

Relations between topological indices of large chemical trees

Ivan Gutman*, Dušica Vidović & Boris Furtula

Faculty of Science, University of Kragujevac, P.O.Box 60, YU-34000 Kragujevac, Serbia & Montenegro
and

Aleksander Vesel

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

Received 16 October 2002

Several approximate relations have recently been established between molecular-graph-based structure descriptors of alkanes, in particular between (a) eigenvalue sum and Hosoya index, (b) greatest graph eigenvalue and connectivity index, (c) Wiener index and smallest positive Laplacian eigenvalue, (d) greatest Laplacian and greatest ordinary graph eigenvalue, (e) Zenkevich and Wiener index, and (f) hyper-Wiener and Wiener index. These all have been found to hold for alkanes with $n=10$ or fewer carbon atoms, and have verified on samples consisting of all alkane isomers. Applying an algorithm for generating trees uniformly by random we have now tested these regularities for very large chemical trees ($n=50$). It has been found that regularities (c) and (f) hold equally well in the case of very large chemical trees, whereas regularities (a), (d) and (e) are applicable, but with significantly attenuated accuracy. Regularity (b) vanishes at large values of n .

In the last few years our research group has systematically searched for relations between molecular-graph-based structure-descriptors, especially in the simplest case of chemical trees. A chemical tree is a tree (= connected acyclic graph)¹ in which no vertex has degree greater than 4; chemical trees are the graph-representations of alkanes¹.

We found several such relations, of which in this work we are concerned with the following six.

- (a) Quite some time ago it was noticed² that there is a linear correlation between the eigenvalue sum ES (= the sum of all positive eigenvalues of the molecular graph) and the logarithm of the Hosoya index Z (= the number of ways in which independent edges can be selected in the molecular graph). However, only quite recently it was established³ that the data points lie not on a single, but on several lines, depending on the number η of zero eigenvalues. An illustrative example is shown in Fig. 1; for further details see ref. 3.
- (b) Two measures of molecular branching, both proposed in the 1970s, are the greatest eigenvalue λ_1 of the molecular graph⁴ and the connectivity index χ (= sum over all pairs of adjacent vertices of the terms $(d_r d_s)^{-1/2}$, where d_r is the degree of the vertex r)⁵. In spite of the fact that both structure-descriptors are frequently encountered in the chemical literature⁶, their mutual relation

was not investigated until quite recently⁷. It was found⁷ that there is a decreasing correlation between λ_1 and χ , of very low quality. On the other hand, the data points are grouped into clusters and within each such cluster a reasonably good increasing linear correlation is observed. The structural parameter on the basis of which the clustering occurs is – as a kind of surprise – $n_3 + 3n_4$, where n_3 and n_4 are the number of tertiary and quaternary carbon atoms, respectively, or – what is the same – the number of vertices of degree 3 and 4 in the molecular graph. An illustrative example is shown in Fig. 2; for further details see ref. 7.

- (c) The existence of an exact mathematical relation between the Wiener index W (= the sum of distance between all pairs of vertices of the molecular graph)⁸ and the Laplacian eigenvalues is known for some time (for details and further references see refs 8 and 9). Based on this relation and approximation $W \approx n/\mu_{n-1}$ was put forward¹⁰, where n is the number of vertices and $\mu_{(n-1)}$ is the smallest positive Laplacian eigenvalue of the respective chemical tree. The example depicted in Fig. 3 shows that the correlation between W and n/μ_{n-1} does indeed exist, but is not particularly good.
- (d) The connection between the greatest graph eigenvalue λ_1 and the greatest Laplacian eigenvalue μ_1 was determined by us in a recent

