MATCH
Communications in Mathematical and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. **84** (2020) 483-497

ISSN 0340 - 6253

Adjusting Geometric–Arithmetic Index to Estimate Boiling Point

Mustapha Aouchiche*, **Vidya Ganesan**

Mathematical Sciences Department, COS, UAEU
Al Ain, UAE
maouchiche@uaeu.ac.ae, vidya_g@uaeu.ac.ae

(Received December 12, 2019)

Abstract

In chemical graph theory, many graph parameters, or topological indices, were proposed as estimates 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 geometric-arithmetic index to get a better estimate of the boiling point. We compare the correlation between the boiling points and both versions of the geometric-arithmetic index using different regression models.

1 Introduction

Mathematical descriptors of molecular structure and properties, such as various topological indices [20], have been widely used in chemical studies. They play a very important role in mathematical chemistry especially in QSAR (quantitative structure-activity relationship) and/or QSPR (quantitative structure-property relationship) related studies. Among those descriptors, a special interest is devoted to so-called topological indices. Many topological indices related to the graph representation of molecular structures were proposed. They are used to understand physicochemical properties of chemical compounds in a simple way, since they sum up some of the properties of a molecule in a single number. During the last decades, a legion of topological indices were introduced and found some applications in chemistry, see *e.g.*, [12, 13, 25].

*Corresponding author

The study of topological indices goes back to the seminal work by Wiener [27] in which he used the sum of all shortest-path distances, nowadays known as the *Wiener index*, of a (molecular) graph for modeling physical properties of alkanes.

Another very important molecular descriptor, was introduced by Randić [19]. It is called the *Randić (connectivity) index* and defined as

$$Ra = Ra(G) = \sum_{uv \in E} \frac{1}{\sqrt{d_u d_v}}$$

where d_u denotes the degree (number of neighbors) of u in G. The Randić index is probably the most studied molecular descriptor in mathematical chemistry. Actually, there are more than two thousand papers and five books devoted to this index (see, *e.g.*, [11, 15–18] and the references therein).

Among other important topological indices, we can cite the *Hosoya topological index* [14] introduced in 1971, the *Szeged index* [10] inroduced in 1994, and the *revised Wiener index* (also called the *revised Szeged index*) by Randić [21] in 2002.

Motivated by the definition of Randić connectivity index, Vukičević and Furtula [26] proposed the *geometric-arithmetic index*. It is so-called since its definition involves both the geometric and the arithmetic means of the endpoints degrees of the edges in a graph. For a simple graph G with edge set $E(G)$, the geometric-arithmetic index $GA(G)$ of a graph G is defined as in [26] by

$$GA(G) = \sum_{uv \in E(G)} \frac{2\sqrt{d_u d_v}}{d_u + d_v} \, .$$

where d_u denotes the degree of u in G.

It is noted in [26] that the predictive power of GA for physico-chemical properties is somewhat better than the predictive power of the Randić connectivity index. In [26], Vukičević and Furtula gave the lower and upper bounds for GA, identified the trees with the minimum and the maximum GA indices, which are the star S_n and the path P_n, respectively. In [28] Yuan, Zhou and Trinajsić gave the lower and upper bounds for GA index of molecular graphs using the numbers of vertices and edges. They also determined the n-vertex molecular trees with the minimum, the second, and the third minimum, as well as the second and the third maximum GA indices. Lower and upper bound on the geometric-arithmetic index in terms of order n, size m, minimum degree δ and/or maximum degree were proved in [22]. Also in [22], GA was compared to other well known topological indices such as the Randić index, the first and second Zagreb indices, the harmonic index and the sum connectivity index. Other lower and upper bounds, on the geometric-arithmetic index, involving the order n the size m, the minimum and the maximum degrees and the second Zagreb index were proved in [6]. In [1], several bounds and comparisons, involving the geometric-arithmetic index and several other graph parameters, were proved. The problem of lower bounding GA over the class of connected graphs with fixed number and minimum degree was discussed in [2, 8, 23]. A comparison between GA and the spectral radius (the largest adjacency eigenvalue) of a connected graph was done in [4]. The chemical applicability of the geometric-arithmetic index was highlighted in [7, 9, 26].

In this paper, we are interested in the study of the correlation between the boiling point, as a molecular property, and an *adjusted version* of the geometric-arithmetic index, as a topological descriptor. To carry out this study, we considered set of data consisting of experimental boiling points of selected saturated hydrocarbons, taken from [24] (see also [3]). We used AutoGraphiX III [5] (available at *https://www.gerad.ca/∼gillesc/*) to compute the values of the topological descriptors of the corresponding molecular graphs.

In the next section, we present a comparison of the geometric-arithmetic index with minimum, maximum and average degrees of a connected graph. Those results, particularly the bounds on the ratio of the geometric-arithmetic index to the average degree, showed similarities with variations in the Randić index. The observed similarities inspired an adjustment in the geometric-arithmetic index. The adjusted topological index is the subject of a statistical study presented in Section 3.

