

Short communication

Topological Indices are Not Necessarily Invariant to Graph Labeling

Anton Perdih

Faculty of Chemistry and Chemical Technology, University of Ljubljana (retired)
Večna pot 113, 1000 Ljubljana, Slovenia

* Corresponding author: E-mail: a.perdih@gmail.com

Received: 28-10-2014

Dedicated to the memory of Prof. Dr. Jurij V. Brenčič.

Abstract

Each element of the Universal matrix U (vertex-degree vertex-distance weighted matrix) represents the mutual contribution of two vertices weighted for the vertex degrees and the distance between them. Regarding different labeling ways of graph vertices, particular matrix elements are not invariant to molecular labeling. Regarding the structural features they are invariants since there are only particular vertex combinations representing particular structural features. Some of the matrix elements are the best topological indices of a physicochemical property in question. Some combinations of matrix elements are very good topological indices of physicochemical properties of octanes regardless how we enumerate their vertices.

Keywords: Boiling point, Octanes, Universal matrix, Graph-theoretical modeling.

1. Introduction

There is a general opinion that the molecular descriptors (topological indices) must be graph invariants.^{1–7} However, on inclusion of vertex-degree vertex-distance weighted indices,^{8–14} especially with variation of value of exponents on vertex degrees and vertex distances,^{11–14} there appear situations, where some combinations of exponents give rise to invariance to molecular labeling, whereas other combinations do not. Some of the latter ones give rise to better correlations with physicochemical properties of alkanes in general or of octanes in particular.^{12–14} There arises the question, whether the use of topological indices of the latter type is legitimate, as well as in which cases they appear.

The question of legitimacy may be answered simply. One may correlate whatever one wish. The question is, however, whether it is done consistently and by exact un-

derstood

0	$1^a 3^b 1^c$	$1^a 3^b 2^c$	$1^a 2^b 3^c$	$1^a 2^b 4^c$	$1^a 1^b 5^c$	$1^a 1^b 2^c$	$1^a 1^b 3^c$
$3^a 1^b 1^c$	0	$3^a 3^b 1^c$	$3^a 2^b 2^c$	$3^a 2^b 3^c$	<i>$3^a 1^b 4^c$</i>	<i>$3^a 1^b 1^c$</i>	<i>$3^a 1^b 2^c$</i>
<i>$3^a 1^b 2^c$</i>	$3^a 3^b 1^c$	0	$3^a 2^b 1^c$	$3^a 2^b 2^c$	<i>$3^a 1^b 3^c$</i>	<i>$3^a 1^b 2^c$</i>	<i>$3^a 1^b 1^c$</i>
$2^a 1^b 3^c$	$2^a 3^b 2^c$	$2^a 3^b 1^c$	0	$2^a 2^b 1^c$	<i>$2^a 1^b 2^c$</i>	<i>$2^a 1^b 3^c$</i>	<i>$2^a 1^b 2^c$</i>
<i>$2^a 1^b 4^c$</i>	$2^a 3^b 3^c$	$2^a 3^b 2^c$	$2^a 2^b 1^c$	0	<i>$2^a 1^b 1^c$</i>	<i>$2^a 1^b 4^c$</i>	<i>$2^a 1^b 3^c$</i>
$1^a 1^b 5^c$	<i>$1^a 3^b 4^c$</i>	<i>$1^a 3^b 3^c$</i>	<i>$1^a 2^b 2^c$</i>	<i>$1^a 2^b 1^c$</i>	0	$1^a 1^b 5^c$	$1^a 1^b 4^c$
$1^a 1^b 2^c$	<i>$1^a 3^b 1^c$</i>	<i>$1^a 3^b 2^c$</i>	<i>$1^a 2^b 3^c$</i>	<i>$1^a 2^b 4^c$</i>	$1^a 1^b 5^c$	0	$1^a 1^b 3^c$
$1^a 1^b 3^c$	<i>$1^a 3^b 2^c$</i>	<i>$1^a 3^b 1^c$</i>	<i>$1^a 2^b 2^c$</i>	<i>$1^a 2^b 3^c$</i>	$1^a 1^b 4^c$	$1^a 1^b 3^c$	0

Figure 1. Universal matrix¹⁴ $U(a,b,c)$ (**D**val matrix^{9,10}) of 2,3-dimethylhexane.

