MATCH
Communications in Mathematical and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. **75** (2016) 29-56

ISSN 0340 - 6253

Simple Alcohols with the Lowest Normal Boiling Point Using Topological Indices*

Mikhail Goubko, Oleg Miloserdov

Institute of Control Sciences

of Russian Academy of Sciences, Moscow, Russia

mgoubko@mail.ru, omilos92@gmail.com

(Received October 28, 2014)

Abstract

We find simple saturated alcohols with the given number of carbon atoms and the minimal normal boiling point. The boiling point is predicted with a weighted sum of the generalized first Zagreb index, the second Zagreb index, the Wiener index for vertex-weighted graphs, and a simple index caring for the degree of a carbon atom being incident to the hydroxyl group. To find extremal alcohol molecules we characterize chemical trees of order n, which minimize the sum of the second Zagreb index and the generalized first Zagreb index, and also build chemical trees, which minimize the Wiener index over all chemical trees with given vertex weights.

1 Introduction

Consider a collection Ω of *admissible molecules* (for example, represented with their structural formulas or chemical graphs), each endowed with $k+1$ significant physical or chemical *properties* (e.g., normal density, normal boiling point, refraction coefficient, retention index, or more exotic and problem-specific ones), and let $P_i(G)$, $i = 0, ..., k$, be the numeric value of the i-th property of a molecule $G \in \Omega$ (e.g., the normal boiling point

*This research is supported by the grant of Russian Foundation for Basic Research, project No 13-07-00389.

value). A typical problem of *molecular design* is the following *optimization problem*:

$$P_0(G) \to \min_{G\in\Omega} (\max_{G\in\Omega}) \qquad (1)$$
$$P_i^{min} \leqslant P_i(G) \leqslant P_i^{max}, i = 1, ..., k.$$

When the functions $P_i(\cdot)$ are only partially known from the experiment, they are replaced with *predicted* figures, relating a chemical graph $G \in \Omega$ to the predicted value $\tilde{P}_i(G)$ of the i-th physical or chemical property ($i = 0, ..., k$) by virtue of numeric characteristics (known as *molecular descriptors*), which can be calculated on basis of a molecular structure. A typical *quantitative structure-property relation* (QSPR) includes several molecular descriptors, and is presented as

$$\tilde{P}_i(G) = \tilde{P}_i(I_1(G), ..., I_m(G)), i = 0, ..., k,$$

where $I_1(G), ..., I_m(G)$ are the values of molecular descriptors for the molecular graph G. The simplest *linear regression* is just a weighted sum of descriptors:

$$\tilde{P}_i(G) = \alpha_{1,i} I_1(G) + ... + \alpha_{m,i} I_m(G), i = 0, ..., k.$$

During recent decades a number of *topological*, *geometrical*, and *quantum-mechanical* molecular descriptors were suggested and studied [13,14,24,25]. Below we limit ourselves to topological descriptors only (see, for instance, the handbook [1]) to study problem (1) as a problem of the *extremal graph theory* [4] .

Exhaustive enumeration of all feasible molecules (the *brute force* approach) can only be used to solve this problem when the feasible set is relatively small; for bigger sets mathematical chemistry suggests a variety of limited search techniques. In numerous papers lower and upper bounds of dozens topological indices over various feasible sets were obtained [5–7,18,19,23,26,28–30], and, in many cases, extremal graphs were characterized. At the same time, the problem (1) of optimization of a composition of indices is still understudied.

In fact, finding lower and upper bounds of individual indices can be a step towards solving problem (1), as a linear combination of lower bounds is a lower-bound estimate of the combination of indices. This estimate can be used in a branch-and-bound algorithm of limited search. Yet, the quality of the estimate may be considerably poor, resulting in lack of efficient cuts in a branch-and-bound algorithm.

Anyway, the common shortcoming of an algorithmic approach to index optimization is that it does not support the analysis of general characteristics of an extremal molecule (i.e., of a corresponding graph). When available, side information would be of great value on why a certain graph is optimal or not, what shape the extremal graphs have, etc. Such information is revealed using analytical tools of discrete optimization.

In this paper we apply recent results in optimization of degree- and distance-based topological indices to find a simple saturated alcohol with the given molecular weight and minimal boiling point. We reduce the property minimization problem to that of minimization of a weighted linear combination of the generalized first Zagreb index, the second Zagreb index, the vertex-weighted Wiener index, and a simple index caring for the degree of a carbon atom being incident to the hydroxyl group. Then we characterize minimizers of this linear combination of indices (see Fig. 5) and of a couple of simpler regressions (see Fig. 2 and 4).

2 Predicting Boiling Points of Simple Alcohols

The normal boiling point of a liquid is determined by its solvation free energy. The solvation free energy can be predicted with high accuracy from computer simulations (see [2,3] for details). The simulation-based approach solves well the "direct problem" of predicting the solvation free energy for a given molecule, but it does not help solving the "inverse problem" of finding the molecule having the minimal solvation free energy (and, consequently, the boiling point). For this reason we predict boiling points of simple saturated alcohols (those having a general formula $C_nH_{2n+1}OH$) with the aid of topological indices.

Alcohols have relatively high boiling points when compared to related compounds due to hydrogen bonds involving a highly polarized hydroxyl group, and branched isomers have lower boiling points than alcohols with the linear structure. Another structural feature affecting the boiling point is the *"oxygen shielding" effect* [21], when atoms surrounding the hydroxyl group partially shield it preventing formation of hydrogen bonds between molecules and, thus, decreasing the boiling point.

We considered several degree-based topological indices (the first Zagreb index M_1 [12], the second Zagreb index M_2 [12], Randić index [22] and the others), which are known to be good metrics of branchiness, and, finally, the generalized first Zagreb index C_1 (see

also [9]) has shown the best results:

$$C_1(G) = \sum_{v \in V(G)} c(d_G(v)), \tag{2}$$

where $V(G)$ is the vertex set of graph G, $d_G(v)$ is the degree of the vertex $v \in V(G)$ in graph G, and $c(d)$ is a non-negative function defined for degrees from 1 to 4. It can alternatively be written as

$$C_1(G) = c(1)n_1(G) + c(2)n_2(G) + c(3)n_3(G) + c(4)n_4(G), \tag{3}$$

where $n_i(G)$ is the atoms' count of degree $i = 1, ..., 4$ in a molecular graph G, and $c(1), ..., c(4)$ are regression parameters.

We also employed the classical *second Zagreb index* [12]

$$M_2(G) = \sum_{uv \in E(G)} d_G(u) d_G(v), \tag{4}$$

where $E(G)$ is the edge set of graph G.

Another index used was the *Wiener index*, which had been the first topological index for boiling point prediction [27] due to its high correlation with the molecule's surface area. To account for heterogeneity of atoms we allow each pair of vertices $u, v \in V(G)$ to have unique weight $\mu_G(u, v)$ and calculate the *pair-weighted Wiener index* as

$$PWWI(G) := \frac{1}{2} \sum_{u,v \in V(G)} \mu_G(u, v) d_G(u, v), \tag{5}$$

where $d_G(u, v)$ is the distance (the length of the shortest path) between vertices u and v in G. For example, we can assign different weights to distances between pairs of carbon atoms and between carbons and the oxygen atom in an alcohol molecule.

Regression tuning has shown the distances between carbon atoms to be irrelevant for the alcohol boiling point, and only distances to the oxygen matter. All such distances are accounted with equal weight, so the pair-weighted Wiener index reduces to the *distance* of the oxygen atom, which was first used for the alcohol boiling point prediction in [21]:

$$WI_{\mathrm{O}}(G) := \sum_{u \in V(G)} d_G(u, \mathrm{O}). \tag{6}$$

In [21] a geometrical descriptor has also been suggested to account for oxygen shielding, but we extend the approach by [15] instead, and introduce a simple topological index

Table 1: Data sample: basic statistics

# carbons	# isomers	min chain len.	max chain len.	min b. p.	max b. p.
2	1	2	2	78	78
3	2	2	3	82.5	97
4	4	2	4	82.4	117.5
5	8	3	5	102	137
6	17	3	6	120	157
7	18	4	7	131.5	175.5
8	10	4	8	147.5	194
9	11	4	9	169.5	215
10	5	5	10	168	231
11	2	10	11	228.5	243
12	1	12	12	259	259
TOTAL	79	2	12	78	259

$S_i(G)$, which is equal to unity when the carbon atom incident to the hydroxyl group in the alcohol molecule G has degree $i = 2, 3, 4$ (we exclude *methanol* from consideration), and is equal to zero otherwise.

We collected a data set of experimental boiling points under normal conditions for 79 simple saturated alcohols having from 2 to 12 carbon atoms and representing various branchiness. Several data sources [8, 15, 17, 21] were combined with priority on Alpha Aesar experimental data to resolve discrepancy. In Table 1 we present basic statistics about the data sample.

Information on boiling points of alcohols including more than 12 carbon atoms is less common and reliable. The complete data set together with the best regressions is available online at [11].

We randomly split the sample into the training set containing 50 cases and the testing set containing 29 cases. Then we examined different linear regressions involving the descriptors mentioned above.[1] The best performance and predictive power was obtained for the linear combination of the oxygen's distance cube root $WI_{\mathrm{O}}(G)^{\frac{1}{3}}$ (with weight b_1^0), the generalized first Zagreb index $C_1(G)$ (with weights $c^0(1), ..., c^0(4)$), the second Zagreb index $M_2(G)$ (weighted by b_3^0), and the simple indicator of the sub-root's degree $S_2(G)$

[1] ChemAxon Instant JChem© was used for index calculation. Authors would like to thank ChemAxon® Ltd (http://www.chemaxon.com) for the academic license.