2 Theoretical aspects

We begin this section by recalling some definitions. In this paper, we consider only simple, undirected and finite graphs, *i.e*, undirected graphs on a finite number of vertices without multiple edges or loops. A graph is (usually) denoted by $G = G(V, E)$, where V is its vertex set and E its edge set. The *order* of G is the number $n = |V|$ of its vertices, and its *size* is the number $m = |E|$ of its edges. For two vertices u and v $(u, v \in V)$, if $uv \in E$, we say u and v are *adjacent* in G. The *degree* of a vertex u, denoted d_u, is the number of vertices adjacent to it in G. A graph G is said to be *regular* of degree d, or d*-regular* if $d_u = d$ for every vertex u in G. The minimum, average and maximum degrees in a graph G are denoted by δ, $\overline{d}$ and Δ, respectively.

As usual, we denote by S_n the star and by K_n the complete graph, each on n vertices.

In the next theorem, we prove a lower and an upper bound on the ratio $GA/\overline{d}$. We also characterize the corresponding extremal graphs, in both cases.

Theorem 2.1. *For any connected graph on $n \geq 3$ with average degree $\overline{d}$ and geometric-arithmetic index GA,*

$$\sqrt{n-1} \leq \frac{GA}{\overline{d}} \leq \frac{n}{2}$$

with equality if and only if G is the star S_n (resp. regular) for the lower (resp. upper) bound.

Proof :

For the lower bound and assuming, without loss of generality, that $d_i \leq d_j$, we have

$$\frac{GA}{\overline{d}} = \frac{\sum_{ij\in E} \frac{2\sqrt{d_i d_j}}{d_i+d_j}}{\frac{2}{n}\sum_{ij\in E} 1} = \frac{\sum_{ij\in E} \frac{2\sqrt{d_i/d_j}}{d_i/d_j+1}}{\frac{2}{n}\sum_{ij\in E} 1} \geq \frac{\sum_{ij\in E} \frac{2\sqrt{1/(n-1)}}{1/(n-1)+1}}{\frac{2}{n}\sum_{ij\in E} 1} = \sqrt{n-1}\ .$$

Equality being reached if and only if $d_i = 1$ and $d_j = n-1$ for all edges $ij \in E$, *i.e.*, if and only if G is the star S_n.

For the upper bound, we have

$$\frac{GA}{\overline{d}} = \frac{\sum_{ij\in E} \frac{2\sqrt{d_i d_j}}{d_i+d_j}}{\frac{2}{n}\sum_{ij\in E} 1} \leq \frac{\sum_{ij\in E} 1}{\frac{2}{n}\sum_{ij\in E} 1} = \frac{n}{2}\ .$$

Equality being reached if and only if $d_i = d_j$ for all edges $ij \in E$, *i.e.*, if and only if G is regular. ■

In the next theorem, we prove a lower and an upper bounds on the ratio GA/Δ. We also characterize the corresponding extremal graphs, in both cases.

Theorem 2.2. *For any connected graph on $n \geq 3$ with maximum degree Δ and geometric-arithmetic index GA,*

$$\frac{2\sqrt{n-1}}{n} \leq \frac{GA}{\Delta} \leq \frac{n}{2}$$

with equality if and only if G is the star S_n (resp. regular) for the lower (resp. upper) bound.

Proof :

For the lower bound, it is well-known that the minimum value of GA over all connected graphs on n vertices is reached only for the star S_n, which also maximizes Δ.

The upper bound, as well as the characterization of the extremal graphs, follows immediately from the corresponding case in Theorem 2.1. ■

Among the the results proved in [1], we recall the following theorem.

Theorem 2.3 ([1]). *For any connected graph G with minimum degree $\delta \geq 2$*

$$\frac{GA}{Ra} \leq n - 1$$

with equality if and only if G is the complete graph K_n.

Experiments with the help of AutoGraphiX led to a conjecture, improving the above theorem, proved in the next proposition.

Proposition 2.4. *For any connected graph G with geometric-arithmetic index GA, Randić index Ra and maximum degree Δ*

$$\frac{GA}{Ra} \leq \Delta$$

with equality if and only if G is Δ-regular.

Proof: It is well-known that $GA \leq m$ with equality if and only G is regular. In addition,

$$Ra = \sum_{ij \in E} \frac{1}{\sqrt{d_i d_j}} \geq \sum_{ij \in E} \frac{1}{\Delta} = \frac{m}{\Delta}$$

with equality if and only if G is Δ-regular.

Combining the above inequalities, we get

$$\frac{GA}{Ra} \leq \frac{m\Delta}{m} = \Delta$$

with equality if and only if G is Δ-regular. ■

Also, among the results proved in [1], we recall the following.

Theorem 2.5 ([1]). *For any connected graph with minimum degree $\delta \geq 2$*

$$GA \geq \delta Ra$$

with equality if and only if G is δ-regular.

To sum up the latter results we have a chain of inqualities which we state in the following theorem.

Theorem 2.6. *For any connected graph on $n \geq 2$ vertices with m edges, we have*

$$\delta \cdot Ra \leq GA \leq m \leq \Delta \cdot Ra \leq (n-1) \cdot Ra \ .$$

Furthermore all equalities hold simultaneously if and only if G is the complete graph K_n; and all equalities, but the last one, hold simultaneously if and only if G is a non-complete Δ-regular graph.

