MATCH
Communications in Mathematical and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. **67** (2012) 369-381

ISSN 0340 - 6253

The Normalized Revised Szeged Index

Mustapha Aouchiche and **Pierre Hansen**

GERAD and HEC Montreal,
3000, chemin de la Côte-Sainte-Catherine
Montreal (QC) Canada H3T 2A7
e-mail: `mustapha.aouchiche@gerad.ca` ; `pierre.hansen@gerad.ca`

(Received September 27, 2010)

Abstract

In chemical graph theory, many graph parameters, or topological indices, were proposed as estimators of molecular structural properties. Often several variants of an index are considered. The aim is to extend the original concept to larger families of graphs than initially considered, or to make it more precise and discriminant, or yet to make its range of values similar to that of another index, thus facilitating their comparison. In this paper, we introduce a new variant of the Szeged index. It is named normalized revised Szeged index, and is obtained by taking the square root of the revised Szeged index divided by the number of edges in the considered graph. The spread of its values is the same as for the Randić index. We also study the correlations between the Szeged indices, as well as the Randić index, and the boiling point of chemical graphs with up to eight vertices.

1. INTRODUCTION

Mathematical descriptors of molecular structure and properties, such as various topological indices [17], have been widely used in chemical studies. Many topological indices related to the graph representation of molecular structures were proposed. Among the earliest and most important descriptors, one can find the *Wiener index* [22] introduced in 1947, the *Hosoya topological index* [10] introduced in 1971, the *Randić index* [19] introduced in 1975 and the *Szeged index* [9] inroduced in 1994 and modified to the *revised Wiener index* (also called the *revised Szeged index*) by Randić [16] in 2002. For the use of the Szeged and the revised Szeged indices see, for example, [14, 15] and the references therein. For other topological indices proposed in the litterature see e.g. [5], [18] and [21]. In this paper, we introduce a modification on the revised Szeged index to get a new topological index called the *normalized revised Szeged index*. The new index, computed from the revised Szeged index, takes into account the number of edges (bonds) of a given graph. It also fits in the same value range as the Randić index. As will be shown in this paper, the normalized revised Szeged index correlates with the boiling point of both cyclic and acyclic molecular graphs, better than both Szeged and revised Szeged indices. First, let us recall some needed definitions.

Let $G = (V, E)$ be a connected graph with vertex set V and edge set E.. For $u, v \in V$, $d(u, v)$ denotes the distance between u and v in G. Let $e = uv \in E$ and define the partition $\{N_u(e), N_v(e), N_0(e)\}$ of the vertices of G with respect to e as follows

$$\begin{aligned} N_u(e) &= \{w \in V : d(u,w) < d(v,w)\}, \\ N_v(e) &= \{w \in V : d(v,w) < d(u,w)\}, \\ N_0(e) &= \{w \in V : d(u,w) = d(v,w)\}. \end{aligned}$$

Let $n_u(e)$, $n_v(e)$ and $n_0(e)$ denote the number of vertices in $N_u(e)$, $N_v(e)$ and $N_0(e)$ respectively. In 1994, Gutman [9] introduced the *Szeged index* of a graph G as a molecular descriptor. It is defined by

$$Sz = Sz(G) = \sum_{e=uv \in E} n_u(e) \cdot n_v(e).$$

It is well known that the Szeged index generalizes the Wiener index to cyclic graphs. In 2002, Randić [16] "*critically examined the Sezged index and found it deficient as a molecular descriptor ... despite it's elegant and attractive mathematical definition*". Indeed, as pointed out by Randić [16], the Szeged index does not take into account the contributions of the vertices at equal distance from the endpoints of an edge. The problem occurs when the graph contains odd cycles. Thus, Randić [16] proposed the *revised Szeged index* that he called the *revised Wiener index* and defined it by

$$Sz^* = Sz^*(G) = \sum_{e=uv \in E} \left(n_u(e) + \frac{n_0(e)}{2} \right) \cdot \left(n_v(e) + \frac{n_0(e)}{2} \right).$$

In the same paper, Randić showed that the revised Szeged (Wiener) index has a better correlation, that the Szeged index, with the boiling point of 45 cycloalkanes.

The best known parameter to be well correlated with the boiling point of chemical compounds is the *Randić index*. It is called so after Milam Randić who introduced it under the name *connectivity index* [19] in 1975. It is defined by

$$Ra(G) = \sum_{uv \in E} \frac{1}{\sqrt{d(u)d(v)}}$$

where $d(u)$ denotes the degree (number of neighbors) of u in G.

