

Estrada index of acyclic molecules

Ivan Gutman^{a,*}, Boris Furtula^a, Biljana Glišić^a, Violeta Marković^a & Aleksander Vesel^b

^aFaculty of Science, University of Kragujevac, P. O. Box 60, 34000 Kragujevac, Serbia

^bDepartment of Mathematics, PEF, University of Maribor, Koroška cesta 160, SI-2000 Maribor, Slovenia
Email: gutman@kg.ac.yu

Received 31 December 2006; revised 5 April 2007

A structure-descriptor EE , recently proposed by Estrada, is examined. If $\lambda_1, \lambda_2, \dots, \lambda_n$ are the eigenvalues of the molecular graph, then $EE = \sum_{i=1}^n e^{\lambda_i}$. In the case of trees with n vertices (that are the graph representations of alkane isomers C_nH_{2n+2}), the main structural factor influencing the differences between the EE -values has been found to be the Zagreb index Zg . The coefficient b in the regression line $EE = aZg + b$ is an almost perfectly linear function of n , implying that in the case of alkanes, EE linearly increases with the number of carbon atoms.

A molecular structure-descriptor, referred to as the 'Estrada index' is described in this paper. It is defined as follows: Let G be the molecular graph^{1,2}. Let n and m be, respectively, the number of vertices and edges of G . If G is acyclic, then it is referred to as³ a '(chemical) tree'. For (chemical) trees, $m=n-1$.

Basic properties of the graph eigenvalues are reported in various books^{1,4}. The eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ of the adjacency matrix of G are said to be the eigenvalues of G and form the spectrum of G . The Estrada index is then:

$$EE = EE(G) = \sum_{i=1}^n e^{\lambda_i} \quad \dots (1)$$

Although introduced quite recently⁵, the Estrada index has, already found numerous applications. It was used to quantify the degree of folding of long-chain molecules, especially proteins⁵⁻⁷. Another, fully unrelated, application of EE was put forward by Estrada and Rodríguez-Velázquez^{8,9}, who showed that it provides a measure of the average centrality of complex (communication, social, metabolic, etc) networks. In addition to this, it was claimed recently¹⁰ that there exists a connection between EE and the concept of extended atomic branching.

Until now, only some elementary mathematical properties of the Estrada index are established^{8,9,11}, and its dependence on molecular structure has not been properly investigated. The present paper is aimed at contributing towards filling this gap.

Theoretical

Recognizing the main structural feature of trees on which the Estrada index depends

In search for the structural features of trees that most significantly influence the value of the Estrada index, our starting point was an apparent (formal) analogy between the Estrada index, as defined by Eq. (1), and the graph energy $E(G)$ as defined by Eq. (2):

$$E = E(G) = \sum_{i=1}^n |\lambda_i| \quad \dots (2)$$

In contrast to the Estrada index, the graph energy was much studied in the past and is still attracting the attention of mathematical chemists (as covered in a recent review¹² and also recent papers¹³⁻¹⁷, of which several are concerned with the energy of trees). The chemical applications of graph energy are also well elaborated^{12,18-21}.

A tree in which the maximum vertex degree is not greater than four is referred to as^{1,3} a 'chemical tree'. Chemical trees provide the graph representations of alkanes. In particular, chemical trees with n vertices represent alkane isomers with the formula C_nH_{2n+2} . Trees possessing vertices of degree greater than four, fail to have a direct chemical interpretation. Nevertheless, in the present study (as well as in the earlier works^{3,22,23} concerned with the properties of alkanes), it was found purposeful to examine all trees with a given (fixed) number of vertices.

By plotting the EE -values of n -vertex trees versus the respective E -values, a peculiar pattern emerges. A characteristic example (pertaining to $n=9$) is shown in Fig. 1.

