# On Interpretation of Well-Known Topological Indices

Milan Randić* and Jure Zupan

Laboratory for Chemometrics, National Institute of Chemistry, Ljubljana, Slovenia

Received July 16, 2000

Many topological indices lack an interpretation in terms of simple physicochemical quantities. We have reexamined the structural interpretation of well-known topological indices: the connectivity index $^1\chi$, the Wiener index W, and the Hosoya topological index Z. We relate the success of various indices in structure−property studies to the degree to which they differentiate contributions from more exposed terminal bonds and more buried interior bonds. When considering bond additive properties of alkanes we find better regressions when greater weights are assigned to terminal CC bonds and lesser weights to internal CC. We suggest here that topological indices be discussed in terms of their partitioning into bond contributions, which for different indices and different bonds will assume different values. With this insight we modified the Wiener index W into a new index W*, in which bond contributions are determined using the reciprocal of the product of the number of atoms on each side of a bond. Similarly we modified the Hosoya index Z into Z* in which the frequency of occurrence of individual CC bonds in the patterns of disjoint bonds are considered. Novel indices are compared with other indices and they show to yield better regressions for boiling points of octane isomers. This suggests a useful classification of topological indices based on the relative magnitudes of the contributions of terminal and interior bonds. To extend such considerations to other indices one needs to consider partitioning of global molecular indices into bond additive terms. A general scheme for partitioning of molecular descriptors into bond contributions is outlined for indices derived from a selection of matrices associated with molecular graphs.

## INTRODUCTION

"In many interesting areas of chemistry we are approaching predictability, but ... I would claim, not understanding." Roald Hoffmann[1]

Mathematical invariants have been employed as molecular descriptors in numerous structure−property and structure−activity studies. They have apparent advantages over alternative QSAR descriptors, such as the hydrophobic and other substituent constants, physicochemical parameters that relate to the electronic effects of a group on a molecule (such as Hammett sigma constants), and the use of molecular properties as descriptors (e.g., molar refractivity). Not only can they be easily computed but also there is a large number of mathematical descriptors to choose from. For example, CODESSA, a software designed by Katritzky, Karelson, and Lobanov,[2] for use in the study of structure versus property or biological activity calculates some 400 molecular descriptors. About half of the descriptors used in CODESSA represent various topological indices, as mathematical invariants are usually referred to.[3−8] However, even though mathematical invariants are well defined, often they are accompanied with an important drawback: a lack of interpretation in terms of simple structural and physicochemical concepts.

Consider, for example, the Wiener number W, which is the first nontrivial structural invariant, proposed over 50 years ago by H. Wiener.[9] As Wiener has shown, this descriptor combined with P (the count of paths of length three) accounts for isomeric variations of many physicochemical properties such as the boiling points of alkanes, alcohols, fatty acids, and such. In the case of paraffins, W was defined as a sum of bond contributions in which the contribution of each bond is given by the product of the number of carbon atoms on one side of the bond and the number of carbon atoms on the other side of the bond. What is the physical interpretation of W? Platt tried to clarify the meaning of W and suggested that it may be related to molecular volume:[10]

"The meaning of Wiener's other parameter W, the "path number", which we might properly call the "Wiener number", is harder to see. It is the sum of the numbers of bonds between all pairs of carbons in the molecular skeleton. It increases for large molecules as $N^3$, where N is the number of carbon atoms in the paraffin. Evidently $\sqrt[3]{W}$ is a sort of mean molecular diameter; and $\Delta W/N^2$, the term in the Wiener formulas, is approximately proportional to the increment in $\sqrt[3]{(W + \Delta W)}$, and represents the increment in mean diameter for a given isomeric change. The $N^2$ acts as a normalizing factor so as to make the $\Delta W$ term for a given type of change almost independent of N."

However, more recently Klein and collaborators[11] have shown that in the case of polymers W increases as $N^{5/2}$ which is not a typical asymptotic behavior for molecular volumes.

