Elements of an Universal Matrix as Topological Indices for Physicochemical Properties of Octanes

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: 10-04-2015

Abstract

Some of the elements of the Universal matrix and their combinations are useful topological indices of physicochemical properties of octanes. Whereas some of the single elements of the Universal matrix give rise to $0.70 < |R| < 0.99$, mutually optimized combinations of only four to six out of 56 of them in the Universal matrix of octanes give rise to $R > 0.99$ and in the worst cases to $R > 0.98$. Also a new measure of goodness of correlation, the information content in the topological index, $IC(\%)$, is introduced. Structural interpretation of some of the physicochemical properties of octanes is demonstrated as well as of the contribution by the most useful elements of the Universal matrix.

Keywords: Information content; Goodness of topological indices; Universal matrix elements; Octanes; Structural interpretation of: Octane Number, van der Waals constant $a_0$, Boiling point, Refractive index, Critical temperature, Critical density, Vapor pressure

1. Introduction

Mathematical methods occupy an eminent place in the field of prediction of properties and activities of chemical compounds, and even materials. These methods, known under the acronym QSPR/QSAR (quantitative-structure-property or structure-activity relationship) use also graph-theoretical descriptors, where molecules are seen as chemical graphs, i.e. as a set of vertices attached to each other by a set of non-metrical connections. These descriptors are proposed as topological indices. They are the simplest means of describing the structure of a molecule, characterizing it by a single number.

There is known a plethora of topological indices. After their compilations, a huge number of new ones has been described and new and new ones are being developed, cf. e.g. A substantial part of topological indices is derived from one or another matrix associated with molecular topology. Ivanciuc presented the Dval matrix and its characteristics, and we have shown that this matrix represents a step in unification of several matrices which had been used to derive topological indices, i.e. of the adjacency matrix, the distance matrix, the reciprocal distance matrix, etc, being thus an Universal matrix. The characteristics of some groups of topological indices derived by means of this generalized vertex-degree vertex-distance matrix have been studied and there was demonstrated the usefulness of some of those new topological indices.

The well known topological indices $W$, $RW$, $\chi$, for example, are composed of the one half of the sum of all 56 matrix elements $u_{(a, b, c)}$ of the Universal matrix, where at $W$ $(a, b, c) = (0, 0, 1)$; at $RW$ $(a, b, c) = (0, 0, -1)$; and at $\chi$ $(a, b, c) = (-1/2, -1/2, -\infty)$.

There arose the question whether particular elements of the Universal matrix as well as their combinations are good topological indices or not. It has been demonstrated that although particular elements of the Universal matrix are not invariant to molecular labelling, they are invariant regarding the structural features of octanes, and the topological indices, which are not invariant to molecular labelling give rise to better correlations than the topological indices, which are invariant to molecular labeling.

For this reason, the elements of the Universal matrix and their mutually optimized combinations have been systematically studied and the results are presented here.
2. Data and Definitions

The origin of data of physicochemical properties (PCP), as well as the notations of octanes have been presented elsewhere.\textsuperscript{14} The data are presented in Appendix 1. Correlation between physicochemical properties of octanes used in present study is presented in Appendix 2. Grouping of physicochemical properties of octanes by their intercorrelation in Appendix 2 and put into subgroups according to the correlation coefficient with the best topological indices (TI) based on grid values of exponents\textsuperscript{14} in TI(a, b, c) are presented in Table 1.

For demonstration of usefulness of elements of the Universal matrix as well as of their combinations, there was chosen from subgroup 1a MON as a physicochemical property having the best correlations with previously tested topological indices. As a less good example was taken from the subgroup 1b $T_c^2/P_c$ representing the van der Waals parameter $a_0$ with constants omitted. From the group 2, BP was chosen. From subgroup 2,3b, Tc was selected and from the subgroup 3a $n_D$. As two of the worst cases were chosen from the group 4 $d_c$, and from the group 5 $\log VP$.

2. 1. Universal Matrix and its Elements

The Universal matrix\textsuperscript{14} $U(a, b, c)$ (first described by Ivanciuc\textsuperscript{12,13} as the Dval matrix) has its elements defined here as follows: $u_{ij}(a, b, c) = v_i^a v_j^b d_{ij}^c$, where $v_i$ and $v_j$ are the vertex degrees of vertices $i$ and $j$, $d_{ij}$ is the distance between them. Each element of the Universal matrix is a function of exponents on vertex degrees and vertex distances, $u_{ij}(a, b, c) = f(a, b, c)$. For easier comparison, the Universal matrix relating to 2,3-dimethylhexane and used here is presented in Appendix 3. The relation between matrix elements from the left side of the Universal matrix and from its right side is simple, for example: $u_{52}(a, b, c) = u_{25}(b, a, c)$. The elements of the Universal matrix, which contain the factor $1^a$ or $1^b$ resp. $1^c$ are given in the form demonstrated here for $u_{32}(a, b, c) = u_{32}(a, b, 1^c)$ to demonstrate that the factors $1^a$ or $1^b$ resp. $1^c$ do not influence the usefulness of the topological index.