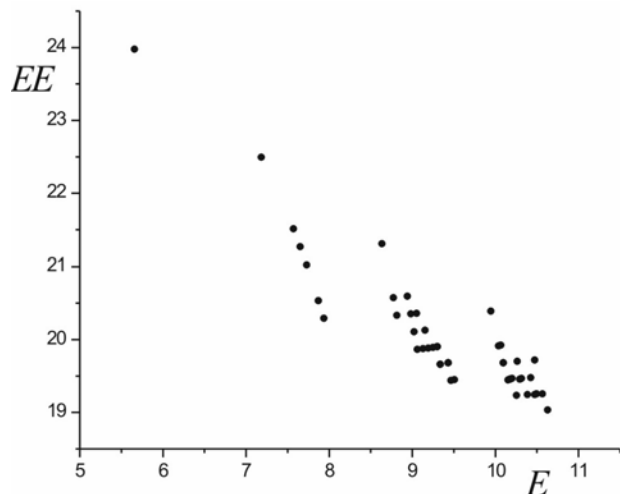


Fig. 1 — The Estrada indices (EE , Eq. (1)) of 9-vertex trees plotted versus the respective energies (E , Eq. (2)) [The clustering of the data points is according to the number of zero eigenvalues. Note that the data points lie on several (almost) horizontal lines].

The (EE, E) -data points are grouped into several clusters, as depicted by Fig. 1. A detailed examination reveals that the parameter according to which this clustering is formed is the number n_0 of zero eigenvalues (sometimes referred to as the 'nullity') of the respective graph. The dependence of the energy of trees on n_0 has been recently thoroughly investigated and is reasonably well understood^{12,24-27} E is a decreasing linear function of n_0 . On the other hand, the Estrada index happens to be almost independent of n_0 (Table 1). Therefore, the observed clustering of the (EE, E) -data points reflects the effect of nullity on graph energy, rather than any structure-dependence of EE .

A clue for recognizing the main structural feature on which EE depends is gained by observing that the (EE, E) -data points lie on several almost horizontal lines, i.e. pertain to almost constant EE -values, i.e. the respective EE -values vary within a remarkably narrow interval. Furthermore, most of these (mutually parallel) lines are found to be equidistant (Fig. 1). This suggests that one has to seek for some property that has the same value for all trees lying on a horizontal line. A significant help in this direction is the observation that the differently branched trees lie on different horizontal lines, and that with increasing

branching, the respective lines are shifted upwards (i.e., in direction of higher EE -values).

There exists a plethora of branching indices^{2,22,28-32}, all aimed at quantifying the intuitive notion of 'branching'. After some trial-and-error search, we recognized that the gross part of the Estrada index of trees (with fixed number of vertices) is determined by one of the simplest and oldest 'branching indices', namely by the so-called 'Zagreb index' or 'Zagreb-group index' (or more precisely, the 'first Zagreb-group index'), which in what follows is denoted by Zg . This structure-descriptor was introduced as early as in 1972, and was first used to describe the effect of branching on total π -electron energy^{33,34}. Its history and various chemical applications are reported in a review³⁵ whereas its main mathematical properties are also reported in a survey³⁶. Some most recent researches on Zg are found in various papers³⁷⁻³⁹.

The Zagreb index is defined as follows: Let G be a graph and v_1, v_2, \dots, v_n be its vertices. Let d_i be the degree of the vertex v_i , that is the number of first neighbours of v_i . Then:

$$Zg = Zg(G) = \sum_{i=1}^n (d_i)^2 \quad \dots (3)$$

Recall, that in the case when G is a chemical graph, Eq. (3) can be written as:

$$Zg = P + 4S + 9T + 16Q$$

where P , S , T and Q denote the number of primary, secondary, tertiary and quaternary carbon atoms, respectively, of the underlying molecule.

By direct checking, we found that all (EE, E) -data points lying on a horizontal line (and thus having nearly equal EE -values) have one and the same Zg -value. A characteristic example illustrating this fact is given in Table 1.

Not only is the Zagreb index the main parameter determining the value of the Estrada index of trees (with fixed number of vertices), but the dependence of EE on Zg is almost perfectly linear. Figure 2 shows the correlation between EE and Zg for (a) $n=8$ and (b) $n=13$. Such correlations have been studied for all values of n up to $n=20$, and include all n -vertex trees.