Bold: Members of index $AB_L(a,b,c)$ resp. $AB_D(a,b,c)$ resp. $AB_{wm}(a,b,c)$ *Italics:* Members of index $C_L(a,b,c)$ resp. $C_D(a,b,c)$ resp. $C_{wm}(a,b,c)$

Normal script: Members of index $D_L(a,b,c)$ resp. $D_D(a,b,c)$ resp. $D_{wm}(a,b,c)$

Subscripts: $_{wm}$ – the index is derived from the whole matrix, $_L$ – the index is derived from the left half of the matrix, $_D$ – the index is derived from the right half of the matrix.

quivocal rules allowing the others to repeat it, and whether one arrives to better and reliable correlations.

Here are illustrated few cases where, on variation of exponent values on vertex degrees and vertex distances in the vertex-degree vertex-distance weighted indices, the results are invariant to molecular labeling and when not.

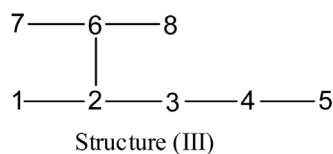
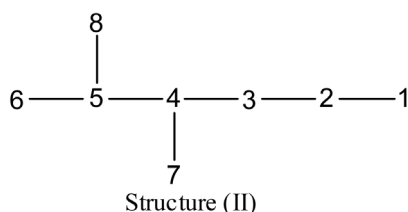
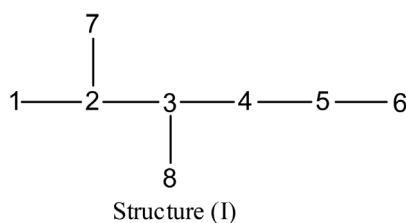
The approach described first in part by Ivanciuc^{9,10} and generalized later by Perdih and Perdih^{13,14} is notated here in the way used in ref.¹⁴

Let us have a Universal matrix¹⁴ $U(a,b,c)$ (first described by Ivanciuc^{9,10} as the **Dval** matrix), where its elements are defined as follows: $u_{ij} = v_i^a \times v_j^b \times d_{ij}^c$, where v_i and v_j are the vertex degrees of vertices i and j , d_{ij} is the distance between them, and the topological indices are derived from the matrix by summation of its elements. In such a case, a topological index TI is the function of exponents on vertex degrees and vertex distances, $TI(a,b,c) = f(a,b,c)$. For the case of 2,3-dimethylhexane the Universal matrix $U(a,b,c)$ is presented in Figure 1.

Having presented the Universal matrix and its elements, let us see what is the situation with the demand of derived topological indices to be invariant to molecular labeling.

2. Results and Discussion

To check whether the topological indices derived in the vertex-degree vertex-distance weighted way are invariants or not, they were compared first for the case of one of the octanes with the quite nonsymmetric structure, the 2,3-dimethylhexane (23M6) – Structures (I), (II), and



(III). Three possibilities of counting its graph elements are presented here – the usual one for 2,3-dimethylhexane (23M6), the reverse one, i.e. as if this compound were 4,5-dimethylhexane (45M6), as well as if it were 2-isopropylpentane (2iPr5). Other tested cases were 3E2M5 (notation variants 3E4M5, 3iPr5), 33M6 (44M6, 2E2M5), 223M5 (344M5, 2iBu4), and 233M5 (334M5, 2iPr2M4).