Table 2: Precision of regressions: Correlation coefficient and standard deviation (SD)

Data set	Basic regression ("0")		Regression I ("I")		Regression II ("II")	
	Corr.	SD, °C	Corr.	SD, °C	Corr.	SD, °C
Training set	0.997	2.98	0.997	3.12	0.996	3.26
Testing set	0.996	3.23	0.995	3.58	0.994	3.99

Table 3: Parameters of regressions

Coefficient	Basic regression ("0")	Regression I ("I")	Regression II ("II")
b_0	35.245	50.626	44.134
b_1	12.233	–	3.851
b_2	9.170	11.295	10.980
b_3	1.486	1.000	–
$c(1)$	–	–	–
$c(2)$	9.514	14.534	17.727
$c(3)$	9.380	20.172	29.673
$c(4)$	–	17.015	36.470

(weighted by b_2^0):

$$BP^0 = b_0^0 + b_1^0 WI_{\rm O}(G)^{\frac{1}{3}} + c^0(2)n_2(G) + c^0(3)n_3(G) + b_2^0 S_2(G) + b_3^0 M_2(G). \qquad (7)$$

Below this regression is referred to as the *basic* one. The optimal values of weights $b_1^0, c^0(1), ..., c^0(4), b_2^0, b_3^0$ (including the constant term b_0^0) calculated with the least squares method are presented in the first column of Table 3. Parameters $c(1)$ and $c(4)$, which weight variables $n_1(G)$ and $n_4(G)$ respectively, appear to be insignificant and can be set to zero in (3) when calculating the generalized first Zagreb index. The precision of the basic regression is shown in Table 2. It is comparable to the best known relations [8, 15, 17, 21]. In the following sections we show how far we can come in analytical and numeric minimization of this combination of indices.

We also considered two simplifications of regression (7), for which alcohol molecules having minimal predicted boiling point can be characterized analytically. The first one (below referred to as "Regression I") is obtained by withdrawing $WI_{\rm O}(G)$ in (7):

$$BP^{\rm I} = b_0^{\rm I} + c^{\rm I}(2)n_2(G) + c^{\rm I}(3)n_3(G) + c^{\rm I}(4)n_4(G) + b_2^{\rm I} S_2(G) + b_3^{\rm I} M_2(G). \qquad (8)$$

See Table 2 for the precision figures of regression (8) under the values of parameters delivering the best approximation to the training set (see the middle column of Table 3).

The second simplified regression (referred to as "Regression II" below) is obtained by withdrawing $M_2(G)$ in (7):

$$BP^{\mathrm{II}} = b_0^{\mathrm{II}} + b_1^{\mathrm{II}} WI_{\mathrm{O}}(G)^{\frac{1}{3}} + c^{\mathrm{II}}(2)n_2(G) + c^{\mathrm{II}}(3)n_3(G) + c^{\mathrm{II}}(4)n_4(G) + b_2^{\mathrm{II}} S_2(G). \quad (9)$$

In Table 2 we show its precision under the optimal values of parameters depicted in the last column of Table 3.

The shortcoming of Regression II is that the term $WI_{\mathrm{O}}(G)^{\frac{1}{3}}$ appears to be insignificant after disposal of $M_2(G)$, being responsible of approximately 1 per cent of the residual sum of squares. Nevertheless, we keep this regression for illustration of joint optimization of $C_1(G)$ and of the pair-weighted Wiener index.

3 Minimization of indices and their combinations

In the present paper we find a simple saturated alcohol isomer with $n-1$ carbon atoms having the lowest predicted boiling point. As the regressions introduced in the previous section are tested only for alcohols containing from 2 to 12 carbon atoms, we restrict our attention to $n \leqslant 14$, where we can expect some accuracy of the obtained results.

For $n \leqslant 14$ admissible sets of all simple saturated alcohol molecules with $n-1$ carbons are not too extensive, and allow for the brute-force enumeration. Moreover, we are sure that no aid of a computer is needed for an organic chemist to draw a molecule being a good approximation to the boiling point minimizer for all $n \leqslant 14$. However, our aim is to show how analytic optimization techniques formalize the professional intuition and help making general conclusions of verifiable reliability.

Let us characterize chemical trees minimizing indices introduced in the previous section and their combinations.

3.1 Degree-based indices

For a simple connected undirected graph G denote with $W(G)$ the set of *pendent vertices* (those having degree 1) of the graph G, and with $M(G) := V(G) \backslash W(G)$ the set of *internal vertices* (with degree > 1) of G.

Definition 1 A simple connected undirected graph of order n is called a *chemical tree* if it has $n-1$ edges and its vertex degrees do not exceed 4. Denote with $\mathcal{T}(n)$ the set of all chemical trees of order n. □

Definition 2 A *pendent-rooted chemical tree* is a chemical tree, in which one pendent vertex is distinguished and called a *root*. A typical pendent-rooted tree is denoted with T_r, with r being its root. A vertex being incident to the root in T_r is called a *sub-root* and is denoted as $sub(T_r)$. Denote with $\mathcal{R}(n)$ the set of all pendent-rooted chemical trees of order n. □

Clearly, the set $\Omega(n-1)$ of all molecules of simple saturated alcohols having $n-1$ carbons coincides with the set $\mathcal{R}(n)$ of pendent-rooted chemical trees of order n (with a root corresponding to the hydroxyl group and the other vertices forming the carbon skeleton of a molecule).

For a topological index $I(\cdot)$ defined on an admissible set of graphs $\mathcal{G}$ introduce the notation $I^*_{\mathcal{G}} := \min_{G\in\mathcal{G}} I(G)$ and let $\mathcal{G}^*_I := \mathrm{Arg}\min_{G\in\mathcal{G}} I(G)$ be the set of graphs minimizing $I(\cdot)$ over $\mathcal{G}$. For example, $\mathcal{T}^*_{M_2}(n)$ is the set of chemical trees of order n minimizing the second Zagreb index $M_2(\cdot)$.

Define also the set $\mathcal{R}_i(n) := \{T \in \mathcal{R}(n) : d_T(sub(T)) = i\}$ of all pendent-rooted trees with a sub-root having degree $i = 2, ..., 4$.

We start with the following obvious statement.

Lemma 1 *$S_i(G)$ achieves its minimum at any pendent-rooted chemical tree with sub-root's degree other than i. In other words, $\mathcal{R}^*_{S_i}(n) = \mathcal{R}(n)\backslash\mathcal{R}_i(n)$.*

Proof is straightforward, as $S_i(G) = 1$ for all $G \in \mathcal{R}_i(n)$, and $S_i(G) = 0$ otherwise. ■

Indices $C_1(G)$ and $M_2(G)$ do not account for heterogeneity of atoms in a molecule, so we can minimize them over the set $\mathcal{T}(n)$ of all chemical trees of order n and then assign the root to an arbitrary pendent vertex of the index-minimizing tree to obtain a pendent-rooted tree, which minimizes the index.

Consider an "ad-hoc" degree-based topological index

$$C(G) := C_1(G) + b_3 M_2(G) = \sum_{v\in V(G)} c(d_G(v)) + b_3 \sum_{uv\in E(G)} d_G(u) d_G(v), \qquad (10)$$

where b_3 is an arbitrary real constant (we keep notation b_3 for compatibility with equations (7), (8)).

Definition 3 A chemical tree $T \in \mathcal{T}(n)$ is *extremely branched*, if its internal vertices have degree 4, except one vertex having degree 2 when $n \bmod 3 = 0$, or one vertex having degree 3 when $n \bmod 3 = 1$. □

Theorem 1 *Assume the following inequalities hold:*

$$c(1) + c(4) + 18b_3 < c(2) + c(3), \tag{11}$$

$$c(1) + c(3) + 8b_3 < 2c(2), \tag{12}$$

$$c(2) + c(4) + 8b_3 < 2c(3). \tag{13}$$

If a chemical tree $T \in \mathcal{T}(n)$ for $n \geqslant 3$ minimizes $C(\cdot)$ over all chemical trees from $\mathcal{T}(n)$, then T is an extremely branched tree. For $n \leqslant 17$ the inequality (11) *can be weakened to*

$$c(1) + c(4) + 17b_3 < c(2) + c(3). \tag{14}$$

PROOF We employ the standard argument of index monotonicity with respect to certain tree transformations. Assume the theorem does not hold, and vertices $u, v \in M(T)$ exist such that $u \neq v$ and $d_T(u), d_T(v) < 4$. Four cases are possible.

1. $d_T(u) = d_T(v) = 2$. Let $v_1, v_2 \in V(T)$ be the vertices incident to v in T. Also, let $u_1, u_2 \in V(T)$ be the vertices incident to u in T, and u_2 lies on the path to the vertex v in T. Without loss of generality assume that

$$d_T(u_1) + d_T(u_2) \geqslant d_T(v_1) + d_T(v_2). \tag{15}$$

   Consider a graph $T' \in \mathcal{T}(n)$ obtained from T by replacing the edge u_1u with the edge u_1v. It is easy to see that T' is a tree. The degree of the vertex u in T' is decreased by one, the degree of vertex v is increased by one, therefore, if $u_2 \neq v$, we have

$$C(T') - C(T) = c(1) + c(3) - 2c(2) + b_3(d_T(u_1) + d_T(v_1) + d_T(v_2) - d_T(u_2)).$$

   From (15) we obtain $C(T') - C(T) \leqslant c(1) + c(3) - 2c(2) + 2b_3 d_T(u_1)$. Since vertex degrees $\leqslant 4$ in a chemical tree, from (12) we have $C(T') - C(T) \leqslant c(1) + c(3) - 2c(2) + 8b_3 < 0$, which contradicts the assumption that T minimizes $C(\cdot)$.

   If $u_2 = v$, in the same manner obtain $C(T') - C(T) \leqslant c(1) + c(3) - 2c(2) + 7b_3$. From (12), it is also negative.