2. 2. Exponent Values

The first step to assess the usefulness of elements of the Universal matrix is the goodness of their correlation with the physicochemical properties of octanes. To assess where approximately the maxima in absolute values of correlation coefficient $R$ are positioned in the space of exponents $a$, $b$, and $c$, a 3D grid of values of exponents was applied and the values of correlation coefficients at those combinations of values were derived. The exponent values $-5$, $-4$, $-3$, $-2$, $-1$, $-0.5$, $-0.3$, $-0.2$, $-0.1$, $0$, $0.1$, $0.2$, $0.3$, $0.5$, $1$, $2$, $3$, $4$, and $5$ were chosen as the grid points in all three dimensions of exponents. The true maximum of the correlation coefficient can then be approached by exponent optimization using also two-digit and, if necessary, three-digit values of exponents besides the grid values. The values of exponents were limited to at most three decimals.

2. 3. Goodness of Topological Indices

To illustrate the potential goodness of topological indices, the correlation coefficient $R$ and standard error $S$

| Group No. | $| R |$ | Physicochemical property, PCP |
|-----------|------|-------------------------------|
| Group 1   |      |                               |
| Subgroup 1a | $| R | > 0.99$  | BON, RON, MON, BP/Tc, Tc/Pc, $\omega$ |
| Subgroup 1b | $0.99 > | R | > 0.95$  | Tc$^2$/Pc, $S$, $R^2$, $C$ |
| Subgroup 1c | $0.95 > | R | > 0.90$  | $\alpha c$ |
| Between the group 1 and 2 |      |                               |
| Subgroup 1,2a | $0.99 > | R | > 0.95$  | $\Delta H_v$ |
| Subgroup 1,2b | $0.95 > | R | > 0.90$  | $A$ |
| Between the group 1 and 3 |      |                               |
| Subgroup 1,3a | $0.99 > | R | > 0.95$  | $P_c$ |
| Group 2     |      |                               |
| Subgroup 2,3a | $0.95 > | R | > 0.90$  | $BP, \Delta H_f^g$ |
| Subgroup 2,3b | $| R | < 0.90$  | $ST$ |
| Group 3     |      |                               |
| Subgroup 3a | $0.99 > | R | > 0.95$  | $n_D, d, V_m$ |
| Subgroup 3b | $0.95 > | R | > 0.90$  | $MR$ |
| Group 4     | $| R | < 0.90$  | $d_c, V_c$ |
| Group 5     | $| R | < 0.90$  | $\log VP, Z_c$ |

Acronyms for physicochemical properties are given in Appendix 1.
are generally used. Here is proposed also another quantity, the information content \( (IC) \) about the physicochemical property in question contained in the topological index (index combination) in question.

The information content \( (IC) \) in the topological index (index combination) in question about the physicochemical property (PCP) of octanes in question is defined as follows:

\[
IC = 1 - \frac{\sum (PCP_{exp} - PCP_{calc})^2 / \sum (PCP_{exp} - PCP_{av})^2)^{1/2}}
\]

where \( PCP_{exp} \) means experimental PCP data of octanes, \( PCP_{calc} \) those calculated from the topological index (index combination) values, and \( PCP_{av} \) is the average of \( PCP_{exp} \).

To the experimental PCP data of octanes, \( PCP_{exp} \), is ascribed the information content \( IC = 1 \), whereas to the average of \( PCP_{exp} \) data of octanes, \( PCP_{av} \), is ascribed the information content \( IC = 0 \) since \( PCP_{av} \) does not contain any information about the contribution of branching in octanes to the value of \( IC \) in question.

The value of \( IC \) contributed by particular matrix elements in the index combination is given normalized in such a way that the sum of all particular \( IC \) is equal to the value of \( IC \) of the topological index combination.

### 2. 4. Topological Index Combination

To assess the usefulness of the topological index combination \( (TI_{comb}) \) composed of two or several elements of the Universal matrix the approach:

\[
TI_{comb} = \sum u_{ij}(a_i, b_j, c_{ij}) \times k_{ij}
\]

was used, where

\[
\sum \abs(k_{ij}) = 1 \text{ and } 0 < \abs(k_{ij}) < 1
\]

and the exponents \( a_i, b_j, c_{ij} \) as well as the smallest \( k_{ij} \) have two significant digits. The exponents \( a_i, b_j, c_{ij} \) as well as the factors \( k_{ij} \) are mutually optimized to reach the highest \( R \) value possible.

### 3. Results

As the first step to assess the usefulness of particular elements of the Universal matrix, \( u_{3ij}(a, b, c) \), as topological indices is the goodness of their correlation with the physicochemical properties of octanes.