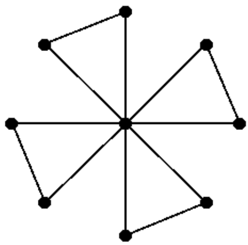

Figure 1. Example of a graph with $GA < \overline{d} \cdot Ra$.

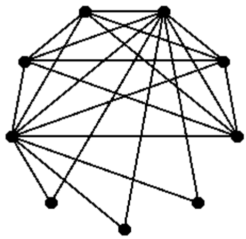

Figure 2. Example of a graph with $GA > \overline{d} \cdot Ra$.

Note that we can not insert $\overline{d} \cdot Ra$ into the chain since there exist graphs with $GA < \overline{d} \cdot Ra$ (see Figure 1 for an example) and others with $GA > \overline{d} \cdot Ra$ (see Figure 2 for an example).

3 Computational study

Let us define the *adjusted* geometric-arithmetic index to be $GA^* = GA/\overline{d}$. With this notation, Theorem 2.1 can be stated as:

$$\sqrt{n-1} \leq GA^* \leq \frac{n}{2}$$

with equality if and only if G is the star S_n (resp. regular) for the lower (resp. upper) bound.

This result reminds of a well-known result about the Randić index:

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

with equality if and only if G is the star S_n (resp. regular) for the lower (resp. upper) bound.

The similarities between both results, for GA^* and for Ra, and the fact that the Randić index [19] was first used to estimate the boiling point motivated a computational comparison between GA, GA^*, Ra and the boiling point (BP). In this section, we report on the results of the study.

We started our study using the two usual models of regression: linear and logarithmic. In addition to considering all the data gathered in one set, each model was studied for different sets of alkanes according the number of cycles in the graphs: trees (acyclic graphs), unicyclic graphs, bicyclic graphs, and graphs with at least 3 cycles. The data for molecular graphs on up to 7 vertices is given in Table 1, on 8 vertices in Table 2, on 9 or 10 virtices in Table 3.

The first observation is that the Randić index is the descriptor having the best correlation with the boiling point in both linear and logarithmic models, except for the set of trees where the GA^* has a slightly better correlation: $R^2 = 0.982173$ for GA^* versus $R^2 = 0.981315$ for Ra in the linear model; $R^2 = 0.975952$ for GA^* versus $R^2 = 0.97552$ for Ra in the logarithmic model.

In all combinations of sets (except unicyclic graphs for which $GA^* = GA/2$) and models, Ra and GA^* have a significantly better correlation than GA has.

The best correlation for each descriptor with both models was reached for the same set of data: the set of trees. Also, the least correlation was reach for the same set: graphs with at least three cycles. The fluctuation in the values of R^2 is the smallest for Ra followed closely by that of GA^*, and was the largest for GA, in both linear and logarithmic models. However, in the logarithmic regression, the fluctuation in R^2 was almost the for Ra and for GA^*.

For the linear model (see Table 4): (i) the correlation between BP and Ra ranges from $R^2 = 0.981315$ for trees (acyclic graphs) to $R^2 = 0.814856$ for cyclic graphs with at least three cycles, for a spread of 0.166459; (ii) the correlation between BP and GA ranges between $R^2 = 0.971045$ for trees and $R^2 = 0.662821$ for cyclic graphs with at least

