FACULTY OF MATHEMATICS
UNIVERSITY OF BELGRADE

Proceedings of the Symposium
Contemporary Mathematics

Devoted to the 125 anniversary of Faculty of Mathematics
and to 190 years of teaching of mathematics in Serbia
18-20 December 1996, Belgrade, Serbia

On the Randić Connectivity Index and Physical Properties of Alkanes

Dušan Repovš[1], Vladimir Bobrovskij[2], and Elena Muranova[3]

[1] Institute of Mathematics, Physics and Mechanics, Ljubljana, Slovenia 1001
[2] Vitebsk State Technological University, Vitebsk, Belarus 210028
[3] Institute of Technical Acoustics, Vitebsk, Belarus 210023

Abstract. We introduce a new method of constructing indices which generalize the classical Randić connectivity index: Our new indices make it possible to get a good correlation between physical and chemical properties of chemical species. Our method is then applied to obtain a new index which gives sufficiently exact formulae for the boiling point temperatures of alkanes. These formulae are simpler and no less exact than the previous ones.

1 Introduction

The application of the graph theory for constructing topological indices, which are then used to predict physical and chemical properties of chemical species, has a long history. More than 20 various indices have been discovered so far, many of which correlate well with physical properties of various substances. Studies of these indices and discoveries of new ones is a very active area of research [3].

In the present paper we investigate the Randić index [2]. Starting from the construction of this index (essentially using its additivity), we study the possibilities of constructing a family of analogous indices, some of which allow one to get a very high correlation with physical properties of certain classes of chemical species.

From the mathematical point of view, this class of indices can be considered as a modification of the Randić index. We also suggest a certain algorithm of choice of a fixed modification which is the most suitable for the solution of the particular problem of predicting physical and chemical properties of substances in each case. Our method of construction of new indices and derivation of formulae is sufficiently universal and it allows further generalizations.

As an example we consider the possibility of applying the Randić index and its possible modifications for determining the boiling point temperature of alkanes. To this end let us recall the notion of the chemical graph for alkanes. This is a graph G in which the set of vertices V coincides with the set of carbon atoms of a given alkane, and where each carbon-carbon bond between atoms c_1 and c_2 determines the edge (c_1, c_2) of the graph G from

the set E of all edges of G. Here the carbon-hydrogen bonds and hydrogen atoms are not depicted in the graph.

In our paper we obtain a functional dependence of the boiling point temperature of alkanes on the modified Randić index. This dependence is considerably simpler than that of previous formulae and is not less exact.

2 The Randić connectivity index and construction of new indices

Recall the definition of the Randić index $R(G)$ of molecular (chemical) graph G, consisting of the set of all vertices V and the set of all edges E (cf. [2]). Recall that the set E consists of some subset of the set of all pairs of vertices of G. The degree of a vertex $i \in V$ of the graph G is the number of edges $v_i \in E$ which come out of this vertex. Consider the function $f : E \to R_+$, defined on the set of all edges E of the graph G. Define $f(\alpha)$ for any edge $\alpha \in E$ joining vertices $i, j \in V$ by the following formula:

$$f(\alpha) = \frac{1}{\sqrt{v_i \cdot v_j}}. \tag{1}$$

Then the Randić index $R(G)$ of the graph G is defined by the following formula:

$$R(G) = \sum_{\alpha \in E} f(\alpha), \tag{2}$$

where the sum runs over all edges $\alpha \in E$ of the graph G.

The discriminating property of the Randić index is not very high. Already for $n = 8$ there exist two pairs of alkanes which have the same Randić index. Moreover, there exist pairs of alkanes for which the Randić indices of different alkanes differ only in the third decimal. For higher n there are many more such examples. This fact limits substantially the possibility of predicting any parameter X in the class of alkanes. In particular, we can get *à priori* estimate of the error by applying the Randić index to determine physical and chemical parameters of any class of chemical species.