The best correlations between tested physicochemical properties (PCP) and \( u_{3ij}(a, b, c) \) elements using grid values of exponents are presented in the form \( R_{max grid} \) (PCP, \( u_{3ij} ) \) as follows: 0.99 > (MON, \( u_{31} ) > (RON, \( u_{32} ) > (PC, \( u_{33} ) > (BON, \( u_{34} ) > (BP/Tc, \( u_{35} ) > (R_2, \( u_{36} ) > (Te^2/Pc, \( u_{37} ) > (B, \( u_{38} ) > (nD, \( u_{39} ) > (\delta c, \( u_{40} ) > (Te/Pc, \( u_{41} ) > 0.90 > (Tc, \( u_{42} ) > (ST, \( u_{43} ) > (V_m, \( u_{44} ) > (d, \( u_{45} ) > (C, \( u_{46} ) > (A, \( u_{47} ) > (S, \( u_{48} ) > (BP, \( u_{49} ) > (CED, \( u_{50} ) > (\Delta Hv, \( u_{51} ) > (Sol.par., \( u_{52} ) > (MR, \( u_{53} ) > (Vc, \( u_{54} ) > 0.85 > (dc, \( u_{55} ) > (\delta c, \( u_{56} ) > 0.80 > (logVP, \( u_{57} ) > (Zc, \( u_{58} ) > 0.70.

Particular elements of the Universal matrix are thus quite good topological indices for some of the tested physicochemical properties. The topological index \( u_{3ij}(a, b, c) \) is the best one in 10 cases, \( u_{35}(a, b, c) \) in 6 cases, \( u_{65}(a, b, c) \) and \( u_{72}(a, b, c) \) in 3 cases, \( u_{52}(a, b, c) \) in two cases, \( u_{31}(a, b, c) \), \( u_{36}(a, b, c) \), \( u_{45}(a, b, c) \), \( u_{66}(a, b, c) \), and \( u_{84}(a, b, c) \) in one case each out of 29 cases.

The usefulness of particular elements of the Universal matrix increases on going from grid values of exponents to two-digit values of them as well as on using mutual optimization of combination of two or more matrix elements used in two-digit values of exponents. This is demonstrated in the case of octanes in Tables 2, 3, and 4 for physicochemical properties MON, \( Tc^2/Pc \), BP, \( nD \), \( Tc \), \( dc \), and \( logVP \).

In Tables 2 through 4 can be seen that using only 4 out of 56 non-diagonal matrix elements of the Universal matrix, after optimization of their exponent values and their relative contribution, there can be achieved in the best tested case (MON) \( R = 0.996, S = 3.10, IC = 96.1\% \) for the combination \( TI_{comb} = -0.8349 \times u_{75}(0.52, 4.3, -3.7) - 0.0268 \times u_{76}(1.18, 4.3, 1.97) - 0.1335 \times u_{42}(0.25, 0.68, 0.147) - 0.0048 \times u_{63}(-1.70, -3.3, 3.6), \) whereas in the worst tested case (dc) \( R = 0.944, S = 0.0028, IC = 67.0\% \) for the combination \( TI_{comb} = -0.02874 \times u_{72} \) (see Table 2).

**Table 2.** Best observed correlation coefficients \( (R) \) between the vertex-degree vertex-distance optimized matrix element (or their combination) and physicochemical property (PCP) of octanes.

| No. of matrix elements | MON | \( Tc^2/Pc \) | BP | \( nD \) | \( Tc \) | \( dc \) | logVP |
|------------------------|-----|-------------|----|--------|--------|-------|-------|
| one, grid*             | -0.975 | 0.927 | -0.871 | -0.922 | -0.897 | -0.837 | 0.753 |
| one, two digit*        | -0.978 | 0.930 | -0.872 | -0.923 | -0.901 | -0.841 | 0.754 |
| two                    | 0.993 | 0.980 | 0.950 | 0.977 | 0.966 | 0.908 | 0.887 |
| three                  | 0.996 | 0.989 | 0.980 | 0.981 | 0.975 | 0.923 | 0.902 |
| four                   | 0.996 | 0.995 | 0.984 | 0.989 | 0.980 | 0.944 | 0.954 |
| five                   | 0.9994 | 0.998 | 0.990 | 0.993 | 0.983 | 0.964 | 0.973 |
| six                    | 0.9996 | 0.999 | 0.995 | 0.995 | 0.986 | 0.986 | 0.986 |

* grid resp. two-digit values of exponents

_Perdih: Elements of an Universal Matrix as Topological Indices_ ...
Extrapolation of the best observed regression data to the structure of 2,2,3,3-tetramethylbutane indicates that its missing MON value would be around 98.5, and if 2,2,3,3-tetramethylbutane would exist at normal pressure and 20 °C in the liquid state, it would have nD of around 1.429 and logVP of around 3.51.