Name	BP	Ra	GA	GA*	Name	BP	Ra	GA	GA*
n1	-161.5	0	0	0	23mn5	89.8	3.180739	5.52068	3.2204
n2	-88.6	1	1	1	22mn5	79.2	3.06066	5.28562	3.08328
n3	-42.1	1.414214	1.88562	1.41422	33mn5	86.1	3.12132	5.37124	3.13322
c3	-32.8	1.5	3	1	223mn4	80.9	2.943376	5.12179	2.98771
n4	-0.5	1.914214	2.88562	1.92375	1bc3	98	3.431852	6.8822	3.4411
2mn3	-11.7	1.732051	2.59808	1.73205	1sbc3	90.3	3.342535	6.74822	3.37411
1mc3	0.7	1.893847	3.82562	1.91281	1m2pc3	93	3.342535	6.74822	3.37411
c4	12.6	2	4	2	12ec3	90	3.38054	6.8048	3.4024
bc110b	8	1.966321	4.91918	1.96767	1m1pc3	84.9	3.267767	6.57124	3.28562
n5	36	2.414214	3.88562	2.42851	1m2ipc3	81.1	3.215214	6.55767	3.27884
2mn4	27.8	2.270056	3.65466	2.28416	1tbc3	80.5	3.105172	6.34934	3.17467
22mn3	9.5	2	3.2	2	11ec3	88.6	3.328427	6.65685	3.32843
1ec3	35.9	2.431852	4.8822	2.4411	1e23mc3	91	3.270056	6.65466	3.32733
12mc3	32.6	2.30453	4.69164	2.34582	1m1ipc3	81.5	3.150482	6.40741	3.20371
11mc3	20.6	2.207107	4.48562	2.24281	11m2ec3	79.1	3.165832	6.43495	3.21748
1mc4	36.3	2.393847	4.82562	2.41281	12m1ec3	85.2	3.188487	6.46399	3.23199
c5	49.3	2.5	5	2.5	1123mc3	78	3.065384	6.31154	3.15577
bc111p	36	2.44949	5.87878	2.44949	1122mc3	76	2.957107	6.08562	3.04281
bc210p	46	2.466326	5.91918	2.46633	1pc4	100.7	3.431852	6.8822	3.4411
s22p	39	2.414214	5.77124	2.40468	1ipc4	92.7	3.30453	6.69164	3.34582
mbc110b	33.5	2.312278	5.63495	2.3479	1e3mc4	89.5	3.325699	6.70781	3.35391
n6	68.7	2.914214	4.88562	2.93137	1e2mc4	94	3.342535	6.74822	3.37411
2mn5	60.3	2.770056	4.65466	2.79279	1ec5	103.5	3.431852	6.8822	3.4411
3mn5	63.3	2.80806	4.71124	2.82674	13mc5	91.3	3.287694	6.65123	3.32562
23mn4	58	2.642734	4.4641	2.67846	12mc5	95.6	3.30453	6.69164	3.34582
22mn4	49.7	2.56066	4.28562	2.57137	11mc5	87.9	3.207107	6.48562	3.24281
1pc3	69	2.931852	5.8822	2.9411	1mc6	101	3.393847	6.82562	3.412815
1ipc3	58.3	2.80453	5.69164	2.84582	c7	118.4	3.5	7	3.5
1e2mc3	63	2.842535	5.74822	2.87411	dcprm	102	3.44949	7.87878	3.44696
1e1mc3	57	2.767767	5.57124	2.78562	bc221h	105.5	3.44949	7.87878	3.44696
123mc3	63	2.732051	5.59808	2.79904	bc311h	110	3.44949	7.87878	3.44696
112mc3	52.6	2.627827	5.37837	2.68919	bc320h	110.5	3.466326	7.91918	3.46464
1ec4	70.7	2.931852	5.8822	2.9411	bc410h	116	3.466326	7.91918	3.46464
13mc4	59	2.787694	5.65123	2.82562	s33h	96.5	3.414214	7.77124	3.39992
12mc4	62	2.80453	5.69164	2.84582	s24h	98.5	3.414214	7.77124	3.39994
11mc4	53.6	2.707107	5.48562	2.74281	2mbc310hx	100	3.37701	7.78521	3.40603
1mc5	51.8	2.893847	5.82562	2.91281	6mbc310hx	103	3.393847	7.82562	3.42371
c6	80.7	3	6	3	mbc211hx	81.5	3.285405	7.56781	3.31092
bc211hx	71	2.94949	6.87878	2.94805	mbc310hx	92	3.312278	7.63495	3.34029
bcpr	76	2.966326	6.91918	2.96537	13mbc111p	71.5	3.12132	7.25685	3.17487
bc220hx	83	2.966326	6.91918	2.96537	14mbc210p	74	3.164214	7.37124	3.22492
bc310hx	81	2.966326	6.91918	2.96537	11ms22p	78	3.164214	7.37124	3.22494
s23hx	69.5	2.914214	6.77124	2.90196	122mbcb	84	3.089152	7.30209	3.19466
mbc210p	60.5	2.812278	6.63495	2.84355	tc410024h	105	3.44949	8.87878	3.45286
13mbcb	55	2.664214	6.37124	2.73053	tc310024h	107	3.44949	8.87878	3.45286
n7	98.5	3.414214	5.88562	3.43328	tc221026h	106	3.44949	8.87878	3.45286
2mn6	90	3.270056	5.65466	3.29855	tc410027h	110	3.483163	8.95959	3.48429
3mn6	92	3.30806	5.71124	3.33156	tc410013h	107.5	3.41745	8.78429	3.41611
3en5	93.5	3.346065	5.76781	3.36456	tec320h	108.5	3.483163	9.95959	3.48586
24mn5	80.5	3.125898	5.42369	3.16382	tec410h	104	3.483163	9.95959	3.48586