The rest of this paper is organized as follows. In the next section, we introduce a new graph parameter called the *normalized revised Szeged index* and computed from the revised Szeged index. We propose it as an estimator of the boiling point of chemical compounds. In Section 3, we prove lower and upper bounds on the normalized revised Szeged index of several classes of graphs: trees with a given number of vertices n, connected unicyclic graphs with a given number of vertices n, connected graphs with a given number of vertices n and a given number of edges m, and connected triangle–free graphs with a given number of vertices. Section 4 is a study of different correlations between the boiling point, of all alkanes and cycloalkanes on up to 8 vertices, on the one hand, and the simple, revised and normalized revised Szeged indices as well as the Randić index on the other hand.

2. Introducing Szs^*

Both Randić and revised Szeged indices are well correlated with the boiling point of chemical compounds. However, the Randić index Ra has a better correlation than the revised Szeged index Sz^*. So, it is natural to look for similarities between Ra and Sz^* in order to improve the correlation between a possible new version of the Szeged index and the boiling point or any other chemical property.

If we consider Ra and Sz^* on the class of all connected graphs on a given number n of vertices, the star S_n is the only graph that minimizes both parameters. Also, the complete graph K_n, among others, maximizes both Ra and Sz^*. Considering the two parameters on the class of all connected graphs with given numbers n of vertices and m of edges, it is proved in [2] that

$$(n-1) \cdot m \leq Sz^*(G) \leq \frac{n^2 \cdot m}{4}.$$

Equality for the lower bound holds if and only if G is the star S_n. The upper bound is reached if and only if G is a *transmission regular graph*, *i.e.*, a graph in which the sum of the distances from each vertex to all others, also called transmission, is a constant.

The above double–inequality can be written as follows

$$n-1 \leq \frac{Sz^*}{m} \leq \frac{n^2}{4},$$

or equivalently

$$\sqrt{n-1} \leq \sqrt{\frac{Sz^*}{m}} \leq \frac{n}{2}.$$

Let us define the *normalized revised Szeged index* by dividing Sz^* by m and then taking the square root, *i.e.*,

$$Szs^* = \sqrt{\frac{Sz^*}{m}}.$$

Thus, we have

$$\sqrt{n-1} \leq Szs^*(G) \leq \frac{n}{2}.$$

These bounds are exactly the well known bounds on the Randić index, but the extremal graphs for the upper bound are not exactly the same. However, there are many common extremal graphs. Recall that the bounds on Ra are

$$\sqrt{n-1} \leq Ra(G) \leq \frac{n}{2}$$

with equality for the lower bound if and only if G is the star S_n, and for the upper bound if and only if G is a (degree) regular graph.

In addition to the similarity in the bounds, Ra and Szs^* have the same value for several families of graphs.

• As already mentioned, $Ra(S_n) = Szs^*(S_n) = \sqrt{n-1}$. The star is also the complete bipartite graph $K_{n-1,1}$, and this equality can be extended to all complete bipartite graphs. Indeed, consider the complete bipartite graph $K_{p,q}$. We have

$$Ra(K_{p,q}) = Szs^*(K_{p,q}) = \sqrt{p \cdot q}.$$

• We have also noted that Ra is maximum for regular graphs and Szs^* is maximum for transmission regular graphs. So, for any regular and transmission regular graph G

$$Ra(G) = Szs^*(G) = \frac{n}{2}.$$

Examples of regular and transmission regular graphs are the complete graph K_n, the cycle C_n the hypercube Q_k on $n = 2^k$ vertices, $K_{2k} \setminus M$, where M is a perfect matching on $n = 2k$ vertices, $K_n \setminus C_n$ for $n \geq 5$, and any regular graph of diameter $D = 2$.

Note that there exist graphs which are transmission regular but not (degree) regular [3], for instance the graph of Figure 1.

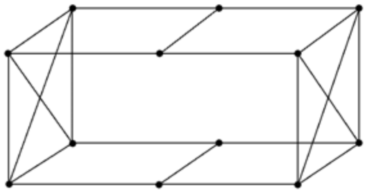

Figure 1: A transmission regular but not regular graph on 12 vertices

3. Bounds on Szs^*

In this section, we prove lower and upper bounds on the normalized revised Szeged index Szs^* of some classes of graphs, and characterize the relevant extremal grphs. Note that the results proved here were first obtained as conjectures using the AutoGraphiX system [1,7]. We first prove upper and lower bounds on Szs^* considered on the set of trees with a given number of vertices n. We also characterize the extremal trees which turn out to be the same as for the Randić index.

Proposition 1. *Let T be a tree on n vertices then*

$$Szs^*(S_n) = \sqrt{n-1} \leq Szs^*(T) \leq Szs^*(P_n) = \sqrt{\frac{n(n+1)}{6}}$$

with equality if and only if G is the star S_n for the lower bound and the path P_n for the upper bound.

Proof : The result follows from the facts that $Sz(T) = Sz^*(T) = W(T)$ and $W(S_n) \leq W(T) \leq W(P_n)$ for any tree T, where $W(T)$ denotes the Wiener index (sum of all distances) of the tree T. ■

The results about the Randić index that are similar to the above proposition are due to Bollobas and Erdös [6] for the lower bound and to Yu [23] for the upper bound. These two bounds are gathered in the following theorem.