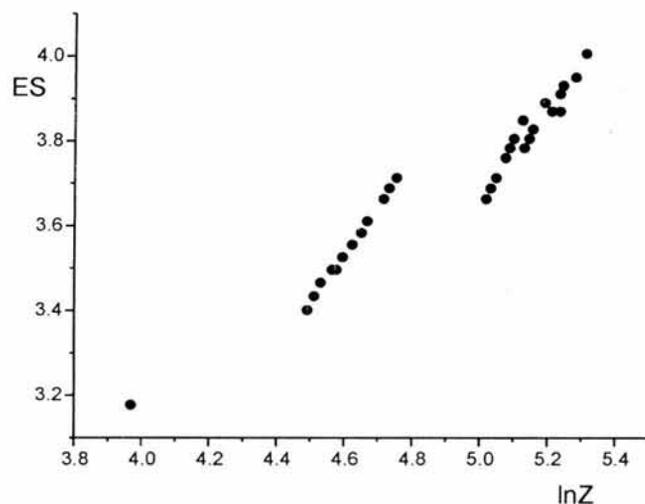


Fig. 1—The eigenvalue sum (ES) vs. the logarithm of the Hosoya index (Z) for the 35 isomeric nonanes. The data points are clustered into three groups, depending on the number of zero eigenvalues (η). The isolated point corresponds to 2,2,4,4-tetramethylpentane, the only nonane species with $\eta=5$. The two other groups of linearly correlated points pertain to $\eta=3$ (middle) and $\eta=1$ (right).

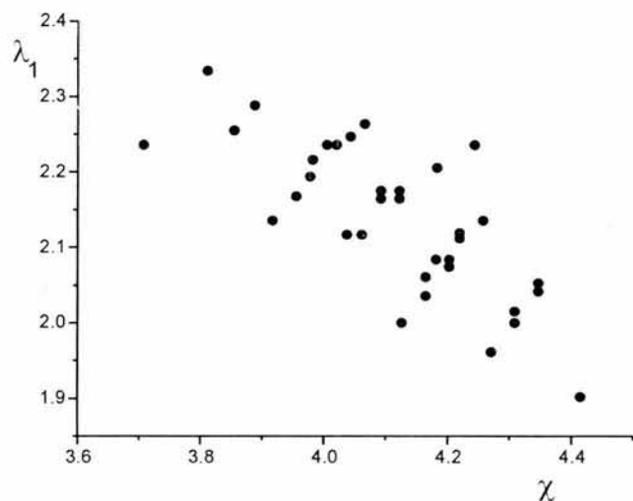


Fig. 2—Correlation between the greatest eigenvalue (λ_1) and the connectivity index (χ) for the isomeric nonanes. Each of the seven groups of data points has a different value of the parameter $n_3 + 3n_4$, equal to 0, 1, 2, 3, 4, 5 & 6 (going from right to left). The single point in the right lower corner, corresponds to the unbranched n -nonane with $n_3 + 3n_4 = 0$; the two points in the left upper corner correspond to 2,2,3,3- and 2,2,4,4-tetramethylpentane, for which $n_3 + 3n_4 = 6$.

work.¹¹ Here the data points are grouped according to d_{max} —the maximal vertex degree. In chemical trees with more than two vertices d_{max} may assume only the values 2, 3, 4; besides, $d_{max}=2$ happens only in the case of unbranched n -alkanes. An illustrative example is shown in Fig. 4; further details are found in ref. 11.

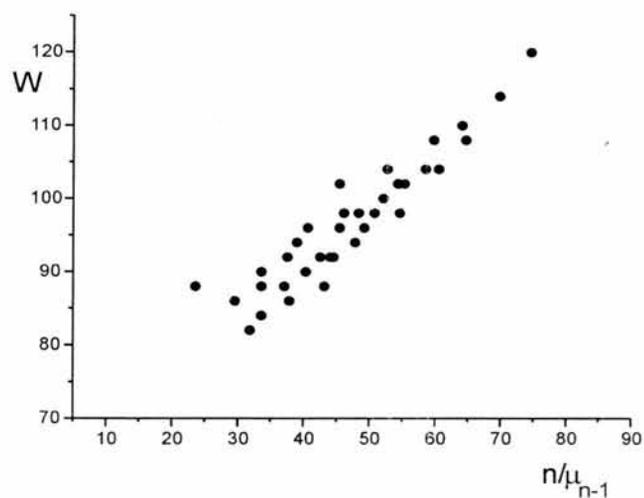


Fig. 3—The Wiener index (W) vs. the reciprocal value of the smallest positive Laplacian eigenvalue (μ_{n-1}) for the isomeric nonanes. The correlation coefficient is $R=0.937$.