2. $d_T(u) = 2$, $d_T(v) = 3$. Let $v_1, v_2, v_3 \in V(T)$ be the vertices incident to v in T. Also, let $u_1, u_2 \in V(T)$ be the vertices incident to u in T, and assume u_2 lies on the path to the vertex v in the tree T. Consider a tree $T' \in \mathcal{T}(n)$ obtained from T by

replacing the edge u_1u with the edge u_1v. By analogy to the previous case, if $u_2 \neq v$, we obtain $C(T') - C(T) = c(1) + c(4) - c(2) - c(3) + b_3(2d_T(u_1) + d_T(v_1) + d_T(v_2) + d_T(v_3) - d_T(u_2))$. Vertex degrees $\leqslant 4$ in a chemical tree. Moreover, $d_T(u_2) \geqslant 2$, since it is an intermediate vertex on the path $u, u_2, ..., v$. Therefore,

$$C(T') - C(T) \leqslant c(1) + c(4) - c(2) - c(3) + 18b_3,$$

and, from (11), $C(T') - C(T) < 0$, which is a contradiction.

To prove the weaker inequality (14) we are enough to prove that $C(T') - C(T) < 0$ for $n \leqslant 17$, since $d_T(u_1) = d_T(v_1) = d_T(v_2) = d_T(v_3) = 4$, and $d_T(u_2) = 2$ is possible only in a tree of order 18 or more (an example is depicted in Fig. 1a), and $C(T') - C(T) = c(1) + c(4) - c(2) - c(3) + 17b_3$ for the tree T depicted in Fig. 1b. If

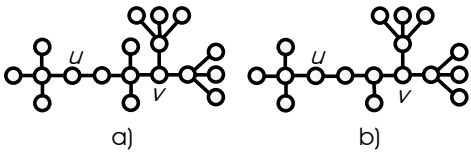

Figure 1: To the proof of inequality (14)

$u_2 = v$, without loss of generality assume that $u = v_3$. Then $C(T') - C(T) = c(1) + c(4) - c(2) - c(3) + b_3(2d_T(u_1) + d_T(v_1) + d_T(v_2) - 2) \leqslant c(1) + c(4) - c(2) - c(3) + 14b_3$. From (11) (or from (14), if $n \leqslant 17$), it is negative.

3. The case of $d_T(v) = 2$, $d_T(u) = 3$ is considered in the same manner.

4. $d_T(u) = d_T(v) = 3$. Let $v_1, v_2, v_3 \in V(T)$ be the vertices incident to v in T. Also, let $u_1, u_2, u_3 \in V(T)$ be the vertices incident to u in T, with u_1 not laying on a path to v in T. Without loss of generality assume that

$$d_T(u_1) + d_T(u_2) + d_T(u_3) \geqslant d_T(v_1) + d_T(v_2) + d_T(v_3). \tag{16}$$

Consider a tree $T' \in \mathcal{T}(n)$ obtained from T by replacing the edge u_1u with the edge u_1v. If $uv \notin E(T)$, then from (16) and $d_T(u_1) \leqslant 4$ we have

$$C(T') - C(T) = c(2) + c(4) - 2c(3) + b_3(d_T(u_1) + d_T(v_1) + d_T(v_2) + d_T(v_3) - d_T(u_2) -$$

$$-d_T(u_3)) \leqslant c(2) + c(4) - 2c(3) + 2b_3 d_T(u_1) \leqslant c(2) + c(4) - 2c(3) + 8b_3,$$

which is less than zero due to (13), and T cannot minimize $C(\cdot)$. If $uv \in E(T)$, in the same way deduce $C(T') - C(T) \leqslant c(2) + c(4) - 2c(3) + 7b_3$, which is negative.

The obtained contradictions prove that no more than one internal vertex in T may have degree less than 4.

As $T \in \mathcal{T}(n)$ and $n > 1$, the well-known equity holds:

$$n_1(T) + 2n_2(T) + 3n_3(T) + 4n_4(T) = 2(n-1). \tag{17}$$

On the other hand,

$$n_1(T) + n_2(T) + n_3(T) + n_4(T) = n. \tag{18}$$

Assume that $n_2(T) = 1$, so that $n_3(T) = 0$. From (18) we have $n_1(T) + n_4(T) = n - 1$, therefore, (17) makes $n = 3 + 3n_4(T)$ and, since $n_4(T) \in \mathbb{N}_0$, $n \bmod 3 = 0$.

In the same manner we show that if $n_3(T) = 1$ then $n \bmod 3 = 1$. If both $n_2(T)$ and $n_3(T) = 0$, then $n \bmod 3 = 2$, and the proof is complete. ∎

Corollary 1 *Under conditions of Theorem* 1*, any tree* T *minimizing* $C(T) = C_1(T) + b_3 M_2(T)$ *over* $\mathcal{T}(n)$ *enjoys the same number* n_i *of vertices of degree* $i = 1, ..., 4$*. Therefore,* $C_1(T) = C_1(T')$ *for any pair of trees* $T, T' \in \mathcal{T}_C^*(n)$*.* □

Corollary 2 *Under conditions of Theorem* 1 *the sets* $\mathcal{T}_C^*(n)$ *for* $n = 4, ..., 14$ *are depicted in Fig.* 2*.* $\mathcal{T}_C^*(n)$ *contains the sole tree for* $n < 14$ *, while* $\mathcal{T}_C^*(14)$ *contains two trees.*

Proof From Corollary 1 we learn that only the value of $M_2(\cdot)$ may vary within $\mathcal{T}_C^*(n)$.

From Theorem 1, for $n \in \{5, 8, 11, 14\}$ an optimal tree is a 4-tree (in which all internal vertices have degree 4). Each of n_1 *stem* edges (those incident to a pendent vertex) adds 4 to the value of M_2, while each of $n_4 - 1$ edges connecting internal vertices adds 16 to the value of M_2. Since n_1 and n_4 are fixed for fixed n, all 4-trees have the same value of $M_2(\cdot)$ (and, therefore, the same value of $C(\cdot)$). Consequently, for for $n = 5, 8, 11, 14$ the set $\mathcal{T}_C^*(n)$ consists of all 4-trees of order n (see Fig. 2).

If $T \in \mathcal{T}_C^*(n)$, and $n \in \{6, 9, 12\}$, one internal vertex $u \in M(T)$ has degree $d_T(u) = 2$, while all others have degree 4. For $n = 6$ only one such tree exists depicted in Fig. 2. It is easy to check that $M_2(\cdot)$ is minimized if vertex u is incident to two internal vertices. Only one such tree exists for $n = 9$ (see Fig. 2), and the same is true for $n = 12$.

For $n \in \{4, 7, 10, 13\}$ any tree $T \in \mathcal{T}_C^*(n)$ has one internal vertex $u \in M(T)$ of degree $d_T(u) = 3$, while all other have degree 4. For $n = 4$ only one such tree exists depicted in Fig. 2, and the same is true for $n = 7$. Again, it is easy to check that, in the context of $M_2(\cdot)$ minimization, vertex u being incident to three internal vertices is strictly preferable to vertex u being incident to one pendent and two internal vertices, which is, in turn, preferred to u having two incident pendent vertices. So, optimal trees for $n = 10, 13$ are depicted in Fig. 2 (black and white filling of circles is explained below). ∎

The same logic allows continuing the sequence of $C(\cdot)$-minimizers to $n > 14$.

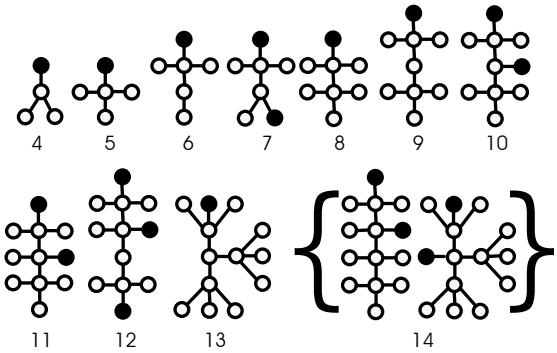

Figure 2: Chemical trees minimizing the "ad-hoc" index $C(\cdot)$ for $n = 4, ..., 14$ (alcohol molecules minimizing $BP^{\mathrm{I}}(\cdot)$ with possible oxygen positions filled with black)

Therefore, Theorem 1 says that, when conditions (11)-(13) hold, chemical trees minimizing $C(\cdot)$ have as many vertices of maximal degree 4 as possible. A similar result can be proved for the modified Wiener index $WI_{\mathrm{O}}(\cdot)$.

3.2 Wiener index

A simple connected undirected graph G is called *vertex-weighted* if each vertex $v \in V(G)$ is endowed with a non-negaive weight $\mu_G(v)$. With μ_G we denote the total vertex weight of the graph G, and $\mathcal{WT}(n)$ stands for the set of all vertex-weighted trees of order n.

Klavžar and Gutman [16] defined the *Wiener index for vertex-weighted graphs* as

$$VWWI(G) := \frac{1}{2} \sum_{u,v \in V(G)} \mu_G(u)\mu_G(v)d_G(u, v).$$

Clearly, $VWWI(\cdot)$ is a special case of the pair-weighted Wiener index $PWWI(\cdot)$ (defined with formula (5)) for $\mu_G(u,v) := \mu_G(u)\mu_G(v)$. The path-weighted Wiener index is poorly studied at the moment, but, fortunately, $WI_{\rm O}(\cdot)$, which is the point of our current interest, can be reduced to the Wiener index for vertex-weighted graphs.