The indices chosen for illustration were taken from ref.¹⁴ as they were defined there; only a change in their notation is introduced here to distinguish more easily whether they are derived from the whole matrix (notation $_{wm}$) or from its left half (notation $_L$) or from its right half (notation $_D$):

- The whole-matrix ($_{wm}$) index¹³ $V_{wm}(a,b,c)$ being the sum of all non-diagonal elements in the Universal matrix $U(a,b,c)$.
- The $V_L(a,b,c)$ index¹³ being the sum of all non-diagonal elements in the left half of the Universal matrix.
- The $AB_{wm}(a,b,c)$ and $AB_L(a,b,c)$ index¹⁴ using non-diagonal elements of the Universal matrix, which contain the mutual contributions of interior vertices only.
- The $C_{wm}(a,b,c)$ and $C_L(a,b,c)$ index¹⁴ using non-diagonal elements of the Universal matrix, which contain the mutual contributions of one interior and one terminal vertex,
- as well as the $D_{wm}(a,b,c)$ and $D_L(a,b,c)$ index¹⁴ using non-diagonal elements, which contain the mutual contributions of terminal vertices only.

The relations between these indices are:¹⁴ $V_L(a,b,c) = AB_L(a,b,c) + C_L(a,b,c) + D_L(a,b,c)$, and $V_{wm}(a,b,c) = V_L(a,b,c) + V_D(a,b,c)$ where $V_D(a,b,c) = V_L(b,a,c)$.

In all tested cases mentioned here, the indices $D_{wm}(a,b,c)$, $D_L(a,b,c)$, and $D_D(a,b,c)$ are invariant to molecular labeling since all their components contain the factors 1^a and 1^b . Other tested indices are invariant to molecular labeling when they are derived from the whole matrix ($TI_{wm}(a,b,c)$), and in general not invariant to molecular labeling when they are derived from one of its halves ($TI_L(a,b,c)$, $TI_D(a,b,c)$). The reason for this is in the relation $TI_{wm}(a,b,c) = TI_L(a,b,c) + TI_D(a,b,c) = TI_L(a,b,c) + TI_L(b,a,c) = TI_D(a,b,c) + TI_D(b,a,c)$.

Among the indices, which are derived from one of the halves of the matrix, they are invariant to molecular labeling when the exponents on vertex degrees, a and b , are equal, i.e. when $a = b$. Part of these cases are the Wiener¹⁵ index W , the Randić⁸ index χ , etc.

Using the Randić⁸ index χ as well as its analogues having the value of exponents $a \neq b$ and the value of exponent $c = -\infty$, gives invariant indices as well. The Randić⁸ index $\chi = V_D(-1/2, -1/2, -\infty) = V_L(-1/2, -1/2, -\infty) = 1/2 V_{wm}(-1/2, -1/2, -\infty)$ is »doubly« invariant: due to $a = b$ and also due to $c = -\infty$. The value of exponent $c = -\infty$ has the consequence that the values of all matrix elements having the distances $d_{ij} > 1$ are set to the value of 0 (zero).

Next question is whether the indices, which are invariant to molecular labeling, and indices, which are not invariant to molecular labeling, both give rise to equally good correlations. This question is answered here by comparing results obtained using the indices $AB_{wm}(a,b,c)$ vs. $AB_L(a,b,c)$, $C_{wm}(a,b,c)$ vs. $C_L(a,b,c)$, and $V_{wm}(a,b,c)$ vs. $V_L(a,b,c)$ index. The subscript $_{wm}$ indicates that the index is derived from the whole matrix, whereas the subscript $_L$ indicates that the index is derived from the left half of the matrix. Not to forget, $TI_L(a,b,c) = TI_D(b,a,c)$. The best results obtained using two-digit values of exponents a , b , and c and expressed as the correlation coefficient R and standard error $S.E.$ are presented in Table 1 for the case of BP (boiling point) of octanes. The boiling point data were taken from Ren.¹⁶

Table 1. Correlations of selected indices derived from the whole ($_{wm}$) Universal matrix $U(a,b,c)$ as well as from its left half ($_L$), with the boiling point (BP) of all (18) octanes.