The illustration, which elements of the Universal matrix, values of their exponents, and their relative contribution give rise to the values presented in Table 2 through 4 for BP of octanes is as follows:

| No. of matrix elements | MON | Tc²/Pc | BP | PCP | nD | Tc | dc | logVP |
|------------------------|-----|--------|----|-----|----|----|----|-------|
| one, grid              | 7.56| 1623   | 3.10| 0.00209 | 3.87 | 0.0047 | 0.117 |
| one, two digit         | 7.11| 1590   | 3.09| 0.00208 | 3.81 | 0.0046 | 0.117 |
| two                    | 4.01| 861    | 1.97| 0.00114 | 2.27 | 0.0036 | 0.082 |
| three                  | 3.12| 652    | 1.26| 0.00103 | 1.94 | 0.0033 | 0.077 |
| four                   | 3.10| 444    | 1.12| 0.00079 | 1.75 | 0.0028 | 0.053 |
| five                   | 1.19| 263    | 0.90| 0.00064 | 1.59 | 0.0023 | 0.041 |
| six                    | 0.91| 194    | 0.60| 0.00054 | 1.45 | 0.0014 | 0.030 |

The sign of the factor k defines the sign of the product $k^{ij}u_u^i v_v^j d_{dij}^c = k^{ij}u_u^i (a, b, c)$.

Four matrix elements:

$-0.995129 \times u_{63}(-3.1, -3.9, -1.92) + 0.002096 \times u_{64}(1.09, 0.0040, 1.12) - 0.002696 \times u_{42}(-0.65, 4.2, -5.9) + 7.9E-05 \times u_{72}(-0.91, 3.2, 0.25), R = 0.984, S = 1.12, IC = 82.3%$

Five matrix elements:

$-0.991771 \times u_{63}(-1.69, -4.2, -1.21) + 0.001858 \times u_{64}(1.11, 0.32, 1.29) - 0.005132 \times u_{42}(-0.20, 4.2, -5.9) + 0.000102 \times u_{72}(-3.2, -0.055) + 0.001137 \times u_{32}(-0.59, 2.5, 1^c), R = 0.990, S = 0.90, IC = 85.8%$

Six matrix elements:

$-0.95898 \times u_{63}(-0.98, -4.2, -0.94) + 0.002541 \times u_{64}(1.21, -0.73, 1.20) - 0.005451 \times u_{42}(0.21, 4.2, -5.9) + 0.000205 \times u_{43}(-\infty, 3.2, -0.44) + 0.001426 \times u_{32}(-1.39, 2.5, 1^c) + 0.031397 \times u_{72}(-0.26, -0.64, 0.80), R = 0.995, S = 0.60, IC = 90.5%$

The contribution of particular matrix elements ($u_{63}, u_{42}, u_{43}, u_{72}, u_{73},$ and $u_{72}$) to the optimized combined topological index derived from them in the case of BP is presented in Figure 1.

Individual goodness of elements of the Universal matrix in their best combination presented in Figure 1 is presented in Table 5, whereas their goodness observed in...
In Table 5 and 6 can be seen that their goodness is better in the best individual cases than in their contribution to the collective result, but the collective goodness is decisively better. Such a situation has been observed in all tested cases.

4. Discussion

Several well known indices, e.g. the Wiener index, the Randić index, etc, are in fact derived from the Universal matrix using the grid values of exponents.

It has been observed that the first digit in the exponent, e.g. 2, defines in most tested cases the first three decimals of the correlation coefficient. The second digit, e.g. 2.3, improves in most tested cases the value of the third to fifth decimal, depending on how far from the best value of the exponent is its one-digit grid value approximation. The third digit of the exponents a, b, and c, e.g. 2.31, improves the value of the fifth or higher decimal of the correlation coefficient. In the space of exponents a, b, and c, there are observed several local maxima of correlation coefficient $R$.

For our purpose, in the first step of assessment three decimals in the correlation coefficient are sufficient, therefore in our first step we use one-digit grid value of exponents. For optimization, five decimals in the value of the correlation coefficient are considered sufficient, therefore
only two digits in the value of exponents \(a, b, c\) and factors \(k\) are needed.

Some of the elements of the Universal matrix \(U(a, b, c)\), i.e. \(u_i(a, b, c)\), proved to be useful topological indices of some physicochemical properties of octanes. Already using grid values of exponents, there are \(0.99 > |R_{\text{mad}}| > 0.95\) at \((\text{MON, } u_{52}) > (\text{RON, } u_{53}) > (\text{Pc, } u_{53}) > (\text{BON, } u_{75})\). The improvements by using optimized two digit values of exponents, by combinations of particular matrix elements with mutually optimized two-digit values of exponents in the form \(\sum k_{ij} u_{ij}(a_i, b_j, c_{ij})\) are demonstrated in Tables 2 through 4.

The improvement of the \(R\) in two of the worst cases tested, from 0.753 to 0.986 in the case of logVP, and from \(-0.837\) to 0.986 in the case of dc, shows that the approach using elements of the Universal matrix as topological indices and especially the combinations of them by using mutually optimized two-digit values of exponents is a promising one.