Table 1. Data for alkanes on up to 7 vertices

Name	BP	Ra	GA	GA*	Name	BP	Ra	GA	GA*
n8	125.7	3.914214	6.88562	3.93464	124mc5	115	3.698377	7.51726	3.758628
2mn7	117.6	3.770056	6.65466	3.80266	1e1mc5	121.5	3.767767	7.57124	3.78562
3mn7	118.9	3.80806	6.71124	3.834992	123mc5	117	3.715214	7.55767	3.778836
4mn7	117.7	3.80806	6.71124	3.834992	113mc5	104.5	3.600954	7.31124	3.655616
25mn6	109.1	3.625898	6.42369	3.67068	112mc5	114	3.627827	7.37837	3.689188
3en6	118.5	3.846065	6.76781	3.867324	1ec6	131.8	3.931852	7.8822	3.9411
24mn6	109.4	3.663902	6.48027	3.703012	14mc6	121.8	3.787694	7.65123	3.825616
23mn6	115.6	3.680739	6.52068	3.726104	13mc6	122.3	3.787694	7.65123	3.825616
34mn6	117.7	3.718744	6.57726	3.758436	12mc6	126.6	3.80453	7.69164	3.84582
22mn6	106.8	3.56066	6.28562	3.59178	11mc6	119.5	3.707107	7.48562	3.742808
3e2mn5	115.6	3.718744	6.57726	3.758436	1mc7	134	3.893847	7.82562	3.912808
234mn5	113.5	3.553418	6.33013	3.617216	c8	149	4	8	4
33mn6	112	3.62132	6.37124	3.640708	bcprm	129	3.94949	8.87878	3.946124
224mn5	99.2	3.416502	6.05466	3.459804	bcp330o	137	3.966326	8.91918	3.96408
3e3mn5	118.2	3.681981	6.45685	3.689632	bcb	136	3.966326	8.91918	3.96408
223mn5	109.8	3.48138	6.17837	3.5305	bc420o	133	3.966326	8.91918	3.96408
233mn5	114.8	3.504036	6.20741	3.547092	bc510o	141	3.966326	8.91918	3.96408
2233mn4	106.5	3.25	5.8	3.314284	2mbc221h	125	3.860173	8.7488	3.88658
1pec3	128	3.931852	7.8822	3.9411	s34o	128	3.914214	8.77124	3.898328
1spec3	117.7	3.842535	7.74822	3.874112	7mbc221h	128	3.87701	8.78521	3.904536
b2mc3	124	3.842535	7.74822	3.874112	2mbc320h	130.5	3.87701	8.78521	3.904536
1nepec3	106	3.578298	7.2822	3.6411	s25o	125	3.914214	8.77124	3.898328
5msbc3	115.5	3.715214	7.55767	3.778836	1mbc221h	117	3.785405	8.56781	3.807916
1e2pc3	108	3.88054	7.8048	3.9024	7mbc410h	138	3.893847	8.82562	3.922496
ib2mc3	110	3.698377	7.51726	3.758628	1mbc410h	125	3.812278	8.63495	3.837756
11m2pc3	105.9	3.665832	7.43495	3.717476	33mbc310hx	115	3.673433	8.4048	3.735468
1m12ec3	108.9	3.726492	7.52057	3.760284	14mbc211hx	91	3.62132	8.53908	3.795144
11m2ipc3	94.4	3.538511	7.2444	3.6222	66mbc310hx	126.1	3.72718	8.25685	3.669712
112m2ec3	104.5	3.517767	7.17124	3.58562	2244mbcb	104	3.488034	8.15897	3.626212
11223mc3	100.5	3.404701	7.04551	3.522748	1223mbcb	105	3.457107	8.08562	3.593608
libc4	120.1	3.787694	7.65123	3.825616	tc510035o	142	3.932653	9.83837	3.935348
p3mc4	117.4	3.825699	7.70781	3.853908	tc510024o	149	3.94949	9.87878	3.951512
1sbc4	123	3.842535	7.74822	3.874112	tc3210o	136	3.94949	9.87878	3.951512
12ec4	119	3.88054	7.8048	3.9024	tc3300o	125	3.966326	9.1918	3.967672
1234mc4	114.5	3.642734	7.4641	3.732052	3mtc2210h	120.5	3.87701	8.65123	3.844992
1133mc4	86	3.414214	6.97124	3.48562	ds2121o	103	3.828427	9.54247	3.816988
1pc5	131	3.931852	7.8822	3.9411	1mtc2210h	111	3.805478	8.46057	3.760252
1ipc5	126.4	3.80453	7.69164	3.84582	ds2022o	115	3.87132	9.65685	3.86274
1e3mc5	121	3.825699	7.70781	3.853908	tec330o	137.5	3.966326	10.9192	3.970612
1e2mc5	124.7	3.842535	7.74822	3.874112					

Table 2. Data for alkanes on 8 vertices

Name	BP	Ra	GA	GA*	Name	BP	Ra	GA	GA*
tc331037n	164.6	4.41582	10.798	4.417349	3cpbc410h	175.5	4.93265	11.8384	4.932655
tc421037n	162.5	4.43265	10.8384	4.433877	tc422025d	219	4.94949	11.8788	4.94949
tc421024n	166	4.43265	10.8384	4.433877	tc521026d	188	4.94949	11.8788	4.94949
tc430037n	161.4	4.44949	10.8788	4.45041	6mtc3220n	189.5	4.86017	11.7448	4.893665
ds2122n	142.5	4.32843	10.5425	4.312832	12cpr1mc3	158.3	4.80548	11.6213	4.8422
11cprc3	147.8	4.41745	10.7843	4.411755	bc310hxsc5	192.7	4.93429	11.8247	4.926955
acprnorb	140.7	4.46633	9.91918	4.463631	ds2024d	160	4.87132	11.6569	4.857025
3mtc3210o	161	4.37701	10.7852	4.412133	38mtc321024o	152.5	4.80453	11.6916	4.87152
ds2023n	147	4.37132	10.6569	4.359623	334mtc2210h	151	4.47296	10.101	4.591355
etc2210h	137.5	4.26679	9.54619	4.295786	133mtc2210h	143.5	4.46698	10.0805	4.58203
33mtc2210n	137.5	4.12103	9.3647	4.214115	177mtc2210h	153	4.483	10.1277	4.603505
17mtc2210h	131	4.14368	9.39373	4.227179	scptc3210o	174	4.89058	12.7171	4.89121
12mtc2210h	128	4.03541	9.16781	4.125515	tec52100d	155	4.82813	11.6503	4.854295
tec33100n	168.3	4.44949	11.8788	4.454541	pec530000d	171	4.96633	13.9192	4.971135
tec4300n	153	4.46633	11.9192	4.469693					