Let A_1 and A_2 be different species, having different molecular graphs with coinciding indices. One should not expect that the exactness of determining the parameter X, for the given class of species, will be greater than the value of the difference $D = |X(A_1) - X(A_2)|$, where $X(A_i)$ is the value of the parameter X for the species $A_i, i \in \{1, 2\}$.

For the parameter X in the class of alkanes, let us take the boiling point temperature T and let us check what maximal exactness one can expect when applying the Randić index. Let n be the number of atoms of the carbon skeleton. For $n \leq 9$, the maximal value D is equal to 3.00 K and when n further increases, its value increases, too. We remark that the margin of error in a real problem is usually greater than this one.

Nevertheless, the Randić index correlates well with those properties which are close to the additivite parameters of alkanes, e.g. the boiling point temperature, density and various heat reactions (cf. [3]).

It follows from the definition of the Randić index that it has practically an additive character. More precisely, the function $f : E \to R_+$ defined above on the edges of the graph, depending on adjacent vertices, and the index equal to the sum of the values of f, evaluated at every edge. Here one should bear in mind two facts. First, by the definition, a chemical graph always has only finitely many edges, i.e. f is defined on a finite set. Second, in any class of species we consider, there are only finitely many types of edges, on which f assumes different values, i.e. the range of f is a finite set, too. For example, in the case of alkanes there are 10 different types of edges α_k $(1 \leq k \leq 10)$, on which f assumes values $f(\alpha_k)$. For any edge $\alpha \in E$, joining vertices $i, j \in V$, we denote by $(v_i, v_j) \in Z_+ \times Z_+$ the pair of numbers, where v_i is the number of edges originating at the vertex i and v_j is the number of edges originating at the vertex j. Then all possible types of edges and values of f are given by the following table:

k	α_k	$f(\alpha_k)$
1	(1,1)	1
2	(1,2)	$1/\sqrt{2}$
3	(1,3)	$1/\sqrt{3}$
4	(1,4)	1/2
5	(2,2)	1/2
6	(2,3)	$1/\sqrt{6}$
7	(2,4)	$1/2\sqrt{2}$
8	(3,3)	1/3
9	(3,4)	$1/2\sqrt{3}$
10	(4,4)	1/4

(3)

It follows from the definition of the Randić index that its value for a particular substance is completely determined by the values of f. One can say that the function f generates the Randić index. It is therefore natural to consider various indices, which are constructed by the same principle as the Randić index, but using a different function f'. At the same time, one can limit the choice of functions to some class F which depends on the goals of the investigation. Such indices can clearly be considered as a modification of the Randić index.

3 Applications: Calculating the boiling point temperature of alkanes

Fundamental results concerning the prediction of the boiling point temperature T of normal alkanes for the period 1842 - 1938 were exposed in [1]. The best one of them is the formula:

$$T(n) = 745.42 \cdot \log(n + 4.4) - 416.31\,, \tag{4}$$

where log stands for $\log_{10}$ and $n \in Z_+$ is the number of atoms of the carbon skeleton. This formula gives for $2 \leq n \leq 19$ the exactness $-2.5\text{K} \leq \Delta \leq 0.7\text{K}$, and for $n = 1$ it gives $\Delta = -18.08\text{K}$ (cf. [1]).

Results of [1] for branch chained alkanes have a quite substantial error and also do not give a unique functional dependence - they only consider different formulae for different homologous series (cf. [1, p.733]). Moreover, in most homologous series only 2 - 4 elements are considered which renders those results almost trivial.

In 1947 Wiener [6] discovered formulae for determing the boiling point temperature of ramified alkanes in the following form:

$$\Delta T = \frac{98}{n^2} \cdot \Delta W + 5.5 \cdot \Delta p, \quad 4 \leq n \leq 10. \tag{5}$$

Here, ΔT is the difference between the boiling point temperature of the ramified and the corresponding linear alkane, n is the number of the atoms of the carbon skeleton, W is the Wiener index, ΔW is the difference between the Wiener indices of the ramified and the corresponding normal alkane, p is the number of polarity, and Δp is the difference between the numbers of polarity of the ramified and the corresponding normal alkane. In this manner, after considering formula (4) it is clear that T depends on three different indices, since Δp is functionally expressed by p and n, whereas ΔW is functionally expressed by W and n. Independently of its complexity, the error is in some cases greater than 10K.