The information content \((IC)\) in the topological index (or index combination) in question about the physicochemical property (PCP) of octanes in question as defined under the heading Data and definitions proved to be linearly and negatively correlated with one of the important measures of goodness of correlation, \(S\). The relation between \(IC\) and \(S\) is as follows:

\[
IC = 1 - S(\text{PCP}_{\text{exp}}; \text{PCP}_{\text{calc}})/ S(\text{PCP}_{\text{exp}}; \text{PCP}_{\text{av}})
\]

Having this relation, there arises the question, which of them is more useful, \(S\) or \(IC\). Each of them has its own type of usefulness. \(IC\) is in some way more illustrative than \(S\) since it directly indicates the information content contained in the tested topological index (index combination). It is an easily comprehensible direct indication of goodness of the topological index (index combination).

\(S\) is an inverse measure. Inverse measures are in general less easy to comprehend. And, \(S\) can not be used for inter-PCP comparisons of goodness of topological indices.

On the other hand, the \(IC\) is not dependent on numerical values of PCP in question and can be used also for inter-PCP comparisons of goodness of topological indices. In this respect it is more similar to the usefulness of the correlation coefficient \(R\) and its use together with \(R\) is suggested. However, \(R\) is dependent on the number of regression parameters, and \(IC\) is not. Therefore \(IC\) is a better criterion for the goodness of model. In order not to mistake \(IC\) data for \(R\) data, it is suggested to express \(IC\) in \%. This way we have three different indications of goodness of correlation, \(-1 < R < 1\), then \(0 < IC < 100\) (%), and \(S\). The parallelism of values of \(R\) and \(IC\) is illustrated in Table 7. Thus, if \(|R| = 0.99\) is considered as the lower limit of sufficient goodness of a topological index,\(^2\) then such a lower limit would be also \(IC = 86\%\). One can, of course, put also a reverse consideration. For example, if one defines that \(IC = 90\%\) or any other \(IC\) value is a proper criterion, then \(|R| = 0.995\) or another \(|R|\) value would result as an additional criterion.

The criterion, how to choose the upper reasonable limit of our demand for \(|R|\) and \(IC\) is the uncertainty of the experimental data. For example, when the values of a physicochemical property are known to three significant digits as e.g. at dc, and when the uncertainty of the third digit is \(\pm 1\), then due to uncertainties in the experimental data it is reasonable to demand \(|R|\) of about 0.995 and \(IC\) of about 90\%. If the uncertainty of the third digit is \(\pm 2\), then due to uncertainties in the experimental data there would be reasonable to demand \(|R|\) of about 0.98 and \(IC\) of about 80\%.

Using \(IC\) there arises the question to which quantity to ascribe not having any information about the differences in the physicochemical property in question among different compounds, for example among isomers of octane. Among octanes, one could suggest its average value as done above, but also the value at \(n\)-octane or even at cyclooctane, which graph contains no vertices of degree one. For practical reasons, since there may not be known the PCP value of a particular octane, it is suggested to ascribe the value of zero information to the average of available data. If we take a different basis for the value of zero information, the \(IC\) data will be slightly different, but all approaching the value of 1 as the correlation is improving.

As a rule of thumb can be concluded that if the correlation coefficient using optimized values of exponents in an element of the Universal matrix is sufficiently good, e.g. \(|R| > 0.99\),\(^2\) then such a topological index can be used as a predictor of values of that physicochemical property. If the correlation coefficient in such a case is not sufficiently good, then the combination of two or more elements of the Universal matrix representing the mutual contribution of graph vertices to the value of the topological index\(^2\) should be tested, mutually optimizing their exponents and their relative contribution.

Let us look at the results from these points of view. If we present in Table 8 the \(IC\) data of individual matrix

| \(|R|\)  | \(IC(\%)\) |
|------|----------|
| 0.9996 | 97.3     |
| 0.99  | 95.5     |
| 0.995 | 90.0     |
| 0.990 | 85.9     |
| 0.980 | 80.0     |
| 0.970 | 75.5     |
| 0.950 | 68.7     |
| 0.900 | 56.0     |
| 0.866 | 50.0     |
| 0.800 | 40.0     |
| 0.714 | 30.0     |
| 0.600 | 20.0     |
| 0.436 | 10.0     |
elements in the best combinations of six of them, the results of which are presented in Table 4, we can see that most of information is contained in the mutually optimized combination of the best three or four matrix elements. In the worst case (dc) it is contained in five of them out of 56 matrix elements.

Here is the question how to continue the improvement. One possibility is to use the brute force optimization testing all possible combinations of matrix elements. Another possibility is to look in the graph of PCP vs. matrix elements combination, which isomers depart the most from the linear regression line. An example is given in Table 9 and 10 for the case of dc, which is one of the worst examples in Tables 2 through 4.