The topic of interpretation of topological indices has received little attention. There are several papers that considered the interpretation of selected topological indices.[11−17] These papers represent an exception rather than a rule, which suggests that the interpretation of topological indices may be rather difficult. In this contribution we decided to consider this challenge and will examine interpretation of three well-known topological indices: The connectivity index,[18] the Wiener index,[9] and the Hosoya topological index.[19]

* Corresponding author phone: (515)292-7411; fax: (515)292-8629. Visiting from the Department of Mathematics and Computer Science, Drake University, Des Moines, IA 50311. Current address: 3225 Kingman Rd., Ames, IA 50014.

10.1021/ci000095o CCC: $20.00 © 2001 American Chemical Society
Published on Web 03/31/2001

**Table 1.** The Count of Paths and the Count of Walks for 2,3-Dimethylhexane

| vertex | $p_1$ | $p_2$ | $p_3$ | $p_4$ | $p_5$ |
|---|---|---|---|---|---|
| 1 | 1 | 2 | 2 | 1 | 1 |
| 2 | 3 | 2 | 1 | 1 | 0 |
| 3 | 3 | 3 | 1 | 0 | 0 |
| 4 | 2 | 3 | 2 | 0 | 0 |
| 5 | 2 | 1 | 2 | 2 | 0 |
| 6 | 1 | 1 | 1 | 2 | 2 |
| 7 | 1 | 2 | 2 | 1 | 1 |
| 8 | 1 | 2 | 3 | 1 | 0 |
| sum | 14 | 16 | 14 | 8 | 4 |
| molecule | 7 | 8 | 7 | 4 | 2 |

| vertex | $w_1$ | $w_2$ | $w_3$ | $w_4$ | $w_5$ |
|---|---|---|---|---|---|
| 1 | 1 | 3 | 5 | 12 | 23 |
| 2 | 3 | 5 | 12 | 23 | 51 |
| 3 | 3 | 6 | 13 | 27 | 56 |
| 4 | 2 | 5 | 9 | 20 | 39 |
| 5 | 2 | 3 | 7 | 12 | 27 |
| 6 | 1 | 2 | 3 | 7 | 12 |
| 7 | 1 | 3 | 5 | 12 | 23 |
| 8 | 1 | 3 | 6 | 13 | 27 |
| sum | 14 | 30 | 60 | 126 | 258 |
| molecule | 7 | 15 | 30 | 63 | 129 |

## THE ELEMENTARY STRUCTURAL CONCEPTS

One should not expect that a topological index should necessarily have an interpretation in terms of standard physicochemical concepts such as molecular surface, molecular volume, compactness, flexibility, foldedness, etc. First, observe that all just mentioned aspects of molecular structure themselves are to a degree ambiguous. More importantly, one should recognize that computational quantities, such as topological indices, should have an interpretation *within* the model in which they are used. Hence we should insist on interpretation of topological indices within models of structural chemistry, rather than models of physical chemistry. To appreciate the distinction consider the well-known concepts of MO model such as Coulson bond orders,[20] HOMO−LUMO gap,[21] and atom−atom polarizability[22] and equally well-known concepts of VB model: Pauling bond orders,[23] Kekulé valence structures,[24] and conjugated circuits.[25,26] Each of these concepts has an important meaning within the respective quantum chemistry model, but neither Kekulé valence structures have any simple interpretation in MO model nor has any the HOMO−LUMO gap simple interpretation in VB model!