To get a complete picture, we also mention the paper [4], where it was shown that there is an almost linear relationship between $\ln T$ and $\ln 2W$ for any alkanes (where W is the Wiener index). Although the exactness in such an approach is quite low, the advantage is the use of only one parameter.

A formula was obtained in [5] for the boiling point temperature T by a combination of the Wiener and the Randić indices, which gives weaker results than [6], but it depends on only two indices.

For normal alkanes there exists a simple relationship between the Randić index R and the number of carbon atoms n:

$$R = \frac{n-3}{2} + \sqrt{2}\,, \tag{6}$$

or, if we express n from here via R and insert it into (4), we get a dependence of the boiling point temperature T on the Randić index R:

$$T(R) = 745.42 \cdot \log(2R + 3 - 2\sqrt{2} + 4.4) - 416.31\,. \tag{7}$$

For convenience of calculation we transform formula (7) into:

$$T(R) = 745.42 \cdot \log(R + 2.2857865) - 191.916\,. \tag{8}$$

We note that this approach is very natural and if we assume that the boiling point temperature correlates with the Randić index, then one can expect that the obtained formula will give acceptable results for ramified alkanes. Let T_{obs} be the boiling point temperature of an alkane and $T(R)$ the boiling point temperature calculated using by formula (8). Then for example, for $n = 7$ the error of calculation $\Delta = T_{obs} - T(R)$ for most alkanes is (its absolute value) less than 2K. Nevertheless, for three alkanes at $n = 7$ this error sharply rises: for 2,3-dimethylpentane we get $|\Delta| = |T_{obs} - T(R)| = 4.95$K, for 3,3-dimethylpentane we get $|\Delta| = 4.76$K, and for 2,2,3-trimethylbutane we get $|\Delta| = 10.43$K.

Thus the existence of carbon atoms which have many (3 or 4) carbon-carbon bonds in some cases significantly lowers the exactness of the results obtained by formula (8) and the calculated temperature is lower than observed. When n further increases, this exactness further diminishes. Already for 2,2,3,3-tetramethylbutane ($n = 8$) the error of calculation by using formula (8) equals to $\Delta = T_{obs} - T(R) = 379.45 - 362.06 = 17.39$K.

In such a situation application of a modified Randić index allows one without problems, to overcome such difficulties and it substantially improves the exactness of calculation. For better clarity we shall restrict ourselves to only those functions, which coincide with f on the normal alkanes and which assume only values of the type $1/\sqrt{n}$, where $n \in N$ is some natural number. In this way the class F considered, contains the function $f : E \to R_+$, defined by formula (3). Here all functions $f' \in F$ are constructed as simply as f. Instead of function f, given by formula (3), we define a new function $f'(\alpha)$ as follows:

$$f'(\alpha_k) = f'(v_i, v_j) = \begin{cases} f(\alpha_k),\ k \leq 6 \\ 1/\sqrt{7},\ k = 7\ ((v_i, v_j) = (2,4)) \\ 1/\sqrt{7},\ k = 8\ ((v_i, v_j) = (3,3)) \\ 1/\sqrt{5},\ k = 9\ ((v_i, v_j) = (3,4)) \\ 1/\sqrt{3},\ k = 10\ ((v_i, v_j) = (4,4)) \end{cases} \tag{9}$$