For every alcohol molecule from $\Omega(n-1)$ (or, equivalently, for every pendent-rooted tree $T_r \in \mathcal{R}(n)$) define a vertex-weighted tree $T(\varepsilon) \in \mathcal{WT}(n)$ by assigning the weight $\mu_{T(\varepsilon)}(v) := \varepsilon$ to each vertex $v \in V(T_r)$ (a carbon atom) except the root r, and assigning the weight $\mu_{T(\varepsilon)}(r) := 1/\varepsilon$ to the root (the oxygen atom). It is easy to see that under these weights $\lim_{\varepsilon\to 0} VWWI(T(\varepsilon)) = WI_{\rm O}(T_r)$. Since $WI_{\rm O}(\cdot)$ is integer-valued, minimizers of $VWWI(\cdot)$ and of $WI_{\rm O}(\cdot)$ coincide for sufficiently small ε.

In [10] the majorization technique suggested by Zhang et al. [29] is used to minimize $VWWI(\cdot)$ over the set of trees with given vertex weights and degrees. Below we recall the notation and selected theorems from [10]. We use them to find the extremal vertex degrees over the set of all trees of order n with fixed vertex weights.

Definition 4 Consider a vertex set V. Let the function $\mu : V \to \mathbb{R}_+$ assign a non-negative weight $\mu(v)$ to each vertex $v \in V$, while the function $d : V \to \mathbb{N}$ assign a natural degree $d(v)$. The tuple $\langle \mu, d\rangle$ is called *a generating tuple* if the following identity holds:

$$\sum_{v\in V} d(v) = 2(|V|-1). \tag{19}$$

Denote with $\overline{\mu} := \sum_{v\in V}\mu(v)$ the total weight of the vertex set V. Let $\mathcal{WT}(\mu) := \{T \in \mathcal{WT}(|V|) : V(T) = V, \mu_T(v) = \mu(v) \text{ for all } v \in V\}$ be the set of trees over the vertex set V with vertex weights $\mu(\cdot)$. For the set $\mathcal{WT}(\mu,d) := \{T \in \mathcal{WT}(\mu) : d_T(v) = d(v) \text{ for all } v \in V\}$ we also require vertices to have degrees $d(\cdot)$. □

Let $V(\mu,d)$ be the domain of functions of a generating tuple $\langle \mu, d\rangle$. Introduce the set $W(\mu,d) := \{w \in V(\mu,d) : d(w) = 1\}$ of *pendent* vertices and the set $M(\mu,d) := V(\mu,d)\backslash W(\mu,d)$ of *internal* vertices.

Definition 5 We will say that in a generating tuple $\langle \mu, d\rangle$ *weights are degree-monotone*, if for any $m, m' \in M(\mu,d)$ from $d(m) < d(m')$ it follows that $\mu(m) \leqslant \mu(m')$, and for any $w \in W(\mu,d)$ we have $\mu(w) > 0$. □

For a generating tuple $\langle \mu, d\rangle$ the *generalized Huffman algorithm* [10] builds a tree $H \in \mathcal{WT}(\mu,d)$ as follows.

Setup. Define the vertex set $V_1 := V(\mu, d)$ and the functions μ^1 and d^1, which endow its vertices with weights $\mu^1(v) := \mu(v)$ and degrees $d^1(v) := d(v)$, $v \in V_1$. We start with the empty graph H over the vertex set $V(\mu, d)$.

Steps $i = 1, ..., q-1$**.** Denote with m_i the vertex having the least degree among the vertices of the least weight in $M(\mu^i, d^i)$. Let $w_1, ..., w_{d(m_i)-1}$ be the vertices having $d(m_i) - 1$ least weights in $W(\mu^i, d^i)$. Add to H edges $w_1 m_i, ..., w_{d(m_i)-1} m_i$.

Define the set $V_{i+1} := V_i \backslash \{w_1, ..., w_{d(m_i)-1}\}$ and functions $\mu^{i+1}(\cdot), d^{i+1}(\cdot)$, endowing its elements with weights and degrees as follows:

$$\mu^{i+1}(v) := \mu^i(v) \text{ for } v \neq m_i, \qquad \mu^{i+1}(m_i) := \mu^i(m_i) + \mu^i(w_1) + ... + \mu^i(w_{d(m_i)-1}),$$
$$d^{i+1}(v) := d^i(v) \text{ for } v \neq m_i, \qquad d^{i+1}(m_i) := 1. \tag{20}$$

Step q**.** Consider a vertex $m_q \in M(\mu^q, d^q)$. By construction, $|M(\mu^q, d^q)| = 1$, $|W(\mu^q, d^q)| = d(m_q)$. Add to H edges connecting all vertices from $W(\mu^q, d^q)$ to m_q. Finally, set $\mu_H(v) := \mu(v)$, $v \in V(H)$.

Theorem 2 **[10]** *If weights are degree-monotone in a generating tuple $\langle \mu, d \rangle$, then $T \in \mathcal{WT}^*_{VWWI}(\mu, d)$ if and only if $T \in \mathcal{WT}(\mu, d)$ and T is a Huffman tree. In other words, only a Huffman tree minimizes the Wiener index over the set of trees whose vertices have given weights and degrees.* □

In the present subsection we study how the value of $VWWI^*_{\mathcal{WT}}(\mu, d)$ changes with degrees $d(\cdot)$. Our results are analogous to those proved by Zhang et al. [29] for the "classical" Wiener index. Following [10], we reformulate the problem for directed trees.

Definition 6 A (weighted) *directed tree* is a weighted connected directed graph with each vertex except the *terminal vertex*[2] having the sole outbound arc and the terminal vertex having no outbound arcs. □

An arbitrary tree $T \in \mathcal{WT}(n)$ can be transformed into a directed tree by choosing an internal vertex $t \in M(T)$, and replacing all its edges with arcs directed towards (a terminal vertex) t. Let us denote with $\mathcal{WD}$ the collection of all directed trees, which can be obtained in such a way, and let $\mathcal{WD}(\mu, d)$ stand for all directed trees obtained from

[2]Typically it is called a *root*, but we will use an alternative notation to avoid confusion with a root of a pendent-rooted tree introduced in the previous subsection.

$\mathcal{WT}(\mu, d)$. Vice versa, in a directed tree from $\mathcal{WD}(\mu, d)$ replacement of all arcs with edges makes some tree from $\mathcal{WT}(\mu, d)$.

If at Step $i = 1, ..., q$ of the generalized Huffman algorithm we add arcs towards the vertex m_i (instead of undirected edges), we obtain a *directed Huffman tree* with the terminal vertex m_q.

Definition 7 For a vertex $v \in V(T)$ of a directed tree $T \in \mathcal{WD}$ define its *subordinate group* $g_T(v) \subseteq V(T)$ as the set of vertices having the directed path to the vertex v in the tree T (the vertex v itself belongs to $g_T(v)$). The *weight* $f_T(v)$ *of a subordinate group* $g_T(v)$ is defined as the total vertex weight of the group: $f_T(v) := \sum_{u \in g_T(v)} \mu_T(u)$. □

Note 1 *If all pendent vertices in* T *have positive weights, then* $f_T(v) > 0$ *for any* $v \in V(T)$. *In particular, it is true for any* $T \in \mathcal{WD}(\mu, d)$, *if weights in* $\langle \mu, d \rangle$ *are degree-monotone.*

The Wiener index is defined for directed trees by analogy to the case of undirected trees: we simply ignore the arcs' direction when calculating distances. Therefore, a tree and a corresponding directed tree share the same value of the Wiener index.

The value of the Wiener index for a directed tree $T_t \in \mathcal{WD}(\mu)$ with a terminal vertex $t \in M(T)$ can be written [10] as:

$$VWWI(T_t) = \sum_{v \in V(T_t) \setminus \{t\}} f_{T_t}(v)(\bar{\mu} - f_{T_t}(v)) = \sum_{v \in V(T) \setminus \{t\}} \chi(f_{T_t}(v)), \qquad (21)$$

where $\chi(x) := x(\bar{\mu} - x)$, and thus, the problems of Wiener index minimization for vertex-weighted trees and for weighted directed trees are equivalent.

Definition 8 Every directed tree T is associated with the *vector of subordinate groups' weights* $\mathbf{f}(T) := (f_T(v))_{v \in V(T) \setminus \{t\}}$, where t is the terminal vertex of T. From equation (21) we see that the vector $\mathbf{f}(T)$ completely determines the value of $VWWI(T)$. □

Definition 9 **[20, 29]** For the real vector $\mathbf{x} = (x_1, ..., x_p)$, $p \in \mathbb{N}$, denote with $\mathbf{x}_\uparrow = (x_{[1]}, ..., x_{[p]})$ the vector where all components of $\mathbf{x}$ are arranged in ascending order. □

Definition 10 **[20, 29]** A non-negative vector $\mathbf{x} = (x_1, ..., x_p)$, $p \in \mathbb{N}$, *weakly majorizes* a non-negative vector $\mathbf{y} = (y_1, ..., y_p)$ (which is denoted with $\mathbf{x} \succeq \mathbf{y}$) if

$$\sum_{i=1}^{k} x_{[i]} \leqslant \sum_{i=1}^{k} y_{[i]} \text{ for all } k = 1, ..., p.$$

If $\mathbf{x}_\uparrow \neq \mathbf{y}_\uparrow$, then $\mathbf{x}$ is said to *strictly weakly majorize* $\mathbf{y}$ (which is denoted with $\mathbf{x} \succ \mathbf{y}$). □

We will need the following properties of weak majorization.