To proceed we have to select the basic structural concepts that will serve as a basis for interpretation of various topological indices. We propose that *paths* and *walks* to be viewed as the *elementary structural concepts* which are well understood and need no explanation. Paths of length one represent bonds, and paths of length two signify two consecutive bonds. Longer paths represent a string of consecutive bonds in a chemical structure. Paths have been advocated by Platt over 50 years ago as potentially useful molecular descriptors.[27] However, it was not until the late 1970s that paths received some attention in chemical graph theory.[28−35]

Walks represent a sequential list of bonds but differ from paths in allowing repetitions.[36−41] At the top of Table 1 we have listed the count of all paths ($p_i$) for 2,3-dimethylhexane and at the bottom of the same table we have listed the count of all walks ($w_i$) for the same molecule; $p_1$ and $w_1$ are equal to vertex degree. In contrast to the count of paths which is finite and smaller than the count of walks, the count of walks can be extended indefinitely. However, computationally the count of paths is more involved,[42] while the count of walks is computationally simple. Walk counts can be obtained by raising the adjacency matrix to higher powers,[43] or by considering extended valence,[44,45] as illustrated in Figure 1 on 2,3-dimethylhexane. Paths and walks characterize different structural aspects of a molecule. One should not overlook that paths give only the number of neighbors at increasing distances from an atom and do not contain information on the relative distributions of the neighbors. Thus the path code 2, 2, 2 corresponds to three different connectivity patterns, which lead to seven possible embeddings in a plane schematically shown in Figure 2. The count of walks, on the other side, is related to the characteristic polynomial of a graph and its spectrum, a topic related to the occurrence of isospectral graphs.[46−60]

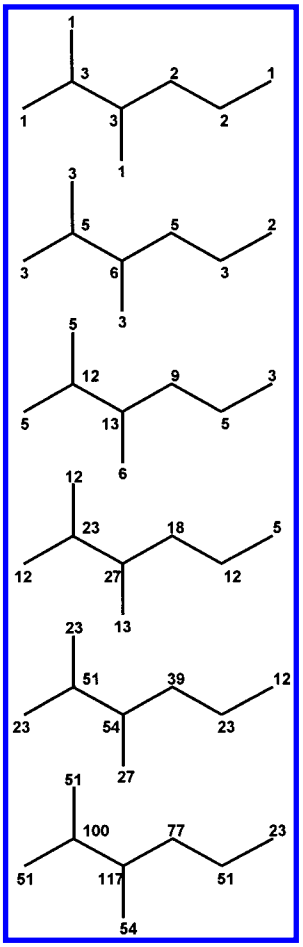


**Figure 1.** Extended valence illustrated for carbon atoms of 2,3-dimethylhexane.

In the following we will try to express the selected molecular descriptors in terms of bond contributions (paths or walks of length one). This will require the partitioning of topological indices into bond contributions, which, as we will see, itself may represent a problem. By accepting paths and walks as the elementary structural concepts, we will see that such an approach will lead us to a novel classification of topological indices that was hitherto completely overlooked, which depends on the relative role of peripheral and interior bonds in a molecule.

In view of the "elementary" status of paths and walks let us point to recently introduced shape descriptors[61,62] $p_i/w_i$,