Theorem 2. *Let T be a tree on n vertices then*

$$Ra(S_n) = \sqrt{n-1} \leq Ra(T) \leq Ra(P_n) = \frac{n-3+2\sqrt{2}}{2}$$

with equality if and only if G is S_n for the lower bound and P_n for the upper bound.

The upper bound on Szs^* on the set of trees is reached only for the path P_n as proved above. If we except P_n, what is the tree that would maximize Szs^*? The next proposition answers this question. Recall that a *comet* $Co_{n,\Delta}$ the tree a btained from a star S_Δ and a path $P_{n-\Delta}$ by joinning an endpoint of the path to the central vertex of the star.

Proposition 3. *Let T be a tree with maximum degree $\Delta \geq 3$. Then*

$$Szs^*(T) \leq \sqrt{\frac{n(n+1)}{6} + \frac{2}{n-1} - 1}$$

with equality if and only if T is the comet $Co_{n,3}$. Moreover the second largest value of Szs^* on the set of trees on n vertices is $Szs^*(Co_{n,3})$.

Proof : It is known that on the set of all trees on n vertices with maximum degree $\Delta \geq 3$, the Wiener index is maximum for and only for the comet $Co_{n,3}$ [4]. As the Szeged index and the Wiener index are equal for a tree, the bound follows as well as the characterization of the corresponding extremal graphs. As proved above, the largest value of Szs^* on the set of all trees on n vertices is attained for and only for the path P_n. So if T is not the path then the maximum degree $\Delta \geq 3$. Thus the second largest value of Szs^* is attained for and only for the comet $Co_{n,3}$. ■

Note that the second largest value of the Randić index over all trees on $n \geq 7$ vertices is not attained for the comet $Co_{n,3}$ but for a tree containing exactly one vertex of degree 3 adjacent to 3 vertices of degree 2 [11].

We now turn to bounding the normalized revised Szeged index Szs^* on the set of unicyclic graphs. In order to characterize the extremal unicyclic graphs related to the lower bound on Szs^*, we need the following definition.

A graph G on $n \geq 4$ vertices is a *turnip* $Tr_{n,g}$ if it is composed of a cycle C_g with $n-g$ pending edges all incident to the same vertex from the cycle.

Proposition 4. *Let G be a unicyclic graph on n vertices then*

$$\left.\begin{array}{ll} \sqrt{(5n^2-4n-6)/(4n)} & \text{if } n \leq 12, \\ \sqrt{n+3-12/n} & \text{if } n \geq 13 \end{array}\right\} \leq Szs^*(G) \leq Szs^*(C_n) = \frac{n}{2}$$

with equality if and only if G is $Tr_{n,3}$ if $n \leq 12$ and $Tr_{n,4}$ if $n \geq 13$ for the lower bound and C_n for the upper bound.

Proof : The lower bound follows from the similar bound on Sz^* proved in [2], namely

$$Sz^*(G) \geq \left\{\begin{array}{ll} 5n^2/4 - n - 3/2 & \text{if } n \leq 12, \\ n^2 + 3n - 12 & \text{if } n \geq 13 \end{array}\right.$$

with equality if and only if G is the turnip $Tr_{n,3}$ if $n \leq 12$ and G is the turnip $Tr_{n,4}$ if $n \geq 13$.

The upper bound follows from the upper bound on Szs^* on general connected graphs and the fact that the cycle C_n is the only unicyclic graph that is transmission regular. ■

Up to now, we proved bounds as functions of the number of vertices n only. We next prove bounds which are functions of both the numbers of vertices n and of edges m of connected graphs.

Proposition 5. *Let G be a connected graph on $n \geq 5$ vertices and m edges, then*

$$Szs^*(G) \leq \sqrt{m} + \frac{n}{2} - \sqrt{n}$$

with equality if and only if G is the cycle C_n.

Proof :

If G is a tree, using Proposition 1, we have

$$Szs^*(G) - \sqrt{m} \leq \sqrt{\frac{n(n+1)}{6}} - \sqrt{n-1} < \frac{n}{2} - \sqrt{n}$$

for all $n \geq 5$. Thus the bound is true and strict.

If G is not a tree, *i.e.*, $m \geq n$, we have

$$Szs^*(G) \leq \sqrt{m} + \frac{n}{2} - \sqrt{n}$$

with equality if and only if $Szs^*(G) = n/2$ and $m = n$, *i.e.*, if and only if G a transmission regular and unicyclic graph. The cycle C_n is the only unicyclic graph to be transmission regular. ■

Theorem 6. *Let G be a connected triangle-free graph on $n \geq 5$ vertices and m edges, then*

$$Szs^*(G) \geq \sqrt{m}$$

with equality if and only if G is a complete bipartite graph $K_{p,q}$.