Numerical work

Table 2 shows the statistical data for the correlations of the form:

$$EE \approx a Zg + b \quad \dots (4)$$

Table 1 – The Estrada indices (EE), Zagreb indices (Zg), and nullities (n_0) of the octane isomers (Species with equal Zg have nearly the same EE -values, whereas EE is almost insensitive to the value of n_0)

Compound	EE	Zg	n_0
<i>n</i> -Octane	16.754	26	0
2-Methylheptane	16.957	28	2
3-Methylheptane	16.967	28	0
4-Methylheptane	16.967	28	2
3-Ethylhexane	16.977	28	0
2,3-Dimethylhexane	17.180	30	2
2,4-Dimethylhexane	17.170	30	2
2,5-Dimethylhexane	17.160	30	2
3,4-Dimethylhexane	17.189	30	0
2-Methyl-3-ethylpentane	17.190	30	2
2,2-Dimethylhexane	17.383	32	2
3,3-Dimethylhexane	17.402	32	2
2,3,4-Trimethylpentane	17.393	32	2
3-Methyl-3-ethylbutane	17.422	32	0
2,2,3-Trimethylpentane	17.616	34	2
2,2,4-Trimethylpentane	17.587	34	4
2,3,3-Trimethylpentane	17.626	34	2
2,2,3,3-Tetramethylbutane	18.054	38	4

obtained for the complete sets of n -vertex trees ($6 \leq n \leq 20$). The correlations are evidently linear (Fig. 2). The coefficients a and b in Eq. (4) were determined by least-squares fitting. For larger values of n , the correlation is slightly curvilinear. For each value of n , the existence of curvilinearity was checked by means of F -test at 99% confidence level. We found that curvilinearity is statistically significant for $n \geq 10$. Table 2 provides the necessary details.

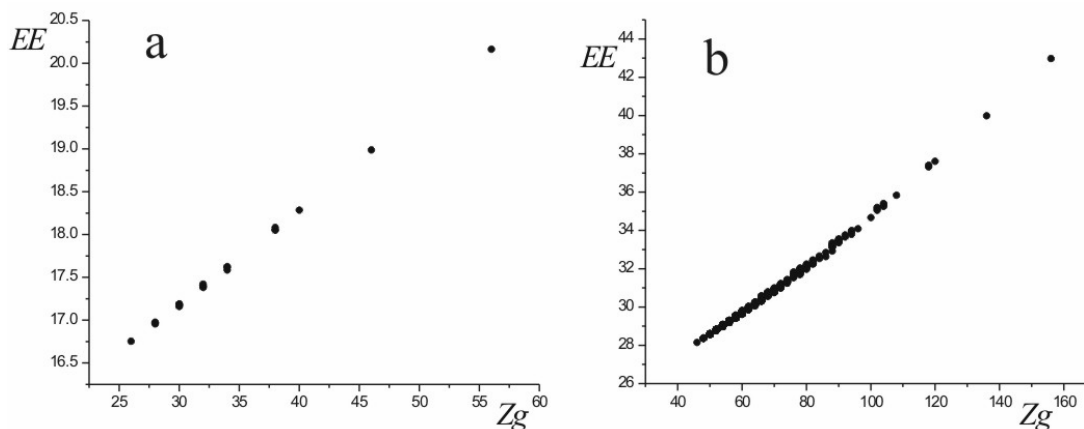


Fig. 2 — Correlation between the Estrada indices (EE) of n -vertex trees and the respective Zagreb indices (Zg) for: (a) $n=8$, and (b) $n=13$. (The correlations are evidently linear. However, for $n \geq 10$ a slight curvilinearity is observed. One should note that because of insensitivity of the Estrada index to other structural features, each point on these diagrams corresponds to *all* n -vertex trees with a given Zg -value).

Dependence of the coefficients a and b in Eq. (4) on the number n of vertices of the underlying trees is shown in Figs 3 and 4.

Figure 3 gives the impression that the coefficient a depends on n in a rather non-linear manner. However, if the actual numerical values of this coefficient are taken into account (Table 2), then we realize that a is almost independent of n , and that for all values of n , its value is $a = 0.11 \pm 0.01$.

The coefficient b in Eq. (4) depends on n in an almost perfectly linear manner (Fig. 4). This implies that the Estrada index of trees is 'in average' a simple linear function of the number of vertices. The coefficient b is much greater than a (by two orders of magnitude, see Table 2). Therefore, if no structural detail of a tree is taken into account, but only its number of vertices, then $EE \approx (1.735 \pm 0.002)n - (0.13 \pm 0.03)$ could be used as a reasonably good approximation.

