Estrada index of acyclic molecules

Ivan Gutman*, Boris Furtula*, Biljana Glišić*, Violeta Marković* & Aleksander Veselb

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

Department of Mathematics, PEF, University of Maribor, Koroska cesta 160, SI-2000 Maribor, Slovenia

Received 31 December 2006; revised 5 April 2007

A structure-descriptor $EE$, recently proposed by Estrada, is examined. If $\lambda_1, \lambda_2, \ldots, \lambda_n$ 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 $Z_g$. The coefficient $b$ in the regression line $EE = a Z_g + b$ 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. Let $n$ and $m$ be, respectively, the number of vertices and edges of $G$. If $G$ is acyclic, then it is referred to as a `(chemical) tree'. For (chemical) trees, $m=n-1$.

Basic properties of the graph eigenvalues are reported in various books. The eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ 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:

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

Although introduced quite recently, the Estrada index has already found numerous applications. It was used to quantify the degree of folding of long-chain molecules, especially proteins. Another, fully unrelated, application of $EE$ was put forward by Estrada and Rodríguez-Velázquez, 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 recently 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, 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):

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

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 review and also recent papers, of which several are concerned with the energy of trees). The chemical applications of graph energy are also well elaborated.

A tree in which the maximum vertex degree is not greater than four is referred to as 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 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 n₀ of zero eigenvalues (sometimes referred to as the `nullity’) of the respective graph. The dependence of the energy of trees on n₀ has been recently thoroughly investigated and is reasonably well understood. E is a decreasing linear function of n₀. On the other hand, the Estrada index happens to be almost independent of n₀ (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.

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:

\[ EE \approx a Zg + b \]  \hspace{1cm} \ldots (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 $n_0$).

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

Compared to 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.

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 mind the power-series expansion of the function $e^\lambda$, one immediately gets:

$$EE \approx (1.735 \pm 0.002)n - (0.13 \pm 0.03).$$
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).

<table>
<thead>
<tr>
<th>( n )</th>
<th># Trees</th>
<th># Isomers</th>
<th>( a )</th>
<th>( b )</th>
<th>( R )</th>
<th>( F )</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>6</td>
<td>5</td>
<td>0.10581</td>
<td>10.286</td>
<td>0.99997</td>
<td>no</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
<td>9</td>
<td>0.10975</td>
<td>12.046</td>
<td>0.9986</td>
<td>no</td>
</tr>
<tr>
<td>8</td>
<td>23</td>
<td>18</td>
<td>0.11321</td>
<td>13.780</td>
<td>0.9996</td>
<td>no</td>
</tr>
<tr>
<td>9</td>
<td>47</td>
<td>35</td>
<td>0.11621</td>
<td>15.496</td>
<td>0.9993</td>
<td>no</td>
</tr>
<tr>
<td>10</td>
<td>106</td>
<td>75</td>
<td>0.11849</td>
<td>17.210</td>
<td>0.9989</td>
<td>Borderline</td>
</tr>
<tr>
<td>11</td>
<td>235</td>
<td>159</td>
<td>0.12039</td>
<td>18.919</td>
<td>0.9984</td>
<td>yes</td>
</tr>
<tr>
<td>12</td>
<td>551</td>
<td>355</td>
<td>0.12168</td>
<td>20.640</td>
<td>0.9979</td>
<td>yes</td>
</tr>
<tr>
<td>13</td>
<td>1301</td>
<td>802</td>
<td>0.12263</td>
<td>22.368</td>
<td>0.9975</td>
<td>yes</td>
</tr>
<tr>
<td>14</td>
<td>3159</td>
<td>1858</td>
<td>0.12320</td>
<td>24.111</td>
<td>0.9973</td>
<td>yes</td>
</tr>
<tr>
<td>15</td>
<td>7741</td>
<td>4347</td>
<td>0.12367</td>
<td>25.855</td>
<td>0.9969</td>
<td>yes</td>
</tr>
<tr>
<td>16</td>
<td>19320</td>
<td>10359</td>
<td>0.12394</td>
<td>27.608</td>
<td>0.9967</td>
<td>yes</td>
</tr>
<tr>
<td>17</td>
<td>48629</td>
<td>24894</td>
<td>0.12414</td>
<td>29.365</td>
<td>0.9967</td>
<td>yes</td>
</tr>
<tr>
<td>18</td>
<td>123867</td>
<td>60523</td>
<td>0.12429</td>
<td>31.124</td>
<td>0.9966</td>
<td>yes</td>
</tr>
<tr>
<td>19</td>
<td>317955</td>
<td>148284</td>
<td>0.12441</td>
<td>32.884</td>
<td>0.9966</td>
<td>yes</td>
</tr>
<tr>
<td>20</td>
<td>823065</td>
<td>366319</td>
<td>0.12452</td>
<td>34.645</td>
<td>0.9966</td>
<td>yes</td>
</tr>
</tbody>
</table>

For all graphs \( G \), \( M_0 = n \) and \( M_1 = 0 \). For all bipartite graphs (among which are also the trees) \( G \),
\[ M_1 = M_2 = M_3 = \cdots = 0. \] In addition to this, for all trees\(^n\), \( M_2 = 2(n-1) \) and \( M_3 = 2 \ Zg - 2n + 2 \).

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

\[
EE = \frac{1}{12} Zg + \frac{23}{12} n - \frac{11}{12}. \quad \ldots (6)
\]

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 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:

\[
EE = 1.735 n - 0.13 + 0.11 Zg
\]

which in all cases, is capable of reproducing \( EE \) with an error less than 0.1%. This accuracy is sufficient for any presently known application \(^{10}\) 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**