Proof : Since G is a triangle-free graph, for any edge $e = uv \in E$

$$(n_u(e) + n_0(e)) \cdot (n_v(e) + n_0(e)) \geq d(u) \cdot d(v)$$

with equality if and only if $d(u) + d(v) = n$. Thus

$$Sz^*(G) \geq \sum_{uv \in E} d(u) \cdot d(v)$$

with equality if and only if $d(u) + d(v) = n$ for every edge uv in G. Assume that G minimizes Sz^* and let v_1v_2 be an edge in G. Then denote by V_1 (resp. V_2) the set of neighbors of v_1 (resp. v_2). Since G is triangle-free, there is no edge between any pair of vertices in V_i, for $i = 1, 2$. Also, $d(u) + d(v) = n$ for any edge in G, thus G is a complete bipartite graph.

Conversely, as seen above, for any complete bipartite graph $Szs^* = \sqrt{m}$. ■

¿From this theorem and the fact that a complete bipartite graph $K_{p,q}$ contains $m = pq$ edges, it is easy to derive the following corollary.

Corollary 7. *Let G be a connected triangle-free graph on $n \geq 3$ vertices with minimum degree δ, then*

$$Szs^*(G) \geq \sqrt{\delta(n - \delta)}$$

with equality if and only if G is the complete bipartite graph $K_{n-\delta,\delta}$.

Note the similarity between the above corollary and the corresponding result about the Randić index proved in [8] and stated as follows.

Theorem 8 ([8, 12]). *Let $G = (V, E)$ be a triangle-free graph of order n with minimum degree $\delta \geq 1$, then*

$$Ra(G) \geq \sqrt{\delta(n - \delta)}$$

with equality if and only if G is the complete bipartite graph $K_{n-\delta,\delta}$.

Liu, Lu and Tian [13] showed that there was a mistake in the proof of the above theorem given by Delorme, Favaron and Rautenbach [8]. A correct proof was recently given by Li and Liu [12].

4. Linear regression

In this section, we report on the correlations between Szs^* and some selected graph invariants. Two types of regression are studied: linear and logarithmic. Each type was studied for different sets of alkanes. First, the two regressions were tested for all alkanes on up to 8 vertices. Among these, the acyclic alkanes were tested alone, and then all cycloalkanes were considered. Finally, the correlations were tested for each of the sets of unicyclic and bicyclic alkanes. The data is given in Table 1 and Table 2, and the graphs corresponding to the chemical compounds considered are given in Figure 2 and

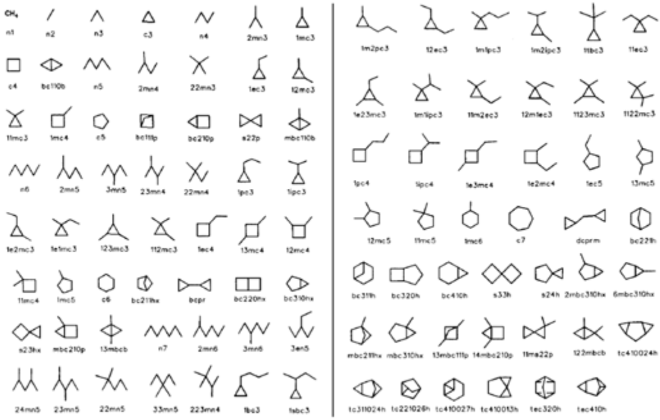

Figure 2: All alkanes on up to 7 vertices

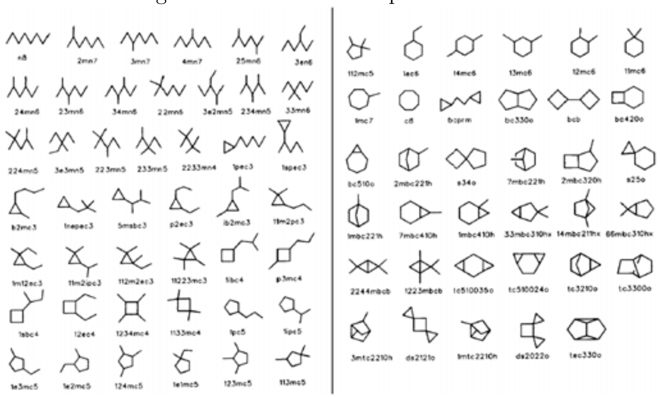

Figure 3: All alkanes on 8 vertices

Figure 3. Note that the graphs with the corresponding name (in chemistry) and boiling points are taken from [20]. The values of Ra, Sz, Sz^* and Szs^* are computed using the AutoGraphiX system [1, 7]. The results are summarized in Table 3.