We define the modified index RM by function f' analogously as R is defined by function f:

$$RM(G) = \sum_{\alpha \in E} f'(\alpha)\,, \tag{10}$$

where the summation is done over all edges of the graph. In this case the relationship of $RM(n)$ is the same as in formula (6) for linear alkanes. Analogous argument as above yields the following formula:

$$T(RM) = 745.42 \cdot \log(RM + 2.2857865) - 191.92 \tag{11}$$

In the table below we have gathered the calculations and the experimental data, obtained for the octane isomers. Here $\Delta_R = T_{obs} - T(R)$ and $\Delta_{RM} = T_{obs} - T(RM)$, and all indices are rounded off to 3 decimals.

octane isomer	T_{obs}	R	RM	$T(R)$	Δ_R	$T(RM)$	Δ_{RM}
n-octane	398.82	3.914	3.914	398.74	0.08	398.74	0.08
2-methylheptane	390.80	3.770	3.770	391.13	-0.33	391.3	-0.33
3-methylheptane	392.08	3.808	3.808	393.15	-1.07	393.15	-1.07
4-methylheptane	390.86	3.808	3.808	393.15	-2.29	393.15	-2.29
2,3-dimethylhexane	388.76	3.681	3.726	386.34	2.42	388.76	0.0
2,4-dimethylhexane	382.58	3.664	3.664	385.41	-2.83	385.41	-2.83
2,5-dimethylhexane	382.25	3.626	3.626	383.34	-1.09	383.34	-1.09
3,4-dimethylhexane	390.86	3.719	3.764	388.39	2.47	390.80	0.06
3-ethylhexane	391.68	3.847	3.847	395.22	-3.54	395.22	-3.54
2-methyl-3-ethylpentane	388.80	3.719	3.764	388.39	0.41	390.80	-2.00
3,3-dimethylhexane	385.42	3.621	3.670	383.06	2.36	385.74	-0.32
2,2,4-trimethylpentane	372.39	3.417	3.441	371.69	0.70	373.05	-0.66
2,2,3,3-tetramethylbutane	379.45	3.250	3.577	362.06	17.39	380.64	-1.19
2,2-dimethylhexane	397.99	3.561	3.585	379.76	0.23	381.08	-1.09
2,3,4-trimethylpentane	386.66	3.553	3.642	379.32	7.34	384.21	2.21
2,2,3-trimethylpentane	382.99	3.481	3.640	375.30	17.69	384.10	-1.11
3-methyl-3-ethylpentane	391.41	3.682	3.755	386.39	5.02	390.33	1.08
2,3,3-trimethylpentane	387.91	3.504	3.687	376.59	11.32	386.66	1.25

It is natural to expect that there exists some "optimal" function g, which depends on the class of species, which generates the best possible correlation with the physical and chemical properties of species of the given class of the modified Randić index. Finding such functions for various classes of substances and their comparison will allow a better understanding of the nature of chemical bonds and possibilities of applications of additive indices.

References

1. Egloff, G., Sherman, J., Dull, R. B. (1940) *Boiling point relationship among aliphatic hydrocarbons.* J. Phys. Chem. **44**, 730-745
2. Randić, M. (1975) *On characterization of molecular branching.* J. Amer. Chem. Soc. **97**:23, 6609-6615
3. Rouvray, D. H. (1983) *Should we have designs on topological indices?.* Chemical Applications of Topology and Graph Theory, R. B. King, Ed. Studies Phys. Theor. Chem. **28**, Elsevier Sci. Publ., Amsterdam, pp. 159-177
4. Rouvray, D. H., Crafford, B. C. (1976) *The dependence of physical-chemical properties on topological factors.* South African J. Sci. **72**, 47-51
5. Zhurakovskij, V. R., Muranov, Yu. V., Slizhevskaya, E. N. (1988) *Topology of molecules and the boiling point temperature of alkanes.* Vesci Akad. Nauk BSSR Ser. Khim. Nauk No.5, pp.47-51 (in Russian)
6. Wiener, H. (1947) *Structural determination of paraffin boiling points.* J. Amer. Chem. Soc. **69**, 17-20