Mathematical analysis

The results of our empirical studies of the structure dependence of the Estrada index of trees (outlined in the preceding sections) can be partially rationalized by means of the following mathematical considerations.

Using the same notation as in Eqs (1) and (2), the k -th spectral moment M_k of a graph G is defined as:

$$M_k = M_k(G) = \sum_{i=1}^n (\lambda_i)^k.$$

Bearing in mind the power-series expansion of the function e^x , one immediately gets⁸:

Table 2 – Statistical data for the correlations described by Eq. (4): n = number of vertices, # trees = number of trees in the sample considered (of which the number of trees corresponding to structural isomers of alkanes is # isomer), R = correlation coefficient, F = result of F -test for curvilinearity of the correlation, at 99% confidence level (The slope a of the correlation is nearly constant; its dependence on n is shown in Fig. 3. The intercept b linearly increases with n , see Fig. 4)

n	# Trees	# Isomers	a	b	R	F
6	6	5	0.10581	10.286	0.99997	no
7	11	9	0.10975	12.046	0.9998	no
8	23	18	0.11321	13.780	0.9996	no
9	47	35	0.11621	15.496	0.9993	no
10	106	75	0.11849	17.210	0.9989	Borderline
11	235	159	0.12039	18.919	0.9984	yes
12	551	355	0.12168	20.640	0.9979	yes
13	1301	802	0.12263	22.368	0.9975	yes
14	3159	1858	0.12320	24.111	0.9973	yes
15	7741	4347	0.12367	25.855	0.9969	yes
16	19320	10359	0.12394	27.608	0.9968	yes
17	48629	24894	0.12414	29.365	0.9967	yes
18	123867	60523	0.12429	31.124	0.9966	yes
19	317955	148284	0.12441	32.884	0.9966	yes
20	823065	366319	0.12452	34.645	0.9966	yes

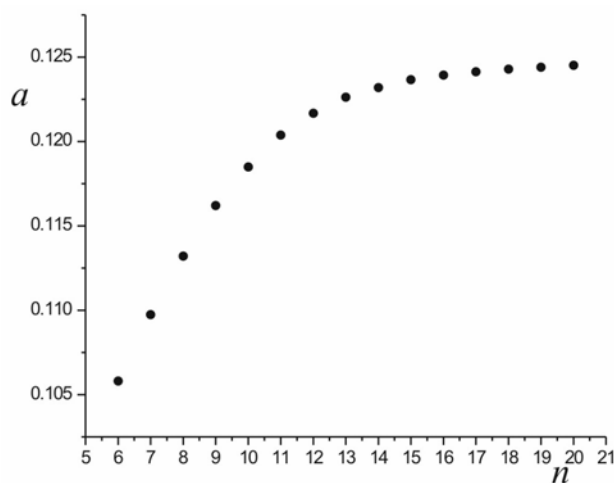


Fig. 3 — Dependence of the slope a of the correlation $EE \approx aZg + b$ on the number n of vertices (Although this dependence appears to be evidently non-linear, it should be noted that a varies within a very narrow interval, i.e., it is almost independent of n).

$$EE = \sum_{k \geq 0} \frac{M_k}{k!} = M_0 + M_1 + \frac{1}{2}M_2 + \frac{1}{6}M_3 + \frac{1}{24}M_4 + \frac{1}{120}M_5 + \frac{1}{720}M_6 + \dots$$