The first observation is that the Randić index is the descriptor having the best correlation, in both linear and logarithmic models, with the boiling point. For the linear model, the BP–Ra correlation ranges between $R^2 = 0.9601$ for the cycloalkanes and $R^2 = 0.9768$ for trees. For the logarithmic model, the BP–Ra correlation ranges between $R^2 = 0.9520$ for all the cycloalkanes, and also for the *bicycloalkanes*, and $R^2 = 0.9895$ for the trees. Thus, when going from the linear to logarithmic model, the correlation increases slightly within

Name	BP	Randic	Sz	Szs	Szs*	Name	BP	Randic	Sz	Szs	Szs*
n1	-161,5	0	0	0	0	23mn5	89,8	3,180739	46	46	2,768875
n2	-88,6	1	1	1	1	22mn5	79,2	3,06066	46	46	2,768875
n3	-42,1	1,414214	4	4	1,414214	33mn5	86,1	3,12132	44	44	2,708013
c3	-32,8	1,5	3	6,25	1,5	223mn4	80,9	2,943376	42	42	2,645751
n4	-0,5	1,914214	10	10	1,825742	1bc3	98	3,431852	51	68,75	3,133916
2mn3	-11,7	1,732051	9	9	1,732051	1sbc3	90,3	3,342535	45	62,75	2,994042
1mc3	0,7	1,893847	8	14,5	1,903943	1m2pc3	93	3,342535	48	62,25	3,099539
c4	12,6	2	16	16	2	12ec3	90	3,38054	47	66,75	3,087995
bc110b	8	1,966321	9	19	1,949359	1m1pc3	84,9	3,267767	45	62,75	2,994042
n5	36	2,414214	20	20	2,236068	1m2ipc3	81,1	3,215214	44	63,25	3,005946
2mn4	27,8	2,270056	18	18	2,12132	1tbc3	80,5	3,105172	41	58,75	2,897043
22mn3	9,5	2	16	16	2	11ec3	88,6	3,328427	43	60,75	2,945942
1ec3	35,9	2,431852	17	26,75	2,313007	1e23mc3	91	3,270056	44	64,25	3,029616
12mc3	32,6	2,30453	16	26,25	2,291288	1m1ipc3	81,5	3,150482	41	58,75	2,897043
11mc3	20,6	2,207107	15	24,75	2,22486	11m2ec3	79,1	3,165832	43	62,75	2,994042
1mc4	36,3	2,393847	28	28	2,366432	12m1ec3	85,2	3,188487	42	61,25	2,95804
c5	49,3	2,5	20	31,25	2,5	1123mc3	78	3,065384	40	60,25	2,933793
bc111p	36	2,44949	36	36	2,44949	1122mc3	76	2,957107	39	58,75	2,897043
bc210p	46	2,466326	24	36,5	2,466441	1pc4	100,7	3,431852	68	68	3,116775
s22p	39	2,414214	14	33,5	2,362908	1ipc4	92,7	3,30453	64	64	3,023716
mbc110b	33,5	2,312278	16	32,5	2,327373	1e3mc4	89,5	3,325699	70	70	3,162278
n6	68,7	2,914214	35	35	2,645751	1e2mc4	94	3,342535	66	66	3,070598
2mn5	60,3	2,770056	32	32	2,529822	1ec5	103,5	3,431852	52	73,25	3,234855
3mn5	63,3	2,80806	31	31	2,48998	13mc5	91,3	3,287694	51	72,75	3,223795
23mn4	58	2,642734	29	29	2,408319	12mc5	95,6	3,30453	49	70,75	3,179173
22mn4	49,7	2,56066	28	28	2,366432	11mc5	87,9	3,207107	48	69,25	3,145291
1pc3	69	2,931852	31	44,5	2,723356	1mc6	101	3,393847	78	78	3,338092
1ipc3	58,3	2,80453	28	41,5	2,629956	c7	118,4	3,5	63	85,75	3,5
1e2mc3	63	2,842535	29	43,5	2,692582	dcprm	102	3,44949	46	81,5	3,191786
1e1mc3	57	2,767767	27	40,5	2,598076	bc221h	105,5	3,44949	62	93,5	3,418699
123mc3	63	2,732051	27	42	2,645751	bc311h	110	3,44949	96	96	3,464102
112mc3	52,6	2,627827	26	40,5	2,598076	bc320h	110,5	3,466326	66	91	3,372684
1ec4	70,7	2,931852	45	45	2,738613	bc410h	116	3,466326	72	95	3,446012
13mc4	59	2,787694	46	46	2,768875	s33h	96,5	3,414214	80	80	3,162278
12mc4	62	2,80453	44	44	2,708013	s24h	98,5	3,414214	47	86	3,278719
11mc4	53,6	2,707107	42	42	2,645751	2mbc310hx	100	3,37701	51	87	3,297726
1mc5	51,8	2,893847	33	49	2,857738	6mbc310hx	103	3,393847	54	87,25	3,302461
c6	80,7	3	54	54	3	mbc211hx	81,5	3,285405	60	88,25	3,321333
bc211hx	71	2,94949	40	61	2,964071	mbc310hx	92	3,312278	49	85	3,259601
bcpr	76	2,966326	27	54	2,77746	13mbc111p	71,5	3,12132	84	84	3,24037
bc220hx	83	2,966326	59	59	2,9032	14mbc210p	74	3,164214	56	79	3,142451
bc310hx	81	2,966326	34	60	2,9277	11ms22p	78	3,164214	38	75,5	3,072051
s23hx	69,5	2,914214	41	54,5	2,790289	122mbcb	84	3,089152	44	72,5	3,010399
mbc210p	60,5	2,812278	38	55,25	2,809423	tc410024h	105	3,44949	53	103,25	3,387067
13mbcb	55	2,664214	26	51	2,699206	tc311024h	107	3,44949	62	101,75	3,362374
n7	98,5	3,414214	56	56	3,05505	tc221026h	106	3,44949	48	108,75	3,476109
2mn6	90	3,270056	52	52	2,94392	tc410027h	110	3,483163	85	109,25	3,484091
3mn6	92	3,30806	50	50	2,886751	tc410013h	107,5	3,41745	51	101,25	3,354102
3en5	93,5	3,346065	48	48	2,828427	tec320h	108,5	3,483163	54	122	3,49285
24mn5	80,5	3,125898	48	48	2,288427	tec410h	104	3,483163	55	117	3,420526