Lemma 2 **[20, 29]** *Consider a positive number $b > 0$ and two non-negative vectors $\mathbf{x} = (x_1, ..., x_k, y_1, ..., y_l)$ and $\mathbf{y} = (x_1 + b, ..., x_k + b, y_1 - b, ..., y_l - b)$, such that $0 \leqslant k \leqslant l$. If $x_i \geqslant y_i$ for $i = 1, ..., k$, then $\mathbf{x} \prec \mathbf{y}$.* □

Lemma 3 **[20, 29]** *If $\mathbf{x} \preceq \mathbf{y}$ and $\mathbf{x}' \prec \mathbf{y}'$, then $(\mathbf{x}, \mathbf{x}') \prec (\mathbf{y}, \mathbf{y}')$, where $(\mathbf{x}, \mathbf{x}')$ means concatenation of vectors $\mathbf{x}$ and $\mathbf{x}'$.* □

Lemma 4 **[20,29]** *If $\chi(x)$ is a increasing concave function, and $(x_1, ..., x_p) \succeq (y_1, ..., y_p)$, then $\sum_{i=1}^{p} \chi(x_i) \leqslant \sum_{i=1}^{p} \chi(y_i)$, and equality is possible only when $(x_1, ..., x_p)_{\uparrow}$ $= (y_1, ..., y_p)_{\uparrow}$.* □

The following lemma establishes an important property of directed Huffman trees:

Lemma 5 **[10]** *For any directed Huffman tree H*

$$vm, v'm' \in E(H), m \neq m', f_H(v) < f_H(v') \Rightarrow f_H(m) < f_H(m'). \tag{22}$$

□

Lemma 6 *Consider generating tuples $\langle \mu, d \rangle$ and $\langle \mu, d' \rangle$ defined on the same vertex set, and let weights be degree-monotone in $\langle \mu, d \rangle$. Let the values of degree functions $d(\cdot)$ and $d'(\cdot)$ differ only for vertices u and v, such that $d(u) \geqslant d(v)$ and $\mu(u) \geqslant \mu(v)$, while $d'(u) = d(u) + 1$, $d'(v) = d(v) - 1$. Then for every directed tree $T \in \mathcal{WD}(\mu, d)$ there exists such a directed tree $T' \in \mathcal{WD}(\mu, d')$ that $\mathbf{f}(T') \succ \mathbf{f}(T)$.*

Proof By Theorem 2 from [10], such a directed Huffman tree $H \in \mathcal{WD}(\mu, d)$ exists, that $\mathbf{f}(H) \succeq \mathbf{f}(T)$.[3] Since $H \in \mathcal{WD}(\mu, d)$ and $d'(v) = d(v) - 1 \geqslant 1$, we know that $d(v) \geqslant 2$ and v has an incoming arc in H from some vertex $v' \in V$. Weights are degree-monotone in $\langle \mu, d \rangle$, therefore, by construction of a Huffman tree, $f_H(u) \geqslant f_H(v)$ and, thus, without loss of generality we can assume that $u \notin g_H(v)$.

Assume that $v \in g_H(u)$. Then a path $(v, m_1, ..., m_l, u)$ exists in H from the vertex v to the vertex u, where $l \geqslant 0$. Consider a directed tree T' obtained from H by deleting the arc $v'v$ and adding the arc $v'u$ instead. It is clear that $T' \in \mathcal{WD}(\mu, d')$, and weights of groups subordinated to vertices $v, m_1, ..., m_l$ decrease by $f_H(v')$ (which is positive by Note 1), while weights of the other vertices do not change. Therefore, by Lemma 2,

$$\mathbf{y} := (f_{T'}(v), f_{T'}(m_1), ..., f_{T'}(m_l)) =$$

[3]Please note the different notation in [10] ($x \preceq_w y$ is used in [10] where we write $x \succeq y$).

$$= (f_H(v) - f_H(v'), f_H(m_1) - f_H(v'), ..., f_H(m_l) - f_H(v')) \succ$$
$$\succ (f_H(v), f_H(m_1), ..., f_H(m_l)) =: \mathbf{x}.$$

If one denotes with $\mathbf{z}$ the vector of (unchanged) weights of groups subordinated to all other non-terminal vertices of H, then, by Lemma 3, $\mathbf{f}(T') = (\mathbf{y}, \mathbf{z}) \succ (\mathbf{x}, \mathbf{z}) = \mathbf{f}(H)$.

Assume now that $v \notin g_H(u)$. Then there are disjoint paths $(u, m_1, ..., m_k, m)$ and $(v, m'_1, ..., m'_l, m)$ (where $k, l \geqslant 0$) in H from vertices u and v to some vertex $m \in M(H)$.

If $f_H(u) > f_H(v)$, then, applying repeatedly formula (22) from Lemma 5, we write $f_H(m_i) > f_H(m'_i), i = 1, ..., \min[k, l]$. It also follows from (22) that $k \leqslant l$, since otherwise $f_H(m_{l+1}) > f_H(m)$, which is impossible, since $m_{l+1} \in g_H(m)$.

Consider a directed tree $T' \in \mathcal{WD}(\mu, d')$ obtained from H by deleting the arc $v'v$ and adding the arc $v'u$ instead. In the tree T' weights of the groups subordinated to the vertices $u, m_1, ..., m_k$ increase by $f_H(v')$ (i.e., $f_{T'}(u) = f_H(u) + f_H(v')$, $f_{T'}(m_i) = f_H(m_i) + f_H(v'), i = 1, ..., k$), weights of the groups subordinated to the vertices $v, m'_1, ..., m'_l$ decrease by $f_H(v')$ (i.e., $f_{T'}(u) = f_H(u) - f_H(v')$, $f_{T'}(m'_i) = f_H(m'_i) - f_H(v'), i = 1, ..., l$), weights of all other vertices (including m) do not change. Therefore, by Lemma 2,

$$\mathbf{y} := (f_{T'}(u), f_{T'}(m_1), ..., f_{T'}(m_k), f_{T'}(v), f_{T'}(m'_1), ..., f_{T'}(m'_l)) =$$

$$= (f_H(u) + f_H(v'), f_H(m_1) + f_H(v'), ..., f_H(m_k) + f_H(v'),$$

$$f_H(v) - f_H(v'), f_H(m'_1) - f_H(v'), ..., f_H(m'_l) - f_H(v')) \succ$$

$$\succ (f_H(u), f_H(m_1), ..., f_H(m_k), f_H(v), f_H(m'_1), ..., f_H(m'_l)) =: \mathbf{x}.$$

If $\mathbf{z}$ is a vector of (unchanged) weights of groups subordinated to all other non-terminal vertices of H, then, by Lemma 3, $\mathbf{f}(T') = (\mathbf{y}, \mathbf{z}) \succ (\mathbf{x}, \mathbf{z}) = \mathbf{f}(H)$.

By construction of the Huffman tree, the situation of $f_H(u) = f_H(v)$ is possible only when $d(u) = d(v)$ and $\mu(u) = \mu(v)$. In this case we cannot use formula (22) to compare subordinate groups' weights of elements of both chains, since all possible alternatives of $k = 0$, or $l = 0$, or any sign of the expression $f_H(m_1) - f_H(m'_1)$ in case of $k, l \geqslant 1$ are possible.

On the other hand, if $f_H(m_1) > f_H(m'_1)$, then formula (22) can be used to show that $k \leqslant l$, $f_H(m_i) > f_H(m'_i), i = 2, ..., k$. In case of the opposite inequality, $f_H(m_1) < f_H(m'_1)$, formula (22) says that, by contrast, $k \geqslant l$, $f_H(m_i) < f_H(m'_i)$, $i = 2, ..., l$. Repeating this argument through the chain, we see that only two alternatives are possible:

- $0 \leqslant p \leqslant k \leqslant l$, $f_H(m_i) = f_H(m'_i)$, $i = 1, ..., p$, $f_H(m_i) > f_H(m'_i)$, $i = p+1, ..., k$. In this case, as above, we can show that for the directed tree $T' \in \mathcal{WD}(\mu, d')$ obtained from H by deleting the arc $v'v$ and adding the arc $v'u$ instead, $\mathbf{f}(T') \succ \mathbf{f}(H)$.
- $0 \leqslant p \leqslant l \leqslant k$, $f_H(m_i) = f_H(m'_i)$, $i = 1, ..., p$, $f_H(m_i) < f_H(m'_i)$, $i = p+1, ..., l$. In this case the same inequality is true for the directed tree $T' \in \mathcal{WD}(\mu, d')$ obtained from H by redirecting to/from vertex v all arcs incident to u, and by redirecting to/from vertex u all arcs incident to v except the arc $v'v$.

Therefore, we proved that a tree $T' \in \mathcal{WD}(\mu, d')$ exists such that $\mathbf{f}(T') \succ \mathbf{f}(H)$. As shown above, $\mathbf{f}(H) \succeq \mathbf{f}(T)$, so, finally, $\mathbf{f}(T') \succ \mathbf{f}(T)$. ∎

Definition 11 A directed tree $T \in \mathcal{WD}(\mu, d)$ with a terminal vertex t is called a *proper tree* if for all $m \in M(T)$, $m \neq t$, $f_T(m) \leqslant \bar{\mu}/2$.

Lemma 7 *Let a function $\chi(x)$ be concave and increasing for $x \in [0, \bar{\mu}/2]$. Consider a pair of generating tuples, $\langle \mu, d \rangle$ and $\langle \mu, d' \rangle$, satisfying conditions of Lemma 6. If $T \in \mathcal{WD}(\mu, d)$ and $T' \in \mathcal{WD}(\mu, d')$ are directed Huffman trees, then*

$$\sum_{v \in V(T') \setminus \{t'\}} \chi(f_{T'}(v)) < \sum_{v \in V(T) \setminus \{t\}} \chi(f_T(v)),$$

where $t \in M(T)$ and $t' \in M(T')$ are terminal vertices of T and T' respectively.

Proof From Lemma 6, such a tree $T'' \in \mathcal{WD}(\mu, d')$ exists that $\mathbf{f}(T'') \succ \mathbf{f}(T)$. Theorem 2 from [10] says that $\mathbf{f}(T') \succeq \mathbf{f}(T'')$, therefore, $\mathbf{f}(T') \succ \mathbf{f}(T)$. Denote for short $n_1 = |V(T)|$, $\mathbf{f}(T) = \mathbf{f} := (f_1, ..., f_{n_1-1})$, $\mathbf{f}(T') = \mathbf{f}' := (f'_1, ..., f'_{n_1-1})$.