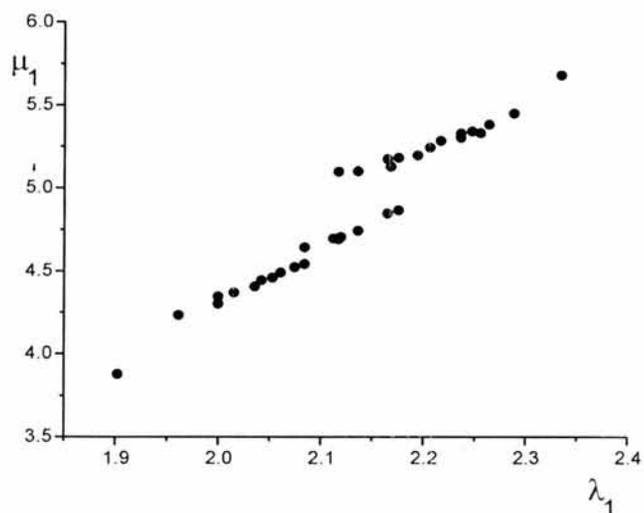


Fig. 4—Correlation between the greatest Laplacian eigenvalue (μ_1) and the greatest ordinary graph eigenvalue (λ_1) for isomeric nonanes. There are two groups of data points, pertaining to $d_{max}=4$ (upper cluster) and $d_{max}=3$ (lower cluster). The isolated point in the lower left corner corresponds to n -nonane with $d_{max}=2$.

(e) Zenkevich proposed¹²⁻¹⁴ a molecular structure-descriptor U , being proportional to the sum of vibrational energies (in the harmonic approximation) of the stretching vibrations of the carbon-carbon bonds of an organic molecule. In the case of alkanes, it could be demonstrated¹⁵ that U is related to the Wiener index. A more detailed examination revealed¹⁶ that the respective data points are grouped, each group having a different value for the number of vertices of degree one (n_1). This regularity is illustrated in Fig. 5; further details are found in refs 15 and 16.

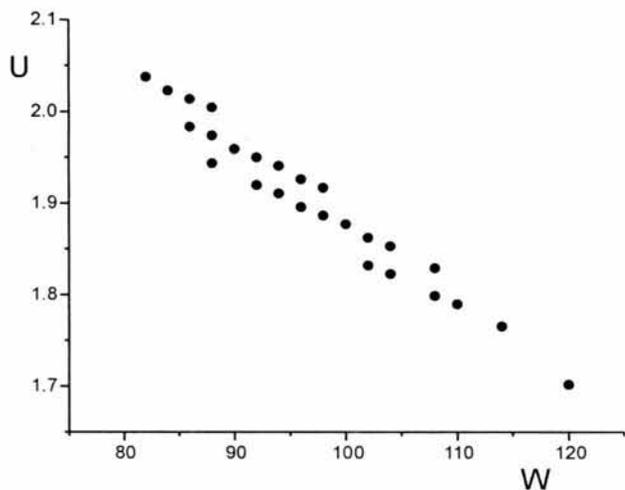


Fig. 5—Correlation between the Zenkevich index (U) and the Wiener index (W) for isomeric nonanes. The data points lie on four parallel lines, corresponding (from top to bottom) to $n_1 = 6, 5, 4$ & 3 , where n_1 is the number of vertices of degree one in the molecular graph (= the number of methyl groups). The isolated point in the lower right corner corresponds to n -nonane with $n_1=2$.