Table 1: Data for alkanes on up to 7 vertices

Name	BP	Randic	Sz	Szs	Szs*	Name	BP	Randic	Sz	Szs	Szs*
n8	125,7	3,914214	84	84	3,464102	124mc5	115	3,698377	72	100,5	3,544362
2mn7	117,6	3,770056	79	79	3,359422	1e1mc5	121,5	3,767767	70	97	3,482097
3mn7	118,9	3,80806	76	76	3,295018	123mc5	117	3,715214	70	98,5	3,508917
4mn7	117,7	3,80806	75	75	3,273268	113mc5	104,5	3,600954	71	99	3,517812
25mn6	109,1	3,625898	74	74	3,251373	112mc5	114	3,627827	67	95	3,446012
3en6	118,5	3,846065	72	72	3,207135	1ec6	131,8	3,931852	109	109	3,691206
24mn6	109,4	3,663902	71	71	3,184785	14mc6	121,8	3,787694	110	110	3,708099
23mn6	115,6	3,680739	70	70	3,162278	13mc6	122,3	3,787694	108	108	3,674235
34mn6	117,7	3,718744	68	68	3,116775	12mc6	126,6	3,80453	106	106	3,640055
22mn6	106,8	3,56066	71	71	3,184785	11mc6	119,5	3,707107	104	104	3,605551
3e2mn5	115,6	3,718744	67	67	3,093773	1mc7	134	3,893847	88	117,5	3,832427
234mn5	113,5	3,553418	65	65	3,047247	c8	149	4	128	128	4
33mn6	112	3,62132	67	67	3,093773	bcprm	129	3,94949	72	117	3,605551
224mn5	99,2	3,416502	66	66	3,070598	bcp330o	137	3,966326	81	131	3,815174
3e3mn5	118,2	3,681981	64	64	3,023716	bcb	136	3,966326	112	112	3,527668
223mn5	109,8	3,48138	63	63	3	bc420o	133	3,966326	132	132	3,829708
233mn5	114,8	3,504036	62	62	2,976095	bc510o	141	3,966326	89	138	3,91578
2233mn4	106,5	3,25	58	58	2,878492	2mbc221h	125	3,860173	89	128,25	3,774917
1pec3	128	3,931852	78	100,5	3,544362	s34o	128	3,914214	92	119	3,636237
1spec3	117,7	3,842535	69	91,5	3,381937	7mbc221h	128	3,87701	83	126	3,741657
b2mc3	124	3,842535	74	98,5	3,508917	2mbc320h	130,5	3,87701	92	124,25	3,715583
1nepec3	106	3,578298	65	87,5	3,307189	s25o	125	3,914214	103	125,5	3,734226
5msbc3	115,5	3,715214	64	86,5	3,288237	1mbc221h	117	3,785405	84	127	3,756476
1e2pc3	108	3,88054	72	97,5	3,49106	7mbc410h	138	3,893847	97	130,5	3,807887
ib2mc3	110	3,698377	69	93,5	3,418699	1mbc410h	125	3,812278	98	127,25	3,760171
11m2pc3	105,9	3,665832	67	92,5	3,400368	33mbc310hx	115	3,673433	76	117	3,605551
1m12ec3	108,9	3,726492	64	89,5	3,344772	14mbc211hx	91	3,62132	87	123	3,696846
11m2ipc3	94,4	3,538511	62	87,5	3,307189	66mbc310hx	126,1	3,72718	76	117	3,605551
112m2ec3	104,5	3,517767	59	84,5	3,25	2244mbcb	104	3,488034	77	107	3,448027
11223mc3	100,5	3,404701	56	82,5	3,211308	1223mbcb	105	3,457107	60	100	3,333333
libc4	120,1	3,787694	93	93	3,409545	tc510035o	142	3,932653	98	151	3,885872
p3mc4	117,4	3,825699	101	101	3,553168	tc510024o	149	3,94949	94	152,5	3,905125
1sbc4	123	3,842535	90	90	3,354102	tc3210o	136	3,94949	86	148,5	3,85357
12ec4	119	3,88054	94	94	3,427827	tc3300o	125	3,966326	84	151	3,885872
1234mc4	114,5	3,642734	92	92	3,391165	3mtc2210h	120,5	3,87701	69	148	3,847077
1133mc4	86	3,414214	92	92	3,391165	ds2121o	103	3,828427	90	135	3,674235
1pc5	131	3,931852	78	105	3,622844	1mtc2210h	111	3,805478	68	147	3,834058
1ipc5	126,4	3,80453	73	100	3,535534	ds2022o	115	3,87132	82	127	3,563706
1e3mc5	121	3,825699	76	104	3,605551	tec330o	137,5	3,966326	99	174	3,977208
1e2mc5	124,7	3,842535	72	100	3,535534						