### Table 8. Information content about the physicochemical property of octanes (IC, %) contained in particular matrix elements in the best combination of six of them.

| Matrix element | MON | Tc²/Pc | PCP | BP | u₉ | Tc | dc | logVP |
|----------------|-----|--------|-----|----|----|----|----|-------|
| best           | 37.8| 37.9   | 25.0| 43.2| 46.8| 27.5| 29.8|
| second best    | 26.5| 28.2   | 23.8| 25.0| 20.1| 15.4| 28.1|
| third best     | 17.0| 21.8   | 20.5| 15.3| 11.4| 13.0| 10.9|
| fourth best    | 11.3| 6.2    | 10.5| 6.0 | 4.0 | 12.9| 5.7 |
| fifth best     | 2.5 | 1.0    | 7.4 | 0.2 | 1.0 | 12.5| 5.1 |
| worst          | 2.1 | 0.4    | 3.3 | 0.1 | 0.1 | 1.8 | 3.7 |
| ΣIC            | 97.3| 95.5   | 90.5| 89.9| 83.4| 83.1| 83.3|

Table 9. The best combination of four matrix elements in the case of dc.

| uᵢⱼ × kᵢⱼ | R   | IC (%) |
|------------|-----|--------|
| u₈₅(1a, –3.7, 2.4) × 0.00035 | 0.626 | 20.6 |
| u₆₆(–3.2, –3, 1) × –0.90395 | 0.592 | 18.2 |
| u₆₇(0.93, –1.62, –3.2) × –0.06696 | –0.532 | 14.4 |
| u₇₈(0.121, 0.55, –1.02) × –0.02874 | 0.522 | 13.8 |
| Σuᵢⱼ × kᵢⱼ | 0.944 | 67.0 |

Table 10. The largest differences dc exp – dc calc for the case of the best four mutually optimized matrix elements.

| Isomer | dc exp – dc calc |
|--------|------------------|
| 33M6   | 0.0046           |
| 223M5  | 0.0045           |
| 234M5  | 0.0031           |
| 4M7    | 0.0019           |
| 22M6   | 0.0016           |
| Oct    | –0.0013          |
| 233M5  | –0.0032          |
| 224M5  | –0.0038          |
| 23M6   | –0.0063          |

In Table 10 we can see that the largest difference is at the octane isomers branched at the vertices No. 2, 3, and 4. The matrix elements containing information about them are u₃₂₂, u₄₂₂, u₄₃₃, etc. The matrix element u₄₃₃ has been already one of the four best ones. Therefore we start testing first u₃₂₂ and u₄₂₂, and continue with other ones containing the information about said vertices. The result using the optimized best combination of six matrix elements gives rise to a correlation, Table 11, \( R = 0.986 \), which is close to \( R = 0.99 \).

### Table 11. The best combination of six matrix elements in the case of dc.

| uᵢⱼ × kᵢⱼ | R   | IC (%) |
|------------|-----|--------|
| u₈₅(1a, –2.7, –0.134) × –0.3545 | 0.758 | 27.5 |
| u₅₄(–2.9, –2.6, –1.39) × 0.1933 | 0.593 | 15.4 |
| u₆₅(0.65, 2.7, 0.88) × –0.0052 | 0.550 | 13.0 |
| u₆₇(–0.075, –3.1, –0.37) × –0.0654 | –0.547 | 12.9 |
| u₇₈(–0.13, –0.38, 2) × 0.3177 | –0.539 | 12.5 |
| u₈₉(0.67, –2.4, 1) × 0.0639 | 0.210 | 1.8 |
| Σuᵢⱼ × kᵢⱼ | 0.986 | 83.1 |

So, the use of mutually optimized combination of elements of the Universal matrix is promising to reach good correlations.

There is also to distinguish, which matrix element contributes the most to good correlation, and which one contributes the most to the «numerical volume» of the combined index. At MON this is not expressed as evidently as at Tc²/Pc, BP, nD, Tc, and especially at dc and logVP. In the case of dc, Table 11, there contributes the matrix element \( u₃₃(1, –2.7, –0.134) \) the most to the observed correlation of the combined index, whereas the matrix element \( u₅₃(–0.13, –0.38, 2) \) contributes the most to the «numerical volume» of the combined index presented in Table 11 as \( Σuᵢⱼ × kᵢⱼ \).

There arises also a principal question, whether the best combination of six matrix elements presented above is an overparametrized situation or not. Counting the number of factors \( kᵢⱼ \) and exponents \( a, b, c \) in \( Σuᵢⱼ(a, b, c) × kᵢⱼ \) being 24 in the case of 18 octanes seems to confirm the overparametrization. However, one must compare this situation from the same point of view, i.e. from the point of view of the Universal matrix, also with the situation in well known topological indices, e.g. the Wiener index.
Wiener index is felt as a single number (single parameter) for each isomer. From the point of view of the Universal matrix one observes that in Wiener index, which is one half of the sum of all (in the case of octanes 56) elements of the Universal matrix, there are contained in the case of octanes in derivation of Wiener index 225 parameters giving rise to a single number of the Wiener index value. This is about one order of magnitude more parameters than in the best combination of six matrix elements presented above. Also in the case of the best combination of six out of 56 matrix elements the result is a single number, as in the case of the Wiener index. The situation in the case of the best combination of six matrix elements presented above is thus, compared to the situation at the Wiener index, not overparametrized.