(f) Theoretical arguments were recently communicated¹⁷, implying that the hyper-Wiener index WW (an extension of the Wiener index concept^{6,8}, introduced by Randić ten years ago¹⁸) and the Wiener index W must be closely correlated. Namely, it was shown¹⁷ that WW is bounded from both below and above by linear functions of W . A detailed numerical analysis of the correlation between WW and W fully confirmed this anticipation¹⁹. A typical result of this kind is shown in Fig. 6.

Approaching Chemical Trees with Large Number of Vertices

The six regularities outlined above and illustrated in Figs 1-6 were established in the case of chemical trees with a relatively small number of vertices (usually $n \leq 10$). We examined all possible such n -vertex trees (corresponding to all possible structural isomers of alkanes with n carbon atoms), and verified the respective regularity for each member of the sample.

For larger values of n such an approach would not be feasible. (Recall that, for instance, there are 366,319 and 1,117,743,651,746,953,270 alkane isomers with 20 and 50 carbon atoms, respectively.)

Therefore, instead of considering the set of all n -vertex chemical trees, we had to use a representative sample thereof. For this an algorithm was employed that for a given value of n , generates n -vertex trees uniformly by random²⁰. By means of this algorithm a

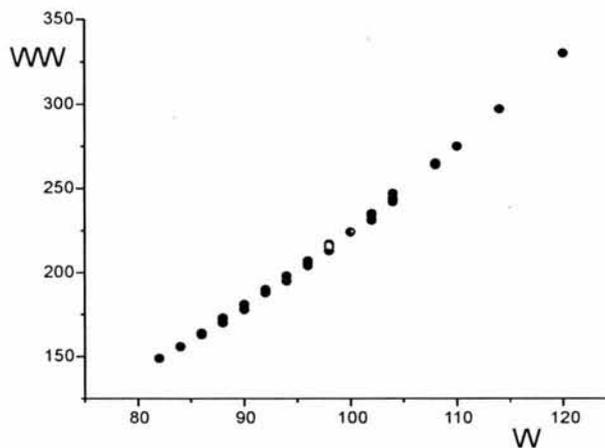


Fig. 6—The hyper-Wiener index (WW) vs. the Wiener index (W) for isomeric nonanes. The correlation coefficient is $R=0.9973$.

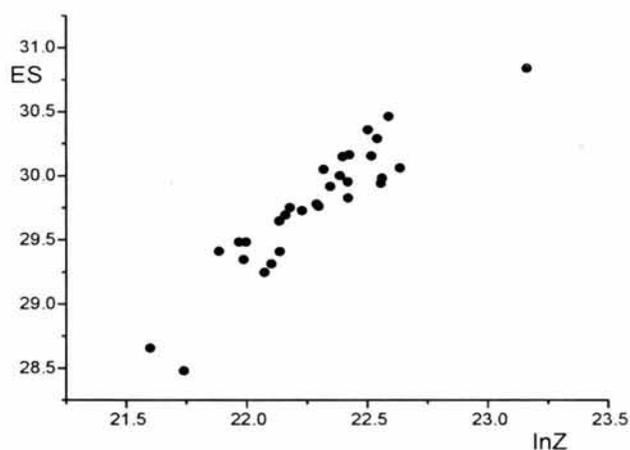


Fig. 7—Correlation between the eigenvalue sum (ES) and the logarithm of the Hosoya index (Z) for a sample consisting of 30 randomly constructed 50-vertex chemical trees. Of these, 2, 8, 14, 4, 1, and 1 trees have $\eta = 2, 4, 6, 8, 10$, and 12 , respectively. Contrary to what could be expected on the basis of Fig. 1, the point in the upper right corner corresponds to a tree with $\eta=4$, and the two points in the lower left corner correspond to a tree with $\eta=10$ (left) and another with $\eta=12$ (right). By means of such a sample no regularity of the type (a) can be envisaged.

sufficiently large number of trees was generated, usually 2000. From these the non-chemical trees (with the property $d_{max} > 4$) were eliminated and the remaining chemical trees—forming the “starting sample”—were further investigated.