from which, after abandoning the sixth- and higher-order terms,

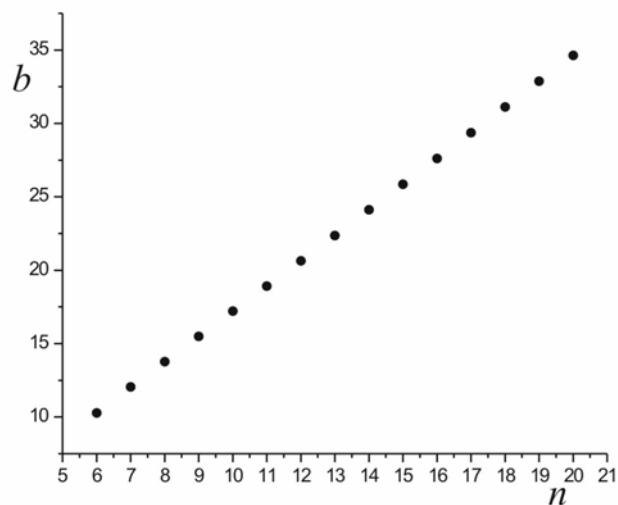


Fig. 4 — Dependence of the intercept b of the correlation $EE \approx aZg + b$ on the number n of vertices. This correlation is almost perfectly linear (correlation coefficient: 0.99999), revealing that the Estrada index of trees is an increasing linear function of the number of vertices. The respective regression line is $b = (1.735 \pm 0.002)n - (0.13 \pm 0.03)$.

$$EE \approx M_0 + M_1 + \frac{1}{2}M_2 + \frac{1}{6}M_3 + \frac{1}{24}M_4 + \frac{1}{120}M_5 + \dots \quad (5)$$

For all graphs^{1,4}, $M_0 = n$ and $M_1 = 0$. For all

bipartite graphs (among which are also the trees)^{1,4}, $M_3 = M_5 = M_7 = \dots = 0$. In addition to this, for all trees³⁶, $M_2 = 2(n-1)$ and $M_4 = 2Zg - 2n + 2$.

When all these relations are substituted back into Eq. (5), we arrive at a remarkably simple approximate expression for the Estrada index:

$$EE \approx \frac{1}{12} Zg + \frac{23}{12} n - \frac{11}{12} \quad \dots \quad (6)$$

In view of the crudeness of the approximation (5), the agreement between the empirically acquired Eq. (4) and the calculated Eq. (6) is remarkably good. Note that formula (6) predicts that the Estrada index of trees is a linear function of the Zagreb index and that the slope of the EE/Zg -line is independent of n . Indeed, the slope $1/12 = 0.083$ in Eq. (6) is quite close to the empirically established $a = 0.11 \pm 0.01$ in Eq. (4).

The term $(23/12)n - 11/12$ in Eq. (6) should be compared with the coefficient b in Eq. (4). Again, $23/12 = 1.917$ is reasonably close to the empirically found value: 1.735 ± 0.002 . Only the constant $11/12 = 0.917$ significantly differs from the calculated 0.13 ± 0.03 . Evidently, this later disagreement would diminish if one would include into Eq. (5) also higher-order terms (M_6, M_8, \dots). However, accuracy of the presently obtained approximation seems to be sufficient already at the level of Eq. (5).

Conclusions

The dependence of the newly conceived Estrada index EE , Eq. (1), on molecular structure has been analyzed in the case of acyclic molecules. We have established the main structural parameters on which EE depends. This is, first of all, the number n of vertices of the molecular graph (i.e., the number of carbon atoms of the underlying alkane). Within groups of isomers (whose n -values are fixed), EE is found to increase with the increasing extent of branching of the carbon-atom skeleton. We have been able to identify the particular 'branching index' responsible for this effect. This turns out to be the Zagreb index Zg . The quantitative details of the dependence of EE on Zg have been established, both empirically and using a pertinent approximate mathematical model.

The Estrada index of a tree with n vertices (i.e., an alkane with n carbon atoms) and with Zagreb index Zg can be computed by means of the approximate expression:

$$EE = 1.735n - 0.13 + 0.11 Zg$$

which in all cases, is capable of reproducing EE with an error less than 0.1%. This accuracy is sufficient for any presently known application⁵⁻¹⁰ of the Estrada index.

The dependence of the Estrada index of acyclic molecules on structural parameters other than n and Zg appears to be insignificant, and its investigation seems to be not purposeful.