Table 2: Data for alkanes on 8 vertices

the class of trees, while it slightly decreases within the set of cyclic graphs. However, the difference does not appear to be significant.

The study confirms that the modified Szeged Szs index correlates better that the Szeged index Sz in the case of a linear model for all the graph classes considered, except of course for the class of trees for which $Szs = Sz$. In the case of the logarithmic model, when we considered all the graphs together, the correlation was slightly better for Sz than for Szs, $R^2 = 0.8967$ against $R^2 = 0.8804$. But, when we considered the class of unicyclic graphs and that of bicyclic graphs, each alone, the correlation was better for Szs. It was $R^2 = 0.8880$ and $R^2 = 0.7912$ for Sz against $R^2 = 0.9469$ and $R^2 = 0.9044$ for Szs in the case of unicyclic and bicyclic graphs respectively. In the case of the class of all cyclic graphs the correlation was $R^2 = 0.8462$ for Sz and $R^2 = 0.8986$ for Szs.

In almost all cases, the boiling point correlates with Szs^* better than with Sz or Szs. In fact the only case where Sz and Szs correlate with BP better than Szs^* is when the class of graphs contains only trees and the model is logarithmic. In that case, the correlation is $R^2 = 0.9686$ for Sz and Szs and $R^2 = 0.9634$ for Szs^*. Globally, the correlation between BP and Szs^* ranges from $R^2 = 0.9058$, for the class of all graphs and the linear model, to $R^2 = 0.9634$, for the class of trees and the logarithmic model.

Comparing the spread of the correlation coefficients for different parameters among the two models gathered, we find that the Randić index has the smallest spread with 0.0375. The second smallest spread corresponds to the normalized revised Szeged index with 0.0576, while the spreads corresponding to Sz and Szs, 0.2794 and 0.3075 respectively, are very large compared to those of Ra and Szs^*.

The spreads of the parameters are the same if we consider only the linear model while their values decrease. Actually the values are 0.0167, 0.0517, 0.179 and 0.2475 for Ra, Szs^*, Sz and Szs respectively. The order changes in the case of a logarithmic model, where the values are 0.0375, 0.055, 0.0882 and 0.1774 for Ra, Szs^*, Szs and Sz.

If we compare the linear and logarithmic models, the logarithmic formula seems to fit better than the linear one for both Szeged and modified Szeged indices. Indeed, for all graph classes considered in the present study, the correlation between BP and Sz, and between BP and Szs is better in the logarithmic model than in the linear model. Also, in the case of trees, the logarithmic model appears to be better than the linear model.

	$Bp = a \cdot X + b$				$Bp = a \cdot \ln(X) + b$			
X	Ra	Sz	Szs	Szs^*	Ra	Sz	Szs	Szs^*
	Linear and logarithmic regressions for all listed graphs							
a	65.7605	1.2979	0.9560	68.2475	173.5428	52.1224	51.2689	176.4932
b	-125.7588	14.1150	17.2091	-120.9368	-112.3499	-112.7021	-122.7882	-105.5079
R^2	0.9681	0.6892	0.6610	0.9058	0.9649	0.8967	0.8804	0..9124
	Linear and logarithmic regressions for all listed trees							
a	67.4928	1.9754	1.9754	85.6040	163.9879	53.3731	53.3731	188.8712
b	-130.4607	-18.4448	-18.4448	-154.7737	-100.8456	-116.7670	-116.7670	-105.6717
R^2	0.9768	0.8682	0.8682	0.9513	0.9895	0.9686	0.9686	0..9634
	Linear and logarithmic regressions for all listed cyclic graphs							
a	61.7626	1.0503	0.8561	65.7610	182.2491	51.4306	58.9833	190.7204
b	-112.2505	30.4603	22.8883	-116.4228	-123.0777	-110.0835	-160.0415	-125.3119
R^2	0.9601	0.7404	0.8006	0.9280	0.9520	0.8462	0.8986	0.9315
	Linear and logarithmic regressions for all listed unicyclic graphs							
a	61.8049	1.1007	1.1673	70.2910	177.0823	51.0513	61.9144	194.5201
b	-113.4510	26.3344	6.1250	-128.1330	-118.1771	-109.7664	-168.0507	-127.7233
R^2	0.9728	0.7615	0.9085	0.9575	0.9648	0.8880	0.9469	0.9561
	Linear and logarithmic regressions for all listed bicyclic graphs							
a	62.1103	0.9832	0.9220	65.6919	190.3087	50.5152	67.4292	196.8301
b	-111.7343	33.4759	12.9827	-118.1816	-131.2320	-106.9898	-201.3849	-134.2765
R^2	0.9618	0.7341	0.8889	0.9151	0.9520	0.7912	0.9044	0.9084