At the first glance, we could simply choose a certain number of trees from the starting sample and check if the above specified regularities still hold. Indeed, such a simple procedure was possible in cases (c) and (f), leading to results shown in Figs 10 and 13. However, in cases (a), (b), (d) and (e) such a direct approach is not purposeful. This is illustrated in Fig. 7 where the ES - $\ln Z$ correlation is shown for 30 randomly constructed 50-vertex chemical trees. No

clustering of the data points can be seen, and the data points corresponding to various values of η are mixed.

In order to overcome this difficulty we checked the regularities (a), (b), (d) and (e) by means of the following procedure. For all trees from the starting sample we computed the values of the structural parameter according to which the clustering of data points might occur, namely η in case (a), $n_3 + 3 n_4$ in case (b), d_{max} in case (d), and n_1 in case (e). In case (d) we then selected from the starting sample 15 trees with $d_{max}=3$ and 15 trees with $d_{max}=4$. In cases (a), (b) and (e) we selected 10 trees with an as small as possible, 10 trees with an as large as possible and 10 trees with a medium value of the respective parameter; details are found in the captions of Figs 8, 9 and 12. Under "as small as possible" and "as large as possible" we mean values that pertain to at least 10 trees in the starting sample. This procedure limits the number of clusters of data points and maximally reduces the overlapping between them.

Results and Discussion

None of the empirical relations (a)-(f) between topological indices of chemical trees could be proven in a rigorous mathematical manner, although each of them was corroborated by some kind of theoretical reasoning. Therefore it was desirable to check the validity of these relations for values of n far beyond what is usually considered in chemical applications. We have performed our tests on $n=50$.

The purpose of our studies is the following. If a regularity, earlier observed for $n \leq 10$, remains valid also for $n=50$, then it is likely that this is a generally valid and mathematically exact relation. Proving this relation remains then a challenge for the future. If, on the other hand, a regularity, earlier observed for $n \leq 10$, is found to be violated for $n=50$, then attempts to prove it should be abandoned.

The results obtained for chemical trees with $n=50$ are shown in Figs 8-13.

By inspecting the data displayed in Figs 8-13 we arrive at the following conclusions:

1. Regularities (c) and (f) hold equally well for small and for large values of n . The correlation between W and n/μ_{n-1} is weak for any value of n , and is not significantly weaker for $n=50$ than for $n \leq 10$. The correlation between WW and W is excellent, and basically linear, for all values of n .

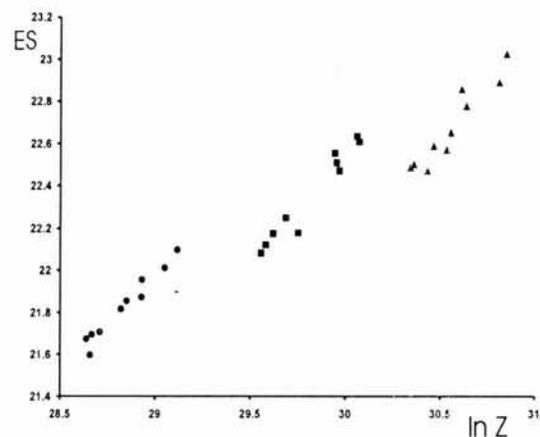


Fig. 8—The eigenvalue sum (ES) vs. the logarithm of the Hosoya index (Z) for 30 randomly selected 50-vertex chemical trees with $\eta=2$ (triangles), $\eta=6$ (squares) and $\eta=10$ (circles); cf. Fig. 1.