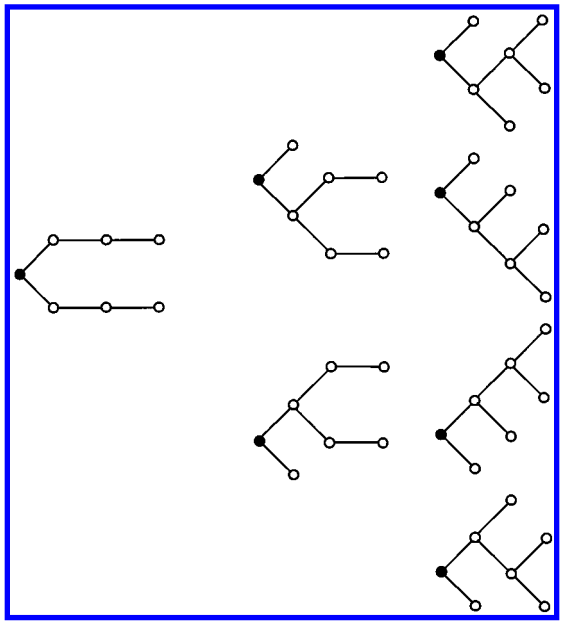

**Figure 2.** Different distribution of the nearest neighbors for a code 2, 2, 2.

**Table 2.** Shape Descriptors Obtained as a Quotient of Path/Walk Counts Illustrated for 2,3-Dimethylhexane

| vertex | $p_1/w_1$ | $p_2/w_2$ | $p_3/w_3$ | $p_4/w_4$ | $p_5/w_5$ | sum |
|---|---|---|---|---|---|---|
| 1 | 1/1 | 2/3 | 2/5 | 1/12 | 1/23 | 2.193478 |
| 2 | 3/3 | 2/5 | 1/12 | 1/23 | 0 | 1.526812 |
| 3 | 3/3 | 3/6 | 1/13 | 0 | 0 | 1.576923 |
| 4 | 2/2 | 3/5 | 2/9 | 0 | 0 | 1.822222 |
| 5 | 2/2 | 1/3 | 2/7 | 2/12 | 0 | 1.785714 |
| 6 | 1/1 | 1/2 | 1/3 | 2/7 | 2/12 | 2.285714 |
| 7 | 1/1 | 2/3 | 2/5 | 1/12 | 1/23 | 2.193478 |
| 8 | 1/1 | 2/3 | 3/6 | 1/13 | 0 | 2.243590 |
| sum | 8 | 4.33333 | 2.30152 | 0.73944 | 0.25362 | 15.62793 |

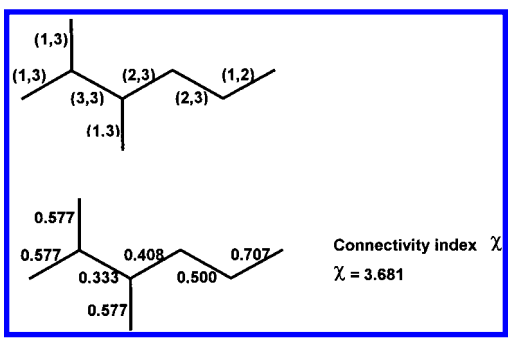


**Figure 3.** The partitioning of the connectivity index of 2,3-dimethylhexane into bond contributions.

where $p_i$ and $w_i$ stand for paths of length i and walks of length i. They are illustrated for 2,3-dimethylhexane in Table 2. It is somewhat surprising that these conceptually so simple indices, which apparently have promise in structure–property-activity studies, have not been discovered earlier.

## THE CONNECTIVITY INDEX[18]

This index is based on differentiation of bonds into (m, n) types, where m, n indicate vertex degrees, i.e., graph theoretical valence of atoms forming the bond in the hydrogen-suppressed graph. For example, 2,3-dimethylhexane has the following bond types: (1, 2), (1, 3), (2, 2), (2, 3), and (3, 3) (Figure 3). In the traditional chemical nomenclature these bond types are written as $CH_3-CH_2$; $CH_3-CH$; $CH_2-CH_2$; $CH_2-CH$; and $CH-CH$. In other molecules as additional types we may have (1, 4), (2, 4), (3, 4), and (4, 4), that is $C-CH_3$; $C-CH_2$; $C-CH$; and $C-C$.

The bond type (m, n) makes the contribution $1/\sqrt{(m\ n)}$ to the magnitude of the connectivity index. As we see from Figure 3 the peripheral bonds make larger contributions to ${}^1\chi$ than the inside bonds. As known, one attributes to peripheral bonds larger bond surface areas, a fact which is reflected in ${}^1\chi$.