4.1. Degeneracy of Elements of the Universal Matrix

A standard question about topological indices is the question of their degeneracy. Elements of the Universal matrix of octanes are quite degenerated. All of them are totally degenerated when a = b = c = 0, being \( u_{ij}(a, b, c) = 1 \), as well as when a = b = c = -\( \infty \), being \( u_{ij}(a, b, c) = 0 \). If a ≠ b ≠ c, there are totally degenerated the following elements of the Universal matrix of octanes: \( u_{21}(a, b, c), u_{31}(a, b, c), u_{71}(a, b, c), u_{42}(a, b, c), u_{61}(a, b, c), u_{53}(a, b, c), u_{64}(a, b, c), \) and \( u_{54}(a, b, c) \). If a = b ≠ c ≠ 0 and a = b = c ≠ \( \infty \), the elements of the Universal matrix of octanes are not totally degenerated. When a ≠ b ≠ c, then the degeneracy is (written in a shorthand way) \( u_{21} > u_{31} > u_{41} > u_{43} > u_{87} > u_{42} > u_{61} > u_{53} > u_{64} > u_{71} > u_{83} > u_{84} > u_{65} > u_{83} > u_{33} > u_{65} > u_{73} > u_{74} > u_{82} > u_{86} > u_{82} > u_{86} > u_{32} > u_{34} > u_{35} > u_{37} > u_{38} > u_{39} > u_{40} > u_{41} > u_{43} > u_{45} > u_{57} > u_{67} > u_{73} > u_{74} > u_{75} > u_{76} > u_{78} > u_{79} > u_{80} > u_{82} > u_{86} \). Lower degeneracy parallels somewhat the higher usefulness of exponents of \( u_{53} > u_{54} > u_{55} > u_{56} > u_{72} > u_{73} > u_{74} > u_{75} > u_{76} > u_{78} > u_{79} > u_{80} > u_{82} > u_{86} \). Degeneracy decreases, in several cases drastically, when mutually optimized exponents are used in combinations of matrix elements. For example, in one of the worst cases of tested physicochemical properties of octanes, dc, in the best observed combination of two of matrix elements, \( u_{43}(0.147, 0.72, -0.32), R = 0.908, S = 0.0036, IC = 58.1\% \), there are four pairs of isomers having equal value of the combined topological index. In the best observed combination of three matrix elements, \( u_{53}(1^e, -0.49, 0.51), u_{56}(\infty, 2.7, 1.17), \) and \( u_{55}(0.20, -2.0, -1.20), R = 0.923, S = 0.0033, IC = 61.5\% \), there are two pairs of isomers having equal value of the combined topological index. In the best observed combination of four matrix elements, in \( u_{53}(1^e, -0.94, 0.70), u_{56}(0.43, 2.8, 1.12), u_{65}(1.38, -3.1, -0.38), \) and \( u_{66}(\infty, -2.5, -2.1), R = 0.943, S = 0.0028, IC = 66.8\% \), there are two pairs of isomers having equal value of the combined topological index as well, whereas in the combination of \( u_{53}(3.2, -3.0, 1^e), u_{56}(0.121, 0.55, -1.02), u_{55}(1^d, -3.7, 2.4), \) and \( u_{62}(0.93, -1.62, -3.2), R = 0.944, S = 0.0028, IC = 67.0\% \), there is observed no degeneracy. In the best observed combinations of five matrix elements, however, in \( u_{53}(1^e, -1.97, 1.25), u_{56}(0.043, 2.7, 1.01), u_{66}(-0.31, -\infty, -0.47), u_{63}(-3.9, -2.6, -1.89), \) and \( u_{66}(0.28, -2.4, 1^e), R = 0.964, S = 0.0023, IC = 73.2\% \), as well as in the combination of \( u_{53}(-4.9, -3.0, 1^e), u_{56}(0.113, -0.020, -1.14), u_{66}(1^d, -2.4, 2.4), u_{66}(0.23, -1.73, -3.5), \) and \( u_{66}(-0.26, 0.124, 1^e), R = 0.960, S = 0.0024, IC = 72.1\% \), there is observed no degeneracy.

These data demonstrate that the degeneracy of topological indices is an important criterion of their goodness but not always decisive.

4.2. Meaning of Exponent Values in Elements of the Universal Matrix

When exponent values for a, b and c in the equation \( u(a, b, c) = v_i^a \times v_j^b \times d^c \) are equal to 1 (one) it means that the values of vertex degrees resp. vertex distances contribute proportionally to their values. An exponent value of >1 means that the contribution of higher vertex degrees resp. vertex distances is exaggerated. An exponent value between 1 and 0 means that the contribution of vertex degrees resp. vertex distances is diminished, i.e. the contribution of higher vertex degrees resp. vertex distances is less than their original value would indicate. An exponent value of 0 (zero) means that different values of vertex degrees resp. vertex distances contribute equally. An exponent value of <0 means that the higher values of vertex degrees resp. vertex distances contribute less than the lower ones. An exponent value of \( \infty \) means that vertex degrees resp. vertex distances higher than 1 do not contribute anything.