Table 3. Data for alkanes on 9 and on 10 vertices

	$Bp = a \cdot X + b$			$Bp = a \cdot \ln(X) + b$		
X	Ra	GA	GA^*	Ra	GA	GA^*
	Linear and logarithmic regressions for all listed graphs					
a	61.37184	18.94278	60.50903	185.2279	121.1098	177.2092
b	-112.612	-40.7462	-111.124	-125.577	-136.741	-116.87
R^2	0.959299	0.78842	0.955586	0.962236	0.856981	0.939629
	Linear and logarithmic regressions for all listed trees					
a	70.71719	36.41813	69.87619	178.2078	130.2453	176.7705
b	-140.977	-116.585	-140.266	-117.538	-132.2453	-117.395
R^2	0.981315	0.971045	0.982173	0.97552	0.973946	0.975952
	Linear and logarithmic regressions for all listed unicyclic graphs					
a	60.64729	30.10739	60.21479	182.2273	182.1453	182.1453
b	-109.444	-110.229	-110.23	-124.46	-252.646	-126.392
R^2	0.969396	0.965125	0.965125	0.960643	0.954635	0.954635
	Linear and logarithmic regressions for all listed bicyclic graphs					
a	61.63346	30.40864	60.80531	188.523	217.1552	186.4009
b	-110.349	-138.296	-108.968	-129.379	-345.3	-128.118
R^2	0.962126	0.947362	0.941305	0.953469	0.941851	0.932613
	Linear and logarithmic regressions for atleast three cycles					
a	50.25068	17.95326	48.95793	208.2723	192.4741	203.1522
b	-69.2629	-42.7935	-64.5593	-155.646	-305.458	-149.001
R^2	0.814856	0.662821	0.799135	0.80804	0.679743	0.792852
	Linear and logarithmic regressions for all listed cycloalkanes graphs					
a	57.05671	18.15172	56.95277	193.7489	147.9678	194.2417
b	-97.0496	-37.7162	-98.2496	-136.535	-196.108	-138.703
R^2	0.952701	0.820597	0.944569	0.949197	0.856507	0.939858

Table 4. Linear and logarithmic regressions

three cycles, with a spread of 0.308224; (iii) the correlation between BP and GA^* ranges between $R^2 = 0.982173$ for trees and $R^2 = 0.799135$ for cyclic graphs with at least three cycles, with a spread of 0.183038.

For the logarithmic model (see Table 4): (i) the correlation between BP and Ra ranges from $R^2 = 0.97552$ for trees (acyclic graphs) to $R^2 = 0.80804$ for cyclic graphs with at least three cycles, for a spread of 0.183038; (ii) the correlation between BP and GA ranges between $R^2 = 0.973946$ for trees and $R^2 = 0.679743$ for cyclic graphs with at least three cycles, with a spread of 0.294203; (iii) the correlation between BP and GA^* ranges between $R^2 = 0.9975952$ for trees and $R^2 = 0.792852$ for cyclic graphs with at least three cycles, with a spread of 0.1831.

The study confirms that the modified geometric index GA^*correlates better that the geometric index GA and significantly close to Randić index in the case of a linear model for all the graph classes under study.

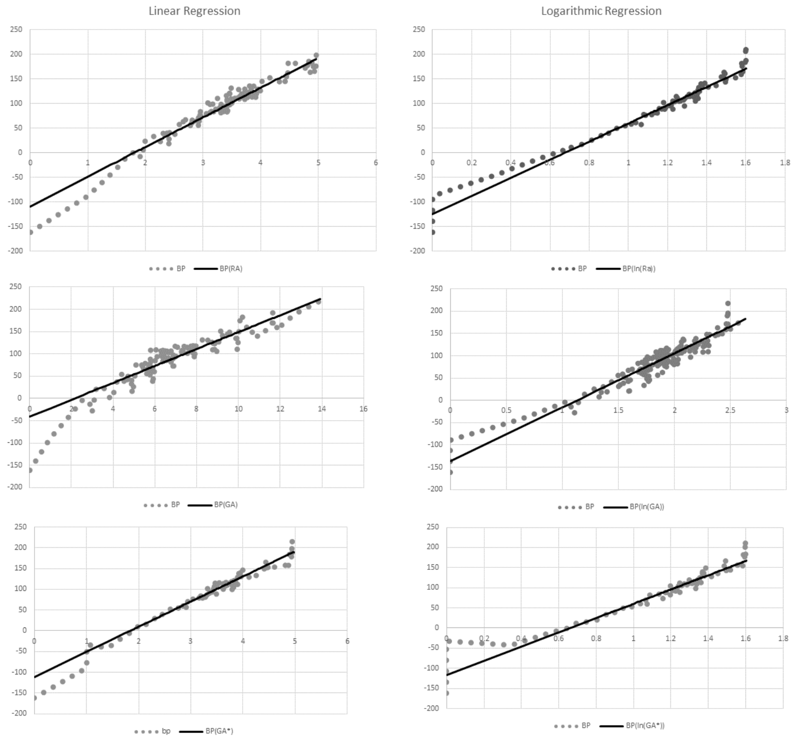

Figure 3. The linear and logarithmic regressions for Ra, GA and GA^* for all listed graphs.