	Best Exponent			R	$S.E.$
	a	b	c		
$AB_{wm}(a,b,c)$	0.40	0.40	-0.48	0.783	3.92
$AB_L(a,b,c)$	2.8	3.9	1.04	-0.915	2.54
$C_{wm}(a,b,c)$	0.174	0.183	$-\infty$	-0.745	4.21
$C_L(a,b,c)$	$-\infty$	3.7	2.0	-0.936	2.23
$V_{wm}(a,b,c)$	$-\infty$	3.7	1.99	-0.931	2.31
$VL(a,b,c)$	$-\infty$	3.7	2.0	-0.936	2.21

Evidently, the indices derived from one of the halves of the matrix are better indices of a physicochemical property than those derived from the whole matrix. In other words, the indices, which are not invariant to molecular labeling, give rise to better results than those, which are invariant to molecular labeling. The indices, which are not invariant to molecular labeling, obviously contain a higher proportion of relevant information about the molecule in question than the indices, which are invariant. Among them, some indices derived from part of a half of the matrix are better indices than those derived from the entire half of the matrix.

The reason for this can be explained simply. Molecules of octanes are in general non-symmetric and non-symmetric matrices describe their structures. Symmetric molecules are special cases. The same seems to be valid for indices derived from their molecular structures. Indices, which are invariants to molecular labeling, are special cases, whereas indices, which are not invariant to molecular labeling, are general.

For practical purposes, it is thus advisable to use also indices, which are not invariant to molecular labeling. One has only to be careful to use for all compounds in question the same system of labeling of vertices as for the derivation of invariant indices.

That the topological indices do not need to be formal invariants, it is possible to demonstrate using particu-

lar matrix elements as topological indices. A matrix element of the Universal matrix, which represents the mutual contribution of two vertices to the value of it,¹⁷ may obtain different notation using different ways of labeling of vertices. However, in any of these cases, there are always the same two vertices that give rise to the same correlation, etc., to a physicochemical property of the compound in question. Let's see it in the case of 2,3-dimethylhexane.

The matrix element u_{72} (or, u_{27}) in the case of labeling of atoms in 2,3-dimethylhexane (Structure (I)) is the same as u_{85} (or, u_{58}) in the case of labeling as if it were 7,8-dimethylhexane (Structure (II)) and also the same as u_{68} (or, u_{86}) in the case of labeling as if it were 2-isopropylpentane (Structure (III)).

Although these matrix elements are seen formally as not being invariant to molecular labeling, they in fact are invariants since it is not the formal labeling of them that defines their relation towards the physicochemical property in question but the characteristics of the structural features, part of which are the vertices involved in mutual contribution to the value of the matrix element. In fact, the physicochemical properties of octanes (and any other compounds) do not depend on formal labeling of vertices in the graphs representing them but on structural features of their molecules, which are represented by elements of the Universal matrix representing their structure. For example, all of these differently enumerated matrix elements representing the same structural feature as the matrix element $u_{72}(-0.170, 0.30, -0.104)$, give rise to $R = -0.872$, $S.E. = 3.09$ to the boiling point of octanes, which is the best value among particular elements of the Universal matrix. These values are worse than in the case of the whole-matrix index $V_{wm}(a,b,c)$ presented in Table 1.

However, already the best combination of two matrix elements being $-0.9979 \times u_{63}(-3.1, -3.6, -2.0) + 0.0021 \times u_{74}(0.91, 0.74, 0.85)$ giving rise to $R = 0.950$, $S.E. = 1.97$, is better than the whole-matrix index $V_{wm}(a,b,c)$ presented in Table 1.

R and $S.E.$ to the boiling point of octanes of additional best combinations of more elements of the Universal matrix are presented in Table 2.

These combinations of matrix elements are each additionally better and in the last three cases $R \geq 0.99$ and

Table 2. Correlation of boiling points of octanes to the best combination of N matrix elements.

N	R	$S.E.$
1	-0.872	3.09
2	0.950	1.97
3	0.980	1.26
4	0.984	1.12
5	0.990	0.90
6	0.995	0.60
7	0.996	0.58

$S.E. < 1\text{ }^{\circ}\text{C}$, much better than in the other QSPR models for alkanes.¹⁸ Therefore, using in the case of BP of octanes the information contained in only six (6) out of fifty-six (56) non-diagonal elements of the Universal matrix gives rise to $R > 0.99$.