It is known (see Lemma 19 in [10]) that each directed Huffman tree with degree-monotone weights is a proper tree, so, $f_T(w) \leqslant \bar{\mu}/2$, $w \in M(\mu, d) \setminus \{t\}$, and $f_{T'}(w') \leqslant \bar{\mu}/2$, $w' \in M(\mu, d') \setminus \{t'\}$. If a vertex $w \in V(T)$ exists, such that $\mu(w) > \bar{\mu}/2$ (there can be at most one such vertex in $V(T)$), then w cannot be an internal vertex in T and a pendent vertex in T', since then conditions of Lemma 6 imply that $w = v$ and $\mu(u) \geqslant \mu(w)$, which is impossible. Therefore, w is either a terminal vertex both in T and in T', or a pendent vertex both in T and in T'. In the latter case $f_T(v) = f_{T'}(v) = \mu(v)$.

Consequently, $f_i, f'_i \leqslant \bar{\mu}/2$ for $i = 1, ..., n_1 - 2$, and if $f_{n_1-1} > \bar{\mu}/2$, then $f_{n_1-1} = f'_{n_1-1} = \mu(w)$.

If $f_{n_1-1} \leqslant \bar{\mu}/2$, the statement of the lemma follows from Lemma 4.

If $f_{n_1-1} > \bar{\mu}/2$, we can write

$$\sum_{v\in V(T)\setminus\{t\}} \chi(f_T(v)) - \sum_{v\in V(T')\setminus\{t'\}} \chi(f_{T'}(v)) =$$

$$= \sum_{i=1}^{n_1-2} \chi(f_i) + \chi(f_{n_1-1}) - \sum_{i=1}^{n_1-2} \chi(f'_i) - \chi(f'_{n_1-1}) = \sum_{i=1}^{n_1-2} \chi(f_i) - \sum_{i=1}^{n_1-2} \chi(f'_i).$$

Since $\mathbf{f}' \succ \mathbf{f}$ and $f_{n_1-1} = f'_{n_1-1}$, we have $(f'_1, ..., f'_{n_1-2}) \succ (f_1, ..., f_{n_1-2})$, and the statement of the lemma again follows from Lemma 4. ∎

Corollary 3 *If generating tuples* $\langle \mu, d\rangle$ *and* $\langle \mu, d'\rangle$ *satisfy conditions of Lemma* 6*, then* $VWWI^*_{\mathcal{WT}}(\mu, d') < VWWI^*_{\mathcal{WT}}(\mu, d)$.

Proof Theorem 3 from [10] says that $\mathcal{WD}^*_{VWWI}(\mu, d)$ consists of all directed Huffman trees. Consider the directed Huffman trees $T \in \mathcal{WD}^*_{VWWI}(\mu, d)$, $T' \in \mathcal{WD}^*_{VWWI}(\mu, d')$. Function $\chi(x) = x(\bar{\mu} - x)$ in (21) satisfies the conditions of Lemma 7, so, from (21), $VWWI(T') < VWWI(T)$. Since every proper directed tree from $\mathcal{WD}(\mu, d)$ has a corresponding tree from $\mathcal{WT}(\mu, d)$, and vice versa, we have $VWWI^*_{\mathcal{WT}}(\mu, d) = VWWI^*_{\mathcal{WD}}(\mu, d)$, and the corollary follows immediately. ∎

In the rest of the section we consider a set V consisting of n vertices with weights $\mu(v)$, $v \in V$. The set V can be thought of as a fixed collection of (heterogeneous) atoms used as building blocks for molecules. All molecules constructed from these building blocks belong to $\mathcal{WT}(\mu)$.

We want to use Corollary 3 to show that, similar to Theorem 1, the vertex-weighted Wiener index is minimized by a tree having as many vertices of the maximum degree as possible. We cannot apply Corollary 3 to the whole collection $\mathcal{WT}(\mu)$ of trees with vertices having fixed weights $\mu(\cdot)$ (unless $\mu(v) \equiv const$), as it inevitably contains trees generated by the tuples with non-degree-monotone weights, for which Corollary 3 is inapplicable. Therefore, we have to carefully limit a set of admissible trees.

Definition 12 Consider an admissible set $\mathcal{M} \subseteq \mathcal{WT}(\mu)$ and denote with $L := \bigcap_{T\in\mathcal{M}} W(T)$ the set of vertices, which are pendent in all trees from $\mathcal{M}$. A generating tuple $\langle \mu, \bar{d}\rangle$ is called *extremal* for $\mathcal{M}$, if $M(\mu, \bar{d})$ consists of $\lceil \frac{n-2}{3} \rceil$ vertices having the highest weights in $V\setminus L$ (i.e., if $u \in M(\mu, \bar{d})$ and $v \in V\setminus L$, then $\mu(u) \geqslant \mu(v)$), at most one

vertex $u \in M(\mu, \bar{d})$ has degree $\bar{d}(u) < 4$ while others having degree 4, and, when exists, u has the minimal weight in $M(\mu, \bar{d})$. □

There can be several extremal tuples for an admissible set, if for some extremal tuple $\langle \mu, \bar{d} \rangle$ such vertices $w \in W(\mu, \bar{d}) \backslash L$ and $m \in M(\mu, \bar{d})$ exist that $\mu(w) = \mu(m)$ (swapping w and m then makes a new extremal tuple). It is clear that weights are degree-monotone in $\langle \mu, \bar{d} \rangle$, and there are only extremely branched trees in $\mathcal{WT}(\mu, \bar{d})$.

Theorem 3 *If $\langle \mu, \bar{d} \rangle$ is an extremal tuple for an admissible set $\mathcal{M} \subseteq \mathcal{WT}(\mu)$ of **chemical** trees with degree-monotone weights, and $H \in \mathcal{WT}(\mu, \bar{d})$ is a Huffman tree, then $VWWI(H) \leqslant VWWI^*_{\mathcal{M}}$, with equality if and only if $\mathcal{M}$ contains H or any other Huffman tree generated by the extremal tuple.*

Proof Consider a vertex-weighted tree $T \in \mathcal{M}$. Weights are degree-monotone in T, so, if $d_T(u) > d_T(v)$ and $\bar{d}(u) < \bar{d}(v)$ for some $u, v \in M(T)$, then $\mu(u) = \mu(v)$. Consequently, such a tree $T' \in \mathcal{WT}(\mu)$ exists that $VWWI(T) = VWWI(T')$, $W(T') = W(T)$, and from $u, v \in M(T)$ and $d_{T'}(u) < d_{T'}(v)$ it follows that $\bar{d}(u) \leqslant \bar{d}(v)$ (T' is constructed from T with a permutation of several vertices of equal weight, which does not affect the index value). So, degrees in the tree T' are "compatible" to those in the tuple $\langle \mu, \bar{d} \rangle$.

Assume $T' \in \mathcal{WT}(\mu, d^0)$ for some generating tuple $\langle \mu, d^0 \rangle$, and $d^0(\cdot) \not\equiv \bar{d}(\cdot)$. Construct such a sequence $\langle \mu, d^0 \rangle, \langle \mu, d^1 \rangle, ..., \langle \mu, d^k \rangle$ of generating tuples with degree-monotone weights that $k \geqslant 1$, $d^k(\cdot) \equiv \bar{d}(\cdot)$, and each pair $\langle \mu, d^i \rangle, \langle \mu, d^{i+1} \rangle$ of sequential tuples meets the requirements of Lemma 6, $i = 0, ..., k-1$.

We build the elements of this sequence one by one. For a tuple $\langle \mu, d^i \rangle$ define $M^+(\mu, d^i) := \{w \in V : d^i(w) < \bar{d}(w)\}$ and $M^-(\mu, d^i) := \{w \in V : d^i(w) > \bar{d}(w)\}$. Set $d^{i+1}(u^i) = d^i(u^i) + 1$ for a vertex u^i having the highest weight in $M^+(\mu, d^i)$, and $d^{i+1}(v^i) = d^i(v^i) - 1$ for a vertex v^i having the least weight in $M^-(\mu, d^i)$, while keeping the degrees of all other vertices.

The proof is clear from Fig. 3, where numbers on arrows show steps of transformation of the initial degree function. By Corollary 3, $VWWI^*_{\mathcal{WT}}(\mu, \bar{d}) < ... < VWWI^*_{\mathcal{WT}}(\mu, d^0) \leqslant VWWI(T') = VWWI(T)$. From Theorem 2 we know that if $T \in \mathcal{WT}(\mu, \bar{d})$, then $VWWI(T) = VWWI^*_{\mathcal{WT}}(\mu, \bar{d})$ if and only if T is a Huffman tree. Therefore, for every $T \in \mathcal{M}$ we have $VWWI^*_{\mathcal{WT}}(\mu, \bar{d}) \leqslant VWWI(T)$ with equality if and only if T is a Huffman tree for $\langle \mu, \bar{d} \rangle$ or another extremal tuple, and the proof is complete. ■

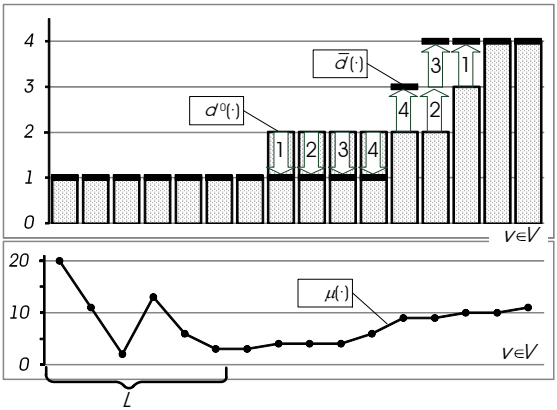

Figure 3: Sequence of degree functions converging to $\bar{d}(\cdot)$. A brace shows L.

