{6}{n(G)-4} \le \frac{5}{3} + \frac{6}{n(G)-4}.
\end{aligned}$$

Hence,

$$OT(G) \le \frac{5}{3}(n(G)-4) + 6 = \frac{5n(G)-2}{3}.$$

Since $OT(G)$ is integer, it follows that $OT(G) \le \left\lfloor \frac{5n(G)-2}{3} \right\rfloor$.

The examples of the extremal graphs obtaining the lower bounds are presented in Fig. 4.

This proves the Theorem. □

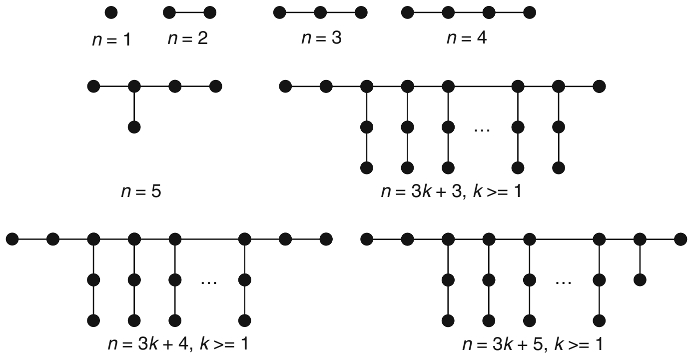

Fig. 4 Extremal graphs obtaining the upper bounds

Acknowledgments The partial support of Croatian Ministry of Science, Education and Sport (grants no. 177-0000000-0884 and 037-0000000-2779) is gratefully acknowledged.

References

1. R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors* (Wiley-VCH, Weinheim, 2000)
2. N. Trinajstić, *Chemical Graph Theory* (CRC Press, Boca Raton, 1992)
3. J. Devillers, A.T. Balaban (eds.), *Topological Indices and Related Descriptors in QSAR and QSPR* (Gordon and Breach, Amsterdam, 1999)
4. M. Karelson, *Molecular Descriptors in QSAR/QSPR* (Wiley-Interscience, New York, 2000)
5. http://www.iamc-online.org/
6. http://www.moleculardescriptors.eu/dataset/dataset.htm
7. D. Vukičević, M. Gašperov, Bond additive modeling 1. Adriatic Indices, Croat. Chem. Acta, submitted
8. D. Vukičević, Bond additive modeling 2. Mathematical properties of max–min rodeg index, Croat. Chem. Acta, submitted
9. D. Vukičević, Bond additive modeling 3. Comparison between the product-connectivity indeks and sum-connectivity index, Croat. Chem. Acta, submitted
10. D. Vukičević, Bond aditive modeling 4. QSPR and QSAR studies of variable Adriatic indices, Croat. Chem. Acta, submitted
11. D. Vukičević, Bond aditive modeling 5. Mathematical properties of variabel sum exdeg indeks, Croat. Chem. Acta, submitted