When we consider all graphs together, the correlation was better for GA^* than for GA with both linear and logarithmic model: $R^2 = 0.955586$ versus $R^2 = 0.78842$ in the linear and $R^2 = 0.939629$ versus $R^2 = 0.856981$ in the logarithmic.

When we considered the different classes separately, except of case for the class of unicyclic graphs for which $GA^* = GA/2$ (in which case the correlation is the same), the correlation of the boiling point was always better with GA^* than with GA: $R^2 = 0.982173$ against $R^2 = 0.971045$ for the linear regression and $R^2 = 0.975952$ against $R^2 = 0.973946$ for the logarithmic, in case of of trees; $R^2 = 0.947362$ against $R^2 = 0.941305$ for the linear regression and $R^2 = 0.941851$ against $R^2 = 0.932613$ for the logarithmic, in case of of bicyclic graphs; $R^2 = 0.799135$ versus $R^2 = 0.662821$ in the linear model and $R^2 = 0.792852$ versus $R^2 = 0.679743$ in the logarithmic, in the case of graphs having more than two cycles.

If we compare the regression models: the linear model seems to fit better than the logarithmic one for modified geometric-arithmetic index GA^*; the logarithmic model works better for the geometric-arithmetic index GA except for the class of unicyclic graphs and that of bicyclic graphs.

Looking to the regression lines, both linear and logarithmic (see Figure 3), for the three indices and for all the listed graphs, we observe: (i) for the linear regression and for GA and GA^* the given boiling points are under the line for small and large values, while they are above the line for medium values; (ii) for the logarithmic regression and for GA and GA^* most of the given boiling points are above the line for small and large values, while they are under the line for most of the medium values. This suggests that we get a better estimate if use a regression model with a curve concave downward, but less concave than the logarithmic curve; (iii) the distribution of the boiling points around the regressions line balanced for Ra better than for GA and GA^*.

Following this observation, we suggest a model under the form $Y = aX^\alpha + b$, with $0 < \alpha < 1$. In our experiments, we considered $\alpha = 0.5 + 0.05k$, for $k = 0, \ldots, 9$. For each pf those values of α, the corresponding curve is concave downward, but not as concave as the logarithmic curve.

α	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60	0.55	0.50
	Randić Index									
a	67.19674	74.8656	83.54435	93.38707	104.5726	117.307	131.8241	148.3826	167.2555	188.7043
b	−118.4945	−128.502	−139.378	−151.221	−164.14	−178.248	−193.657	−210.465	−228.732	−248.435
R^2	0.933082	0.936452	0.93921	0.941522	0.942848	0.943007	0.941601	0.93809	0.931724	0.921472
	Geometric-Arithmetic Index									
a	22.44576	26.19183	30.62349	35.88273	42.14553	49.63105	58.61337	69.43627	82.53093	98.43521
b	−50.5286	−58.5721	−67.4075	−77.1533	−87.9508	−99.9683	−113.406	−128.501	−145.528	−164.794
R^2	0.82394	0.831963	0.8339924	0.847763	0.855391	0.862682	0.869454	0.875434	0.880215	0.883178
	Adjusted Geometric-Arithmetic Index									
a	98.43521	75.79759	84.5032	94.36253	105.5592	118.2972	132.8116	149.364	168.2372	189.706
b	−164.794	−132.824	−143.573	−155.271	−168.029	−181.961	−197.184	−213.805	−231.898	−251.467
R^2	0.98279	0.985274	0.987191	0.98838	0.988628	0.987644	0.985035	0.980262	0.972584	0.96097

Table 5. Results for the model $BP = a.X^\alpha + b$, with $\alpha = 0.5 + 0.05k$, for $k = 0, \ldots, 9$.

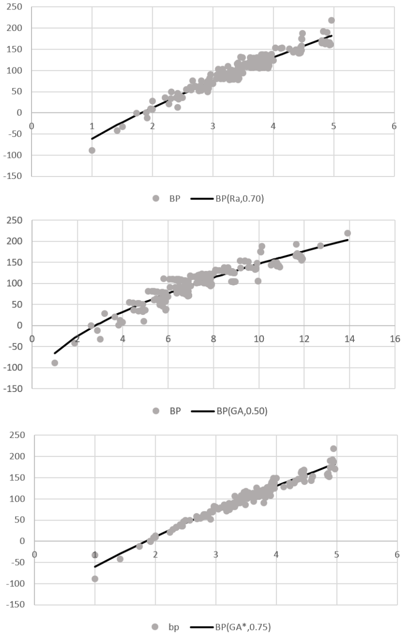

Figure 4. The best regression for Ra, GA and $GA*$, $BP = aY + b, Y = X^{\alpha}$.