Table 3: Linear and logarithmic regressions

References

[1] M. Aouchiche, J. M. Bonnefoy, A. Fidahoussen, G. Caporossi, P. Hansen, L. Hiesse, J. Lacheré, A. Monhait, Variable neighborhood search for extremal graphs. 14. The AutoGraphiX 2 system, in: L. Liberti, N. Maculan (Eds.), *Global Optimization: From Theory to Implementation*, Springer, New York, 2006, pp. 281–310.

[2] M. Aouchiche, G. Caporossi, P. Hansen, T. Pisanski, Notes about the Szeged, Szeged star and Wiener indices in graphs, in preparation.

[3] M. Aouchiche, G. Caporossi, P. Hansen, Variable neighborhood search for extremal graphs. VIII. Variations on Graffiti 105, *Congr. Numer.* **148** (2001) 29–144.

[4] M. Aouchiche, P. Hansen, Automated results and conjectures on average distance in graphs, in: A. Bondy, J. Fonlupt, J. L. Fouquet, J. C. Fournier, J. L. Ramírez Alfonsín (Eds.), *Graph Theory in Paris*, Birkhäuser, Basel, 2007, pp. 21–36.

[5] A. T. Balaban, Topological indexes and their uses – A new approach for the coding of alkanes, *J. Mol. Struct.* **165** (1988) 243–253.

[6] B. Bollobas, P. Erdős, Graphs of extremal weights, *Ars Comb.* **50** (1998) 225–233.

[7] G. Caporossi, P. Hansen, Variable neighborhood search for extremal graphs: I. The AutoGraphiX system, *Discr. Math.* **212** (2000) 29–44.

[8] C. Delorme, O. Favaron, D. Rautenbach, On the Randić index, *Discr. Math.* **257** (2002) 29–38.

[9] I. Gutman, A formula for the Wiener number of trees and its extension to graphs containing cycles, *Graph Theory Notes New York* **27** (1994) 9–15.

[10] H. Hosoya, Topological index. A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons, *Bull. Chem. Soc. Jpn.* **44** (1971) 2332–2339.

[11] X. Li, I. Gutman, *Mathematical Aspects of Randić-Type Molecular Structure Descriptors*, Univ. Kragujevac, Kragujevac, 2006.

[12] X. Li, J. Liu, Complete solution to a conjecture on the Randić index, *Discr. Math.* **309** (2009) 6322–6324.

[13] H. Liu, M. Lu, F. Tian, On the Randić index, *J. Math. Chem.* **38** (2005) 345–354.

[14] T. Pisanski, M. Randić, Use of the Szeged index and the revised Szeged index for measuring network bipartivity, *Discr. Appl. Math.* **158** (2010) 1936–1944.

[15] T. Pisanski, J. Žerovnik, Edge-contributions of some topological indices and arboreality of molecular graphs, *Ars Math. Contemp.* **2** (2009) 49–58.

[16] M. Randić, On generalization of Wiener index for cyclic structures, *Acta Chim. Slov.* **49** (2002) 483–496.

[17] M. Randić, Topological indices, in: P. von Rague Schleyer (Ed.), *Encyclopedia of Computational Chemistry*, Wiley, London, 1998, pp. 3018–3032.

[18] M. Randić, Generalized molecular descriptors, *J. Math. Chem.* **7** (1991) 155–168.

[19] M. Randić, On characterization of molecular branching, *J. Amer. Chem. Soc.* **97** (1975) 6609–6615.

[20] G. Rücker, C. Rücker, On topological indices, boiling points, and cycloalkanes, *J. Chem. Inf. Comput. Sci.* **39** (1999) 788–802.

[21] R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley–VCH, Weinheim, 2000.

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

[23] P. Yu, An upper bound on the Randić of trees, *J. Math. Study* **31** (1998) 225–230.