The physical interpretation of the connectivity index follows from its close relationship to chemical bonds (paths of length one). The higher order connectivity indices are similarly constructed by considering instead of bonds longer paths (consecutive bonds) in a molecule. The connectivity index[18] and the higher order connectivity indices[63] are nothing but weighted paths, where the selected weighting procedure assigns a dominant role to terminal bonds and a lesser weight to less exposed inner bonds. When heteroatoms are present, some way of discriminating atoms of different kinds needs to be considered. Kier and Hall proposed a particular generalization of the connectivity indices into the valence connectivity indices.[64] Valence connectivity indices have distinct contributions assigned to different heteroatoms, hence it was not surprising that in many cases the valence connectivity indices resulted in a better regression when compared with the regression using the simple connectivity index. However, it is more surprising to find that in several regressions involving molecules having heteroatoms the simple connectivity index offered a better regression than the valence connectivity index. For example, ${}^1\chi$ gives as good a correlation for the molar refraction $R_m$ of ethers as that obtained with ${}^1\chi^v$,[65] and it gives a better regression for van der Waals constants of mixed compounds.[66] According to Kier and Hall[66] "It is typical of properties related to polarizabilities that they do not improve the correlation when only two variables are used." Other properties in which ${}^1\chi$ is a better descriptor than ${}^1\chi^v$ includes molar magnetic susceptibilities of alcohols,[67] heat of vaporization of aliphatic alcohols,[68] boiling points of aliphatic alcohols,[69] boiling points of aliphatic ethers,[70] the boiling points of primary and secondary aliphatic amines,[71] solubility of aliphatic alcohols,[72] partition coefficient of aliphatic ethers,[73] and so on.[74] The list extends into the structure–activity area also, as is illustrated by the toxicity of ethers.[75,76] It is needless to say that there are many structure–property-activity relationships involving heteroatoms in which ${}^1\chi^v$ is a better descriptor that ${}^1\chi$.

This paradoxical situation that the connectivity index ${}^1\chi$ can produce better (even if not satisfactory) simple regressions for compounds having heteroatoms than ${}^1\chi^v$ apparently has not attracted much attention in the past. However, only recently the situation became better understood. Introduction of the variable connectivity indices[77,78] that itself was overlooked for almost 10 years has shown that different properties require different optimal weights.[79–84] It was observed that in some cases the weights assigned to the valence connectivity indices depart more from the optimal weights than the "nonweighted" simple connectivity index. For example, Randić and Dobrowolski[84] found that optimal weights for correlation of the boiling points of amines are obtained when the diagonal entries in the adjacency matrix assume values $x = 1.25$ and $y = -0.65$ for carbon and nitrogen atoms, respectively. As a consequence the variable

**Table 3.** The Distance, Matrix, the Reversed Distance Matrix, and the Reciprocal Distance Matrix Illustrated for 2,3-Dimethylhexane

| | | | | Distance Matrix | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | row sum |
| 1 | 0 | 1 | 2 | 3 | 4 | 5 | 2 | 3 | 20 |
| 2 | | 0 | 1 | 2 | 3 | 4 | 1 | 2 | 14 |
| 3 | | | 0 | 1 | 2 | 3 | 2 | 1 | 12 |
| 4 | | | | 0 | 1 | 2 | 3 | 2 | 14 |
| 5 | | | | | 0 | 1 | 4 | 3 | 18 |
| 6 | | | | | | 0 | 5 | 4 | 24 |
| 7 | | | | | | | 0 | 3 | 20 |
| 8 | | | | | | | | 0 | 18 |
| | | | | Reverse Distance Matrix | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 1 | row sum |
| 1 | 0 | 7 | 6 | 5 | 4 | 3 | 6 | 5 | 36 |
| 2 | | 0 | 7 | 6 | 5 | 4 | 7 | 6 | 42 |
| 3 | | | 0 | 7 | 6 | 5 | 6 | 7 | 44 |
| 4 | | | | 0 | 7 | 6 | 5 | 6 | 42 |
| 5 | | | | | 0 | 7 | 4 | 5 | 38 |
| 6 | | | | | | 0 | 3 | 4 | 32 |
| 7 | | | | | | | 0 | 5 | 36 |
| 8 | | | | | | | | 0 | 38 |
| | | | | Reciprocal Distance Matrix | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | row sum |
| 1 | 0 | 1 | 1/2 | 1/3 | 1/4 | 1/5 | 1/2 | 1/3 | 3.11667 |
| 2 | | 0 | 1 | 1/2 | 1/3 | 1/4 | 1 | 1/2 | 4.58333 |
| 3 | | | 0 | 1 | 1/2 | 1/3 | 1/2 | 1 | 4.83333 |
| 4 | | | | 0 | 1 | 1/2 | 1/3 | 1/2 | 4.16667 |
| 5 | | | | | 0 | 1 | 1/4 | 1/3 | 3.66667 |
| 6 | | | | | | 0 | 1/5 | 1/4 | 2.73333 |
| 7 | | | | | | | 0 | 1/3 | 3.11667 |
| 8 | | | | | | | | 0 | 3.25000 |