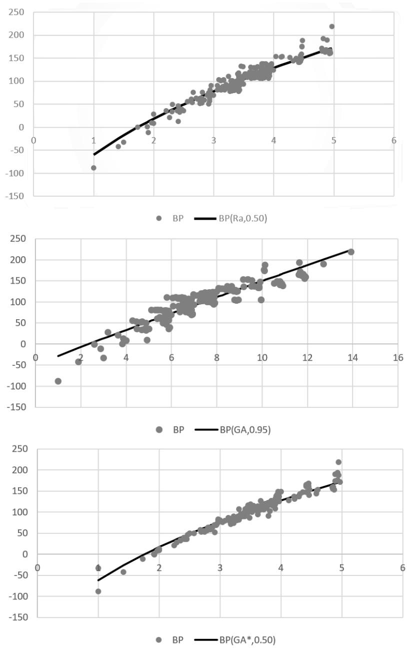

Figure 5. The worst regression for Ra, GA and $GA*$, $BP = aY + b, Y = X^{\alpha}$.

After the computation experiments (see Table 5), the first observation was that the model is worst than the linear model for the Randć index Ra for all values of α. For the geometric-arithmetic index GA, the correlation in the new model was significantly better than that in the linear model for all values of α, while compared to the logarithmic model, it improves for the values of α less than 0.75. For the modified geometric-arithmetic index GA^{*}, the model with α gives significantly better correlation than both linear and logarithmic ones and for all values of α. In addition, the correlation of GA^{*} with the boiling point BP in the $Y = aX^{\alpha} + b$ model is better then that of GA and Ra in any other model.

The best correlation (see Figure 4) for Ra, GA and GA^{*} in the $Y = aX^{\alpha} + b$ model was obtained for $\alpha = 0.7$, 0.5 and 0.75, respectively; while the worst (see Figure 5) was obtained for $\alpha = 0.5$, 0.95 and 0.5, respectively.

References

[1] A. Ali, A. A. Bhatti, Z. Raza, Further inequalities between vertex–degree–based topological indices, *Int. J. Appl. Comput. Math.* **3** (2017) 1921–1930.

[2] M. Aouchiche, I. El Hallaoui, P. Hansen, Geometric–arithmetic index and minimum degree of connected graphs, *MATCH Commun. Math. Comput. Chem.* **83** (2020) 179–188.

[3] M. Aouchiche, P. Hansen, The normalized revised Szeged index, *MATCH Commun. Math. Comput. Chem.* **67** (2012) 369–381.

[4] M. Aouchiche, P. Hansen, Comparing the geometric–arithmetic index and the spectral radius of graphs, *MATCH Commun. Math. Comput. Chem.* in press.

[5] G. Caporossi, Variable neighborhood search for extremal vertices: The AutoGraphiX-III system, *Comput. Oper. Res.* **78** (2017) 431–438.

[6] K. C. Das, I. Gutman, B. Furtula, On the first geometric–arithmetic index of graphs, *Discr. Appl. Math.* **159** (2011) 2030–2037.

[7] K. C. Das, I. Gutman, B. Furtula, Survey on geometric–arithmetic indices of graphs, *MATCH Commun. Math. Comput. Chem.* **65** (2011) 595–644.

[8] T. Divnić, M. Milivojević, L. Pavlović, Extremal graphs for the geometric–arithmetic index with given minimum degree, *Discr. Appl. Math.* **162** (2014) 386–390.

[9] B. Furtula, I. Gutman, Geometric–arithmetic indices, in: I. Gutman, B. Furtula (Eds.), *Novel Molecular Structure Descriptors – Theory and Applications*, Univ. Kragujevac, Kragujevac, 2010, pp. 137–172.

[10] 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.

[11] I. Gutman, B. Furtula (Eds.), *Recent Results in the Theory of Randić Index*, Univ. Kragujevac, Kragujevac, 2008.

[12] I. Gutman, B. Furtula (Eds.), *Novel Molecular Structure Descriptors – Theory and Applications I*, Univ. Kragujevac, Kragujevac, 2010.

[13] I. Gutman, B. Furtula (Eds.), *Novel Molecular Structure Descriptors – Theory and Applications II*, Univ. Kragujevac, Kragujevac, 2010.

[14] 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.

[15] L. B. Kier, L. H. Hall, *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York 1976.

[16] L. B. Kier, L. H. Hall, *Molecular Connectivity in Structural–Activity Analysis*, Wiley, New York, 1986.

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

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

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

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

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

[22] J. M. Rodríguez, J. M. Sigarreta, On the geometric–arithmetic index, *MATCH Commun. Math. Comput. Chem.* **74** (2015) 103–120

[23] M. Sohrabi–Haghighat, M. Rostami, Using linear programming to find the extremal graphs with minimum degree 1 with respect to geometric–arithmetic index, *Appl. math. Eng. Manag. Tech.* **3** (2015) 534–539.

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

[25] R. Todeschini, V. Consonni, *Molecular Descriptors for Chemoinformatics*, Wiley–VCH, Weinheim, 2009.

[26] D. Vukičević, B. Furtula, Topological index based on the ratios of geometrical and arithmetical means of end vertex degrees of edges, *J. Math. Chem.* **46** (2009) 1369–1376.

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

[28] Y. Yuan, B. Zhou, N. Trinajstić, On geometric–arithmetic index, *J. Math. Chem.* **47** (2010) 833–841.