Corollary 4 *Trees, where $WI_O(\cdot)$ achieves its minimum over the set of pendent-rooted trees of order $n = 4, ..., 14$, are depicted in Fig. 4.*

PROOF As we already argued, $WI_O(\cdot)$ can be seen a special case of $VWWI(\cdot)$ for the vertex set where all vertices have sufficiently small weight ε except a root having weight $1/\varepsilon$. Let r be the root in all considered pendent-rooted trees. Then the set of pendent-rooted trees satisfies conditions of Theorem 3 with $L = \{r\}$, and the statement of the corollary follows from Theorem 3 and the fact that for a given n the value of the index cannot be improved only for the extremal degree function $\bar{d}(\cdot)$, the one distinct from 1 and 4 at no more than one vertex. The concrete degree function for each n is justified from (17), as in Theorem 1. From [10] we know that, for a given degree function $\bar{d}(\cdot)$, Huffman trees (shown in Fig. 4) minimize $VWWI(\cdot)$. ∎

The extension of the above results to trees with vertex degrees limited to any $\Delta \geqslant 3$ is straightforward.

4 Minimal Boiling Point

Below we combine the theorems from the previous section to find an alcohol molecule with the minimal (predicted) boiling point. The main tool will be the following obvious proposition.

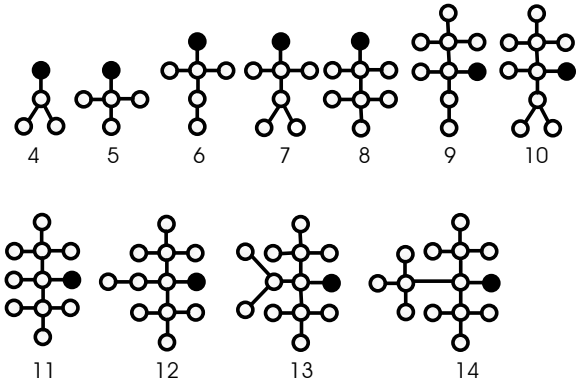

Figure 4: Rooted chemical trees minimizing $WI_{\mathrm{O}}(\cdot)$ for $n = 4, ..., 14$ (alcohol molecules minimizing $BP^{\mathrm{II}}(\cdot)$ with oxygens filled with black)

Proposition 1 *For an arbitrary admissible set $\mathcal{M}$ of graphs and a pair of indices $I(G)$ and $J(G)$ defined for all graphs $G \in \mathcal{M}$, if $\mathcal{M}_I^* \cap \mathcal{M}_J^* \neq \emptyset$, then $\mathcal{M}_{I+J}^* = \mathcal{M}_I^* \cap \mathcal{M}_J^*$.*

Proof is straightforward, since $\mathcal{M}_I^* \cap \mathcal{M}_J^*$ is a collection of graphs where indices $I(\cdot)$ and $J(\cdot)$ attain their minima simultaneously. ∎

We start with the simplest Regression I (see Equation (8)) being a linear combination of degree-based indices – the "ad-hoc index" $C^{\mathrm{I}}(\cdot)$ defined with Equation (10) and $S_2(\cdot)$ index, which penalizes presence of a sub-root of degree 2 in a pendent-rooted tree.

Proposition 2 *The collection $\mathcal{R}_{BP^{\mathrm{I}}}^*(n)$ of pendent-rooted trees (alcohol molecules) with the minimal predicted boiling point BP^{I} for $n = 4, ..., 14$ consists of the trees depicted in Fig. 2 with black circles representing a possible root (oxygen atom) positions in a tree (alcohol molecule).*

Proof Using Table 3 one easily checks conditions (12)-(14) of Theorem 1 to hold for the "ad-hoc index" $C^{\mathrm{I}}(\cdot)$ and $n \leqslant 14$. Therefore, by Corollary 2, trees from $\mathcal{T}_{C^{\mathrm{I}}}^*(n)$ are depicted in Fig. 2. We know that the "ad-hoc" index does not care for the root position in a tree, and, therefore, $\mathcal{R}_{C^{\mathrm{I}}}^*(n)$ can be obtained from $\mathcal{T}_{C^{\mathrm{I}}}^*(n)$ by assigning the root to any pendent vertex of all optimal trees.

From Lemma 1, $S_2(\cdot)$ is minimized with pendent-rooted trees whose sub-root has degree 3 or 4. Each tree in Fig. 2 has a pendent vertex being incident to a vertex of degree 3 or 4, so, $\mathcal{R}^*_{C^{\rm I}}(n) \cap \mathcal{R}^*_{S_2}(n) \neq \emptyset$, and conditions of Proposition 1 hold for all $n = 4, ..., 14$. Therefore, $\mathcal{R}^*_{BP^{\rm I}}(n)$ consists of the trees from Fig. 2 with the root assigned to a pendent vertex incident to a vertex of degree 3 or 4. Possible roots are filled with black color in Fig. 2. ∎

The similar reasoning can be carried out for Regression II, which is defined with Equation (9) and adds up from the generalized first Zagreb index $C_1^{\rm II}(\cdot)$, the cube root of the distance $WI_{\rm O}(\cdot)^{\frac{1}{3}}$ of the oxygen atom in an alcohol molecule, and, again, from $S_2(\cdot)$ index.

Proposition 3 *The collection* $\mathcal{R}^*_{BP^{\rm II}}(n)$ *of rooted trees* (*alcohol molecules*) *with the minimal predicted boiling point* $BP^{\rm II}$ *for* $n = 4, ..., 14$ *consists of the trees depicted in Fig.* 4 *with a black circle representing a root* (*oxygen atom*) *position in a tree* (*alcohol molecule*).

Proof Using Table 3 we check that conditions (11)-(13) of Theorem 1 hold, therefore, from Corollary 1, the generalized first Zagreb index $C_1^{\rm II}(\cdot)$ is minimized with **any** extremely branched tree.

According to Corollary 4, $\mathcal{R}^*_{WI_{\rm O}}(n)$ includes only extremely branched trees depicted in Fig. 4. Since the cube root is a monotone function, by Proposition 1, $\mathcal{R}^*_{WI_{\rm O}^{1/3}+C_1^{\rm II}}(n) = \mathcal{R}^*_{WI_{\rm O}}(n)$.

The sub-root of all trees depicted in Fig. 4 has degree 3 or 4. Therefore, $\mathcal{R}^*_{WI_{\rm O}}(n) \cap \mathcal{R}^*_{S_2}(n) = \mathcal{R}^*_{WI_{\rm O}}(n)$, and, finally, by Proposition 1, $\mathcal{R}^*_{BP^{\rm II}}(n)$ consists of rooted trees from Fig. 4 with roots filled black. ∎

The Basic regression (see Equation (7)) combines the "ad-hoc" index $C^0(\cdot)$, the cube root of the oxygen's distance $WI_{\rm O}(\cdot)^{\frac{1}{3}}$, and $S_2(\cdot)$ index. From Corollary 4 we know that $WI_{\rm O}(\cdot)$ is minimized with an extremely branched tree, and so is $WI_{\rm O}(\cdot)^{\frac{1}{3}}$. At the same time, neither of the inequalities (11)-(14) from Theorem 1 holds for $C^0(\cdot)$, and we cannot be sure that $C^0(\cdot)$ is minimized with an extremely branched tree. Therefore, we cannot prove formally that $BP^0(\cdot)$ achieves its minimum at some extremely branched tree.

Nevertheless, since for both simplified regressions ($BP^{\rm I}$ and $BP^{\rm II}$) an extremely branched tree appears to be optimal, optimality of an extremely branched tree for the

Basic regression $BP^0(\cdot)$ is believed to be a credible hypothesis, which we state below formally.

Conjecture 1 *If $BP^0(T) = BP^{0*}_{\mathcal{R}}(n)$ for some pendent-rooted tree $T \in \mathcal{R}(n)$, then T is an extremely branched tree.*

Proposition 4 *If Conjecture* 1 *is supposed to hold, the collection $\mathcal{R}^*_{BP^0}(n)$ of rooted trees (alcohol molecules) with the minimal predicted boiling point BP^0 for $n = 4, ..., 14$ consists of the trees depicted in Fig.* 5 *with a black circle representing a root (oxygen atom) position in a tree (alcohol molecule).*

PROOF From Conjecture 1, BP^0-minimizer is an extremely branched tree. From Fig. 2 and 4 we see that for $n = 4, 5, 6, 8$ optimal trees for $WI_{\mathrm{O}}(\cdot)$ and for $C^0(\cdot)$ coincide, while for $n = 7, 11, 14$ $\mathcal{R}^*_{WI_{\mathrm{O}}}(n) \subset \mathcal{R}^*_{C^0}(n)$. Since all sub-roots in $\mathcal{R}^*_{WI_{\mathrm{O}}}(n)$ have degree 3 or 4, from Proposition 1 we conclude that $\mathcal{R}^*_{BP^0}(n) = \mathcal{R}^*_{WI_{\mathrm{O}}}(n)$ for such n.

For $n = 9, 10, 12, 13$, if Conjecture 1 is assumed to hold, both WI_{O}-minimizers and C^0-minimizers are extremely branched trees depicted in Fig. 2 and 4, but $\mathcal{R}^*_{WI_{\mathrm{O}}}(n) \cap \mathcal{R}^*_{C^0}(n) = \emptyset$, and we cannot use Proposition 1. At the same time, the sets of extremely branched rooted trees of order n are remarkably small for $n = 9, 10, 12, 13$, and are completely enumerated in Fig. 6 along with their predicted normal boiling points BP^0. The tree with the least boiling point for each n is framed in Fig. 6. Combining the above findings we obtain Fig. 5. ∎

Predictions of the minimal boiling point based on the Wiener index slightly differ from those based on the "ad-hoc" index $C^{\mathrm{I}}(\cdot)$. Minimizers of both indices are extremely branched trees. For $n = 4, 5, 6, 8$ optimal trees for $WI_{\mathrm{O}}(\cdot)$ and for $C^{\mathrm{I}}(\cdot)$ coincide. For $n = 7, 11, 14$ $\mathcal{R}^*_{WI_{\mathrm{O}}}(n) \subset \mathcal{R}^*_{C^{\mathrm{I}}}(n)$, so, $WI_{\mathrm{O}}(\cdot)$ appears to be a refinement of $C^{\mathrm{I}}(\cdot)$, and Regression II refines Regression I.