connectivity indices for amines are larger than the corresponding $^1\chi$ indices in which nitrogen and carbon are not differentiated. In contrast, according to Kier and Hall, the simple connectivity index has to be decreased for nitrogen in primary amines by about 0.42 (in secondary amines by 0.30 and in tertiary amines by 0.24).[85] Hence in this case the valence connectivity index was shifted in the "wrong" direction so to speak, and in such situations the simple connectivity $^1\chi$ makes a better regression than $^1\chi^v$, even though the regression need not be satisfactory and additional descriptors may be invoked.

## WIENER INDEX

The Wiener index, W, was introduced as one of two descriptors to describe variations in the number of physicochemical properties of alkanes.[9,86−89] The other parameter, P, represented the count of paths of length three and was to account for the crowding of methyl groups in highly branched paraffins. Wiener offered a simple algorithm for calculating W:[9] For each bond multiply the number of carbon atoms on each side of the bond and add all bond contributions. Hosoya[19] has pointed out that W can be obtained from the distance matrix as the sum of matrix elements above the main diagonal. In Table 3 we show the distance matrix for 2,3-dimethylhexane, from which one obtains W = 70.

For acyclic molecules W can be obtained alternatively as the dot product of the path vector P and the path length vector L = (1, 2, 3,...).[90] For 2,3-dimethylhexane P = (7, 8, 7, 4, 2, 0, 0), the components giving the number of path of length one, two, three, etc.:

$$W = \langle P|L\rangle = 7 \times 1 + 8 \times 2 + 7 \times 3 + 4 \times 4 + 2 \times 5 = 70$$

The Wiener number W represents a global molecular index. As we see more distant atom pairs make larger contribution to W than adjacent atom pairs. Despite this counterintuitive aspect W appears to be a very good descriptor for numerous molecular properties. How can this be understood?