References

- Gutman I & Polansky O E, *Mathematical Concepts in Organic Chemistry* (Springer, Berlin) 1986.
- Randić M, *Indian J Chem*, 42A (2003) 1207.
- Gutman I, Vidović D, Furtula B & Vesel A, *Indian J Chem*, 42A (2003) 1241.
- Cvetković D, Doob M & Sachs H, *Spectra of Graphs - Theory and Application* (Johann Ambrosius Barth Verlag, Heidelberg) 1995.
- Estrada E, *Chem Phys Lett*, 319 (2000) 713.
- Estrada E, *Bioinformatics*, 18 (2002) 697.
- Estrada E, *Proteins*, 54 (2004) 727.
- Estrada E & Rodríguez-Velázquez J A, *Phys Rev*, 71E (2005) 56103.
- Estrada E & Rodríguez-Velázquez J A, *Phys Rev*, 72E (2005) 46105.
- Estrada E, Rodríguez-Velázquez J A & Randić M, *Int J Quantum Chem*, 106 (2006) 823.
- Gutman I, Estrada E & Rodríguez-Velázquez J A, *Croat Chem Acta*, 80 (2007).
- Gutman I, *J Serb Chem Soc*, 70 (2005) 441.
- Ramane H S, Walikar H B, Rao S B, Acharya B D, Hampiholi P R, Jog S R & Gutman I, *Appl Math Lett*, 18 (2005) 679.
- Lin W, Guo X & Li H, *MATCH Commun Math Comput Chem*, 54 (2005) 363.
- Yan W, Ye L, *MATCH Commun Math Comput Chem*, 53 (2005) 449.
- Indulal G & Vijayakumar A, *MATCH Commun Math Comput Chem*, 55 (2006) 83.
- Zhou B, *MATCH Commun Math Comput Chem*, 55 (2006) 91.
- Gutman I, *Indian J Chem*, 40A (2001) 929.
- Gutman I, Vidović D, Cmiljanović N, Milosavljević S & Radenković S, *Indian J Chem*, 42A (2003) 1309.
- Furtula B, Gutman I & Turković N, *Indian J Chem*, 44A (2005) 9.
- Gutman I, Milosavljević S, Furtula B & Cmiljanović N, *Indian J Chem*, 44A (2005) 13.
- Gutman I, Lepović M, Vidović D & Clark L H, *Indian J Chem*, 41A (2002) 457.
- Gutman I & Vidović D, *Indian J Chem*, 41A (2002) 893.
- Gutman I, Cmiljanović N, Milosavljević S & Radenković S, *Chem Phys Lett*, 383 (2004) 171.
- Gutman I, Cmiljanović N, Milosavljević S & Radenković S, *Monatsh Chem*, 135 (2004) 765.

- 26 Gutman I, Stevanović D, Radenković S, Milosavljević S & Cmiljanović N, *J Serb Chem Soc*, 69 (2004) 777.
- 27 Fiorini S, Gutman I & Sciriha I, *Lin Algebra Appl*, 397 (2005) 245.
- 28 Gutman I, Yeh Y N, Lee S L & Luo Y L, *Indian J Chem*, 32A (1993) 651.
- 29 Kuanar M, Kuanar S K, Mishra B K & Gutman I, *Indian J Chem*, 38A (1999) 525.
- 30 Gutman I, Araujo O & Morales D A, *Indian J Chem*, 39A (2000) 385.
- 31 Todeschini R & Consonni V, *Handbook of Molecular Descriptors* (Wiley-VCH, Weinheim) 2000.
- 32 Perdih A & Perdih B, *Indian J Chem*, 42A (2003) 1246.
- 33 Gutman I & Trinajstić N, *Chem Phys Lett*, 17 (1972) 535.
- 34 Gutman I, Ruščić B, Trinajstić N & Wilcox C F, *J Chem Phys*, 62 (1975) 3399.
- 35 Nikolić S, Kovačević G, Miličević A & Trinajstić N, *Croat Chem Acta*, 76 (2003) 113.
- 36 Gutman I & Das K C, *MATCH Commun Math Comput Chem*, 50 (2004) 183.
- 37 Vukičević D & Trinajstić N, *MATCH Commun Math Comput Chem*, 53 (2005) 111.
- 38 Zhang S, Wang W & Cheng T C E, *MATCH Commun Math Comput Chem*, 56 (2006) 579.
- 39 Zhou B & Stevanović D, *MATCH Commun Math Comput Chem*, 56 (2006) 571.