Similarly to particular matrix elements, the indices¹⁴ $A(a,b,c)$, $B(a,b,c)$, $AB(a,b,c)$, and $C(a,b,c)$ are not invariant regarding different ways of labeling of vertices. They are, however, invariant regarding the structural features, which they reflect. Due to this and in order to avoid confusions in marking of indices, it is advisable to use only one and the same system of their labeling, for example the system of IUPAC Nomenclature.

3. Conclusion

Particular elements of the Universal (vertex-degree vertex-distance weighted) matrix can be used as topological indices. Regarding different ways of labeling of vertices presented above, particular matrix elements are not invariant to molecular labeling. Regarding the structural features, which they represent, they are invariants since regardless how we enumerate them, there are only particular combinations of vertices representing particular structural features.

Some combinations of matrix elements, weighted for the vertex degrees, the distance between them, and their relative contribution are very good descriptors of physicochemical properties of octanes regardless how we enumerate them.

4. Acknowledgement

Valuable discussions with M. Randić are thankfully acknowledged.

Povzetek

Vsak element Univerzalne matrike U predstavlja vzajemen prispevek dveh točk z uteženima njunima vrednostma in uteženo vrednostjo razdalje med njima. Glede na različne načine oštevilčenja točk posamezni elementi matrike niso invariantni. Glede na strukturne značilnosti, ki jih predstavljajo, pa so invariantni, ker vsak par točk predstavlja eno strukturno značilnost molekule. Nekateri elementi matrike so boljši topološki indeksi za posamezne fizikalno-kemijske lastnosti kot drugi. Nekatere utežene kombinacije elementov Univerzalne matrike U so zelo dobri topološki indeksi za fizikalno-kemijske lastnosti oktanov ne glede na to, kako oštevilčimo njihove točke.

5. References

1. A. T. Balaban (Ed.), From chemical topology to three-dimensional geometry. Plenum Press, New York and London, 1997.
2. L. Kier, L. Hall, Molecular structure description. Academic Press, San Diego, 1999.
3. R. Todeschini, V. Consonni, Handbook of Molecular Descriptors. Wiley-VCH, Weinheim, 2000.
<http://dx.doi.org/10.1002/9783527613106>
4. M. Karelson, Molecular Descriptors in QSAR/QSPR. John Wiley & Sons, New York, 2000.
5. J. Devillers, A. T. Balaban (Eds.), Topological indices and related descriptors in QSAR and QSPR. Gordon and Breach, Amsterdam, 2000.
6. H. Timmerman, R. Todeschini, V. Consonni, R. Mannhold, H. Kubinyi, Handbook of Molecular Descriptors. Wiley-VCH, Weinheim, 2002.
7. R. Todeschini, V. Consonni, Molecular Descriptors for Chemoinformatics (2 volumes), Wiley-VCH, Weinheim, 2009.
<http://dx.doi.org/10.1002/9783527628766>
8. M. Randić, *J. Am. Chem. Soc.* **1975**, 97, 6609–6615.
<http://dx.doi.org/10.1021/ja00856a001>
9. O. Ivanciuc, *Rev. Roum. Chim.* **1999**, 44, 519–528.
10. O. Ivanciuc, *Rev. Roum. Chim.* **2000**, 45, 587–596.
11. A. Perdih, B. Perdih, *Acta Chim. Slov.* **2002**, 49, 67–110.
12. A. Perdih, B. Perdih, *Acta Chim. Slov.* **2003**, 50, 95–114.
13. A. Perdih, B. Perdih, *Acta Chim. Slov.* **2004**, 51, 589–609.
14. A. Perdih, F. Perdih, *Acta Chim. Slov.* **2006**, 53, 180–190.
15. H. Wiener, *J. Am. Chem. Soc.* **1947**, 69, 17–20.
<http://dx.doi.org/10.1021/ja01193a005>
16. B. Ren, *J. Chem. Inf. Comput. Sci.* **1999**, 39, 139–143.
<http://dx.doi.org/10.1021/ci980098p>
17. A. Perdih, B. Perdih, *Indian J. Chem.* **2003**, 42A, 1219–1226.
18. Z. Mihalić, N. Trinajstić, *J. Chem. Educ.* **1992**, 69, 701–712.
<http://dx.doi.org/10.1021/ed069p701>