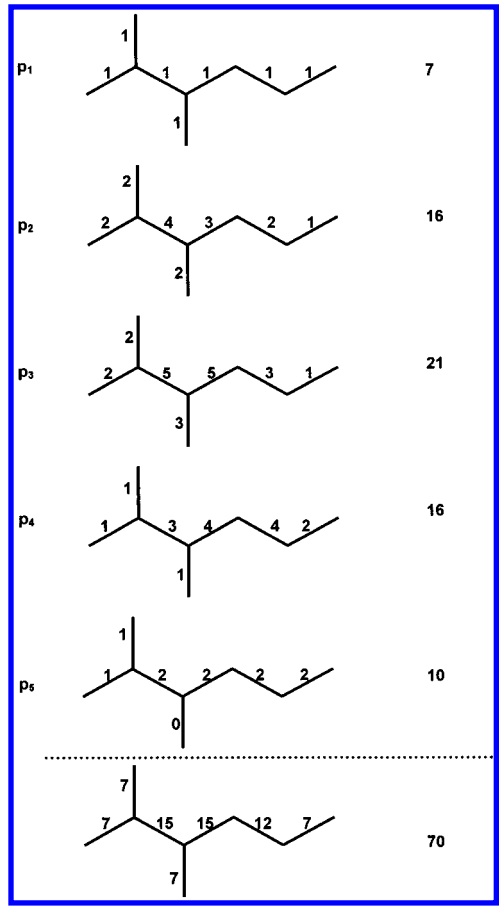


**Figure 4.** Participation of individual CC bonds of 2,3-dimethylhexane in paths of different length.

To interpret W we need to partition the Wiener index into bond contributions. This is not difficult in view of the algorithm proposed by Wiener to calculate W. Each bond makes a contribution equal to the product of the number of atoms on each side of the bond considered. For 2,3-dimethylhexane we obtain

$$W = 1 \times 7 + 1 \times 7 + 1 \times 7 + 1 \times 7 + 3 \times 5 + 3 \times 5 + 2 \times 6$$

$$W = 7 + 7 + 7 + 7 + 15 + 15 + 12 = 70$$

The above partitioning of W is illustrated in Figure 4. Observe that now, contrary to what was found for the connectivity index, the central CC bonds in 2,3-dimethylhexane make greater contributions, i.e., are assigned more important role in governing molecular properties.

## THE REVERSED WIENER INDEX M

Randić[90] has observed an interesting and hitherto overlooked aspect of structural characterization of octane isomers related to the Wiener index. Consider vector L* = (7, 6, 5, 4, 3, 2, 1), which in comparison to vector L has reversed weights: the shortest paths gets the largest weight, etc. The dot product, $\langle P|L^*\rangle = 7 \times 7 + 8 \times 6 + 7 \times 5 + 4 \times 4 + 2 \times 3 = 154$, defines a novel topological index, that we

will label as M (an inverted W). Other alternative weights are possible. Recently Balaban and co-workers[91,92] considered as the weights the difference between the graph diameter (which is given by the longest path in a molecule) and the distance. Such a choice for weighting also reverses the role of the nearest neighbors and the more distant neighbors.

One may expect that by reversal of the weight the novel descriptor M, associated with the reversed distance matrix (shown in the middle part of Table 3), may show a better regression when compared to W. When restricting comparison to isomers unexpectedly one finds that W and M are strictly collinear. That being the case we may speculate that the success of the Wiener number as molecular descriptor may well be due to its parallelism with M, the index which is defined by reversing the weights of close and distant neighbors in a molecule, rather than to a model in which distant neighbors play a greater role (as implied by the distance matrix).

However, on the other hand, at least in the case of isomers when both M and W are collinear, we should expect that partitioning of M will show considerable parallelism with that of W, resulting in a model which again gives to terminal CC bonds a lesser weights. To test this we have to partition M to bond contributions, but how are we to obtain a partitioning of M? This in itself presents a problem, because there is no simple algorithm that immediately allows one to obtain individual bond contributions to M. Because both W and M can be defined through the scalar product of vectors, to find a partitioning for M we will reconsider the partitioning of W using the definition of W as scalar product of path vectors P and L. In Figure 4 we illustrated in the case of 2,3-dimethylhexane for paths of different length contributions of bonds involved. Bonds which are involved in only a single path obtain the value 1. Bonds which are involved in two paths obtain the value 2, etc. Hence the number associated with each bond indicates in how many paths the bond was included as a component. Clearly when counting paths of length one each bond is involved just once. For example, when considering paths of length two the terminal bonds involving carbon atoms 1, 7, and 8 are each present in two paths. For example, bond 1−2 is involved in paths 1−2−3 and 1−2−7. The CC bond 2−3 is involved in four paths of length two (1−2−3; 2−3−4; 2−3−8; and 7−2−3) and so on. The numbers at the right, which are obtained by adding all bond numbers shown on each diagram, give the partitioning of W for contributions of paths of different length. If, however, we add contributions arising from paths of different length for individual bonds we obtain partitioning of W by individual bonds, as is shown in the molecular diagram at the bottom of Figure 4.