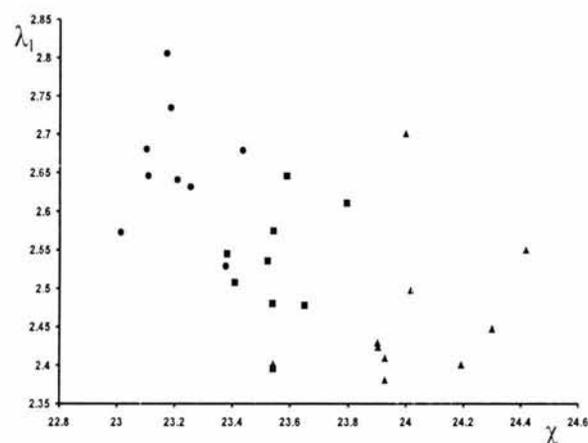


Fig. 9—The greatest eigenvalue (λ_1) vs. the connectivity index (χ) for 30 randomly selected 50-vertex chemical trees with $n_3 + 3 n_4=15$ (triangles), $n_3 + 3 n_4=20$ (squares) and $n_3 + 3 n_4=25$ (circles); cf. Fig. 2.

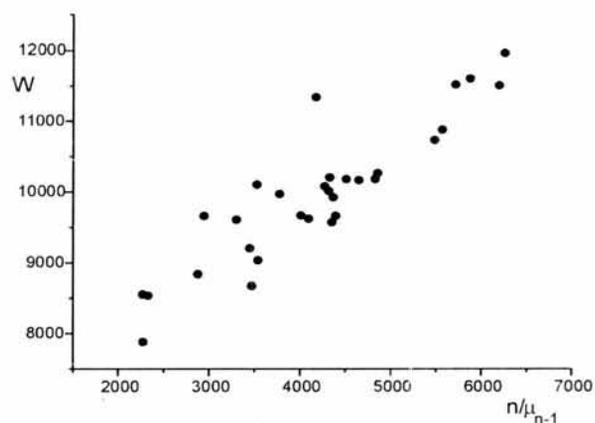


Fig. 10—The Wiener index (W) vs. the reciprocal value of the smallest positive Laplacian eigenvalue (μ_{n-1}) for 30 randomly selected 50-vertex chemical trees; cf. Fig. 3. The respective correlation coefficient is $R=0.901$.

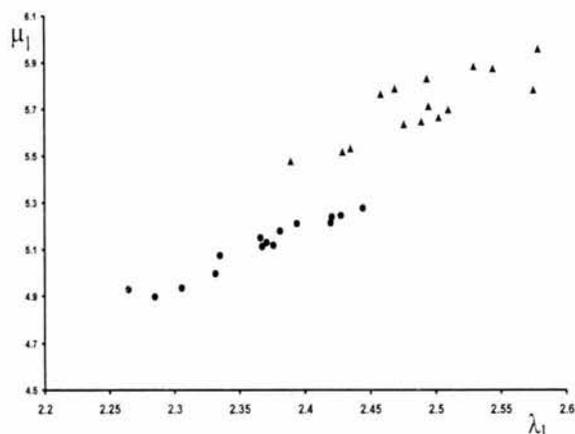


Fig. 11—The greatest Laplacian eigenvalue (μ_1) vs. the greatest ordinary eigenvalue (λ_1) for 30 randomly selected 50-vertex chemical trees with $d_{\max}=3$ (circles) and $d_{\max}=4$ (triangles); cf. Fig. 4.

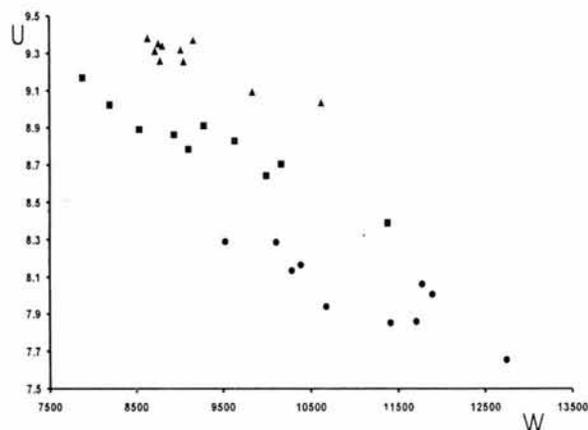


Fig. 12—The Zenkevich index (U) vs. the Wiener index (W) for 30 randomly selected 50-vertex chemical trees with $n_1=13$ (circles), $n_1=18$ (squares) and $n_1=23$ (triangles); cf. Fig. 5.