4.3. Structural Interpretation of Some of the Physicochemical Properties of Octanes Based on Elements of the Universal Matrix

Next question is, whether the elements of the Universal matrix, which represent particular structural features, in our case of octanes, enable the structural interpretation of their physicochemical properties. Structural interpretation of Octane Number, which is a PCP governed by a series of chemical reactions, has already been performed, cf. e.g. [21,22]. Structural interpretation of the elements of the Universal matrix, which give rise to the best observed correlation with MON data is presented in Appendix 4.
The Boiling point (BP) is governed by the intermolecular attractions and collisions as well. It decreases with increasing branching that gives at octanes the sequence of BP: Oct > 3M7 > 3Et6 > 3Et3M5 > 34M6 > 4M7 > 2M7 > 3E2M5 > 23M6 > 233M5 > 24M5 > 223M5 > 25M6 > 225M6 > 2233M4 > 224M5. The sequence of values of nD is as follows: 3Et3M5 > 233M5 > 223M5. Structural interpretation of the elements of the Universal matrix, which give rise to the best observed correlation with Tc^2/Pc data is presented in Appendix 5.

The sequence of values of Critical temperature, Tc, is 3Et3M5 > 233M5 > 223M5 > 24M6 > 224M5 > 2233M4. Above the general trend are positioned Oct and 233M5, below it 24M6, 224M5, and 2233M4. Structural interpretation of the elements of the Universal matrix, which give rise to the best observed correlation with Tc^2/Pc data is presented in Appendix 5.

The Refractive index n_r is a volumetric PCP. The sequence of values of n_r is as follows: 3Et3M5 > 233M5 > 24M6 > 3Et6 > 3Et3M5 > 233M5 > 34M6 > 4M7 > 2M7 > 3E2M5 > 23M6 > 233M5 > 24M5 > 223M5 > 25M6 > 225M6 > 2233M4 > 224M5. The sequence of values of Critical density, dc, data is presented in Appendix 6. The sequence of values of dc is 223M5 > 3Et2M5 > 33M6 > 3Et6 > 3Et3M5 > 233M5 > 24M5 > 2233M4 > 34M6 > 23M6 > 33M6 > 223M5 > 24M6 > 225M6 > 2233M4 > 224M5. Above the general trend are positioned Oct and 233M5, below it 24M6, 224M5, and 2233M4. Structural interpretation of the elements of the Universal matrix, which give rise to the best observed correlation with Tc data is presented in Appendix 8.

Structural interpretations of MON, Tc^2/Pc, BP, n_d, Tc, dc, and logVP are presented, as well as interpretations of what contribute to it particular matrix elements, which are members of the best combined topological indices that are mutually optimized combinations of six matrix elements.

5. Conclusions

Particular elements of the Universal matrix and especially the mutually optimized combinations of few (four to six out of 56) of them can be used as good topological indices, correlating to tested physicochemical properties to R > 0.985 even in the worst tested cases.

Besides R and S, an additional quantity useful to illustrate the potential goodness of topological indices is proposed, the information content (IC). IC is linearly and negatively correlated to S. It is an easily comprehensible direct indication of goodness of the topological indices (index combination) and is not dependent on numerical values of PCP in question, so it can be used also for inter-PCP comparisons of goodness of topological indices.

Structural interpretations of MON, Tc^2/Pc, BP, n_d, Tc, dc, and logVP are presented, as well as interpretations of what contribute to it particular matrix elements, which are members of the best combined topological indices that are mutually optimized combinations of six matrix elements.

6. References

1. L. Pogliani, Indian J. Chem. 2003, 42A, 1347–1353.
2. Z. Mihalić, N. Trinajstić, J. Chem. Educ. 1992, 69, 701–712. http://dx.doi.org/10.1021/ed069p701
3. A. T. Balaban (Ed.), From chemical topology to three-dimensional geometry. Plenum Press, New York and London, 1997.
Povzetek

Kot topološki indeksi za fizikokemijske lastnosti oktanov so uporabni tudi elementi Univerzalne matrike in kombinacije po nekaj od njih. Medtem ko nekateri posamezni elementi Univerzalne matrike dajo $0.70 < |R| < 0.99$, pa medsebojno optimirane kombinacije po 4 do 6 od 56 elementov Univerzalne matrike oktanov dajo $R > 0.99$ in v najslabših primerih $R > 0.98$. Uvedeno je tudi novo merilo za oceno, kako dobra je korelacija, to je vsebnost informacije v topološkem indeksu, $IC (%)$. Narejena je tudi strukturna interpretacija nekaterih fizikalno-kemijskih lastnosti oktanov ter doprinosu posameznih elementov matrike.