To partition M into bond contributions we duplicate the process just outlined but multiplying bond numbers by appropriate weights. The paths of length one are multiplied by 7, the paths of length two by 6/2, the paths of length three by 5/3, the paths of length four by 4/4, and the paths of length five by 3/5. In the quotient giving the contribution per bond, the denominator 2, 3, 4, 5 arises because the length of paths increases. By adding all the bond contributions we obtain the partition of M associated with paths of different length shown at the right side of each diagram. By adding all contributions for individual bonds we obtain the sought bond partitioning of M, shown at the bottom of Figure 5.

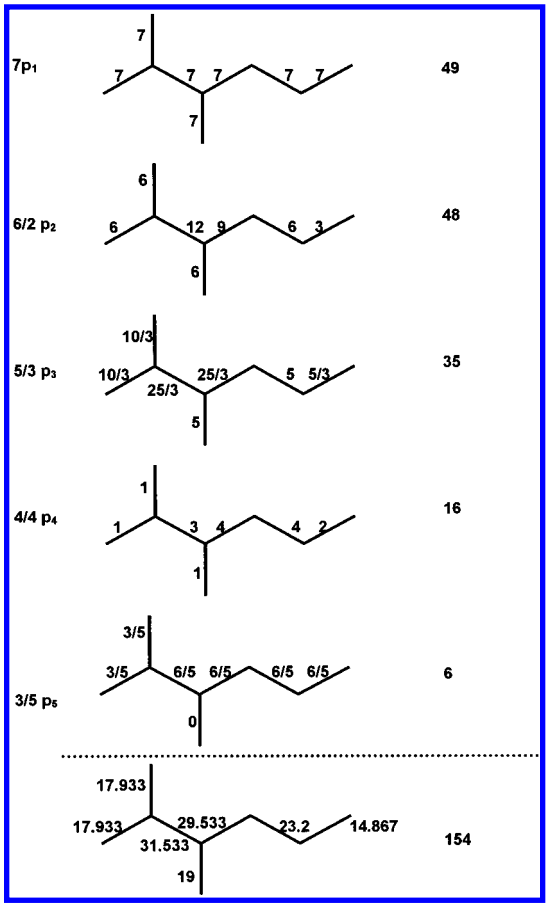


**Figure 5.** Participation of individual CC bonds of 2,3-dimethylhexane in paths of different length for reversed Wiener number M.

As we see central (interior) CC bonds have a greater weight than the terminal CC bonds, just as was the case with W.

## HARARY INDEX H

To bypass the apparent disadvantage of the distance matrix Plavšić and co-workers[93] constructed as an alternative the reciprocal distance matrix (shown in Table 3); the same matrix was constructed independently by Balaban and co-workers.[94] It was expected that by giving a lesser weight to more distant atoms one will arrive at a modification of the Wiener index that will show better regression capabilities when compared to the Wiener index. The new index H was named "Harary index," to honor Frank Harary who has introduced the distance matrix into Graph Theory.[94] In Figure 6 we illustrate the construction and the partitioning of the Harary index, for which again, perhaps unexpectedly, we obtain smaller contributions for terminal bonds than for internal bonds. Again consider separately contributions arising from paths of different length as was illustrated previously on W. Then multiply the corresponding bond contributions by $(1/2)^2$, $(1/3)^2$, $(1/4)^2$, and $(1/5)^2$, respectively. The first factor 1/m is determined by the length of the path in order to obtain contribution per bond. The second factor 1/m comes from the assumed reciprocal dependence on distance.

Hence, as we have seen the topological indices W, M, and H derived from the distance matrix, the reversed distance matrix, and the reciprocal distance matrix, respectively, give more weight to internal CC bonds instead of terminal CC