For $n = 9, 10, 12, 13$ minimizers of $WI_{\mathrm{O}}(\cdot)$ and of $C^{\mathrm{I}}(\cdot)$ differ. By Corollaries 2 and 4, extremely branched trees minimize both indices. They have the only internal vertex v of non-maximal degree, but $C^{\mathrm{I}}(\cdot)$ suggests settling this vertex in the very center of a molecule surrounding it with the other internal vertices, while $WI_{\mathrm{O}}(\cdot)$ says vertex v must be a *stem vertex* incident to only one internal vertex. Therefore, predictions of Regression II differ from those of Regression I.

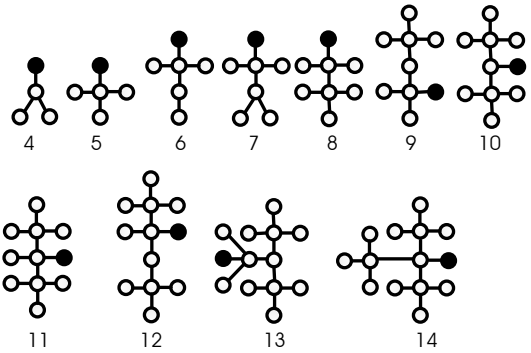

Figure 5: Pendent-rooted chemical trees (alcohol molecules) minimizing $BP^0(\cdot)$ for $n = 4, ..., 14$ with black circles being roots (oxygen atoms)

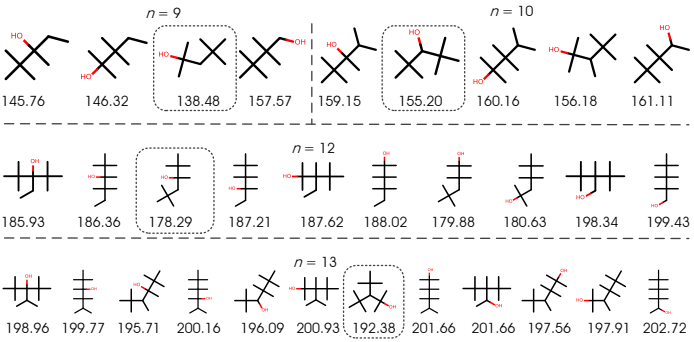

Figure 6: Extremely branched pendent-rooted trees of order $n = 9, 10, 12, 13$ and their BP^0 values (minima are framed)

Basic regression, which contains a weighted linear combination of the Wiener index and the "ad-hoc" index, represents a sort of intermediate behavior between these extremal trends (at least for extremely branched trees under Conjecture 1). The weight of the Wiener index in the regression is not enough to move the minimal tree sufficiently from the trees depicted in Fig. 2, and Basic regression becomes a yet another refinement of Regression I.

5 Conclusion

The focus of this paper is development of optimization techniques for combinations of some well-known and novel topological indices over chemically interesting sets of graphs. We derived conditions under which an extremely branched tree minimizes the sum of the second Zagreb index and of the generalized first Zagreb index. We also found minimizers of the vertex-weighted Wiener index over the set of chemical trees with given vertex weights.

We enumerated index minimizers for moderate (up to 14) non-hydrogen atom count in a molecule, and combined them in several regressions of different complexity to forecast a simple alcohol molecule with the lowest boiling point.

For simpler regressions (Regressions I and II) we managed to obtain a complete analytical characterization of extremal alcohol molecules, while for the most complex (yet the most precise) "basic" regression we had to limit our attention to extremely branched trees (see Conjecture 1) and employed the brute-force enumeration to find molecules of low-boiling alcohols.

Forecasts based on different regressions slightly differ, but they all comply with the collected experimental data on normal boiling points of simple alcohols.

Finally, let us sketch several promising directions of future research. An obvious shortcoming of this paper is Conjecture 1, which is explained but is not proven formally. To justify it, we need to refine sufficiently our optimization techniques, *viz*, to optimize jointly the Wiener index and the Zagreb indices.

On the other hand, there is wide space for investigation of popular combinations of topological indices forecasting important physical and chemical properties of compounds.

References

[1] A. T. Balaban, J. Devillers (Eds.), *Topological indices and related descriptors in QSAR and QSPAR*, CRC Press, Boca Raton, 2000.

[2] U. Bren, V. Martínek, J. Florián, Decomposition of the Solvation Free Energies of Deoxyribonucleoside Triphosphates Using the Free Energy Perturbation Method, *J. Phys. Chem. B* **110** (2006), 12782–12788.

[3] U. Bren, D. Janežič, Individual Degrees of Freedom and the Solvation Properties of Water, *J. Chem. Phys.* **137** (2012) 024108.

[4] B. Bollobás, *Extremal graph Theory*, Academic Press, London, 1978.

[5] A. A. Dobrynin, R. Entringer, I. Gutman, Wiener index of trees: theory and applications, *Acta Applicandae Mathematica* **66** (2001) 211–249.

[6] M. Fischermann, A. Hoffmann, D. Rautenbach, L. Székely, L. Volkmann, Wiener index versus maximum degree in trees, *Discrete Applied Mathematics* **122** (2002) 127–137.

[7] B. Furtula, I. Gutman, H. Lin, More trees with all degrees odd having extremal Wiener index, *MATCH Commun. Math. Comput. Chem.* **70** (2013) 293–296.

[8] R. García–Domenech, J. Gálvez, J. V. de Julián–Ortiz, L. Pogliani, Some new trends in chemical graph theory, *Chem. Rev.* **108** (2008) 1127–1169.

[9] M. Goubko, Minimizing degree-based topological indices for trees with given number of pendent vertices, *MATCH Commun. Math. Comput. Chem.* **71** (2014) 33–46.

[10] M. Goubko, Minimizing Wiener index for vertex-weighted trees with given weight and degree sequences, *MATCH Commun. Math. Comput. Chem.* **75** (2016) 3–27.

[11] `http://www.mtas.ru/upload/library/GoubkoMilosOnline14.pdf` (last accessed Oct 30, 2014).

[12] I. Gutman, N. Trinajstić, Graph theory and molecular orbitals. Total π-electron energy of alternant hydrocarbons, *Chem. Phys. Lett.* **17** (1972) 535–538.

[13] I. Gutman, B. Furtula (Eds.), *Novel molecular structure descriptors — Theory and Applications I*, Univ. Kragujevac, Kragujevac, 2010.

[14] I. Gutman, B. Furtula (Eds.), *Novel molecular structure descriptors — Theory and Applications II*, Univ. Kragujevac, Kragujevac, 2010.

[15] D. Janežič, B. Lučić, S. Nikolić, A. Miličević, N. Trinajstić, Boiling points of alcohols — a comparative QSPR study, *Internet Electronic Journal of Molecular Design* **5** (2006), 192–200.

[16] S. Klavžar, I. Gutman, Wiener number of vertex-weighted graphs and a chemical application, *Discrete Appl. Math.* **80** (1997) 73–81.

[17] M. Kompany–Zareh, A QSPR study of boiling point of saturated alcohols using genetic algorithm, *Acta chimica slovenica* **50** (2003) 259–273.

[18] X. Li, Y. Shi, A survey on the Randic index, *MATCH Commun. Math. Comput. Chem.* **59** (2008) 127–156.

[19] H. Lin, Extremal Wiener index of trees with all degrees odd, *MATCH Commun. Math. Comput. Chem.* **70** (2013) 287–292.

[20] A. W. Marshall, I. Olkin, *Inequalities: theory of majorization and its applications, mathematics in science and engineering*, V. 143, Academic Press, New York, 1979.

[21] P. Penchev, N. Kochev, V. Vandeva, G. Andreev, Prediction of boiling points of acyclic aliphatic alcohols from their structure, *Traveaux Scientifiquesd'Universite de Plovdiv* **35** (2007) 53–57.

[22] M. Randić, On characterization of molecular branching, *J. Amer. Chem. Soc.* **97** (1975) 6609.

[23] J. Rada, R. Cruz, Vertex-degree-based topological indices over graphs, *MATCH Commun. Math. Comput. Chem.* **72** (2014) 603–616.

[24] R. Todeschini, V. Consonni, *Handbook of molecular decriptors*, Wiley-VCH, Weinheim, 2000.

[25] R. Todeschini, V. Consonni, *Molecular decriptors for chemoinformatic*, Wiley-VCH, Weinheim, 2009.

[26] H. Wang, The extremal values of the Wiener index of a tree with given degree sequence, *Discrete Appl. Math.* **156** (2008) 2647–2654.

[27] H. Wiener, Structrual determination of paraffin boiling points, *J. Am. Chem. Soc.* **69** (1947) 17–20.

[28] K. Xu, M. Liu, K. C. Das, I. Gutman, B. Furtula, A survey on graphs extremal with respect to distance-based topological indices, *MATCH Commun. Math. Comput. Chem.* **71** (2014) 461–508.

[29] X.-D. Zhang, Q.-Y. Xiang, L.-Q. Xu, R.-Y. Pan, The Wiener index of trees with given degree sequences, *MATCH Commun. Math. Comput. Chem.* **60** (2008) 623–644.

[30] B. Zhou, I. Gutman, Relations between Wiener, hyperWiener and Zagreb indices, *Chem. Phys. Lett.* **394** (2004) 93–95.