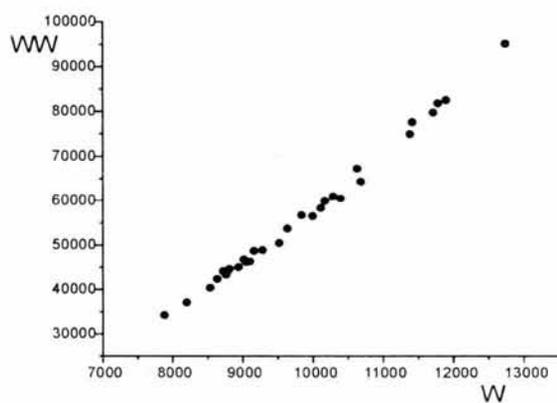


Fig. 13—The hyper-Wiener index (WW) vs. the Wiener index (W) for 30 randomly selected 50-vertex chemical trees; cf. Fig. 6. The correlation coefficient is $R=0.9963$.

2. Regularities (a), (d) and (e) hold also for very large chemical trees. There is, however, a much more pronounced dispersion of the data points and, consequently, near-lying clusters partially overlap. In order to make the clustering of the data points easier to envisage, in Figs 8 and 12 we have included data for chemical trees with only three distinct values of the clustering parameters, that differ as much as possible.
3. Even with such a special selection of chemical trees, the clustering of the data points in the case (b) is no more visible at $n=50$, meaning that the earlier reported⁷ correlation between λ_1 and χ disappears when the number of vertices is sufficiently large.

In summary, we established that regularities (c) and (f) hold equally well in the case of very large chemical trees, that regularities (a), (d) and (e) are applicable, but with somewhat reduced accuracy, while regularity (b) vanishes.

References

- 1 Gutman I & Polansky O E, *Mathematical concepts in organic chemistry* (Springer-Verlag, Berlin), 1986.
- 2 Hosoya H, Hosoi K & Gutman I, *Theor chim Acta*, 38 (1975) 37.
- 3 Gutman I, Vidović D & Hosoya H, *Bull chem Soc Japan*, 75 (2002) 1723.
- 4 Cvetković D M & Gutman I, *Croat chem Acta*, 49 (1977) 115.
- 5 Randić M, *J Am chem Soc*, 97 (1975) 6609.
- 6 Todeschini R & Consonni V, *Handbook of molecular descriptors* (Wiley-VCH, Weinheim), 2000.
- 7 Gutman I & Vidović D, *Theor Chem Acc*, 108 (2002) 98.
- 8 Gutman I, Yeh Y N, Lee S L & Luo Y L, *Indian J Chem*, 32A (1993) 651.
- 9 Gutman I, Lee S L, Chu C H & Luo Y L, *Indian J Chem*, 33A (1994) 603.
- 10 Gutman I, Vidović D & Furtula B, *Indian J Chem*, 42A (2003) 000.
- 11 Gutman I & Vidović D, *Indian J Chem*, 41A (2002) 893.
- 12 Zenkevich I G, *Rus J phys Chem*, 73 (1999) 797.
- 13 Zenkevich I G, *Zh anal Khim*, 55 (2000) 1091
- 14 Zenkevich I G, *Zh org Khim*, 37 (2001) 283.
- 15 Gutman I & Zenkevich I G, *Z Naturforsch*, 57a (2002) 000.
- 16 Gutman I, Vidović D, Furtula B & Zenkevich I G, *J Serb chem Soc*, 68 (2003) 000.
- 17 Gutman I, *Chem Phys Lett*, 364 (2002) 352.
- 18 Randić M, *Chem Phys Lett*, 211 (1993) 478.
- 19 Furtula B & Gutman I, to be published.
- 20 Wilf H, *J Algorithms*, 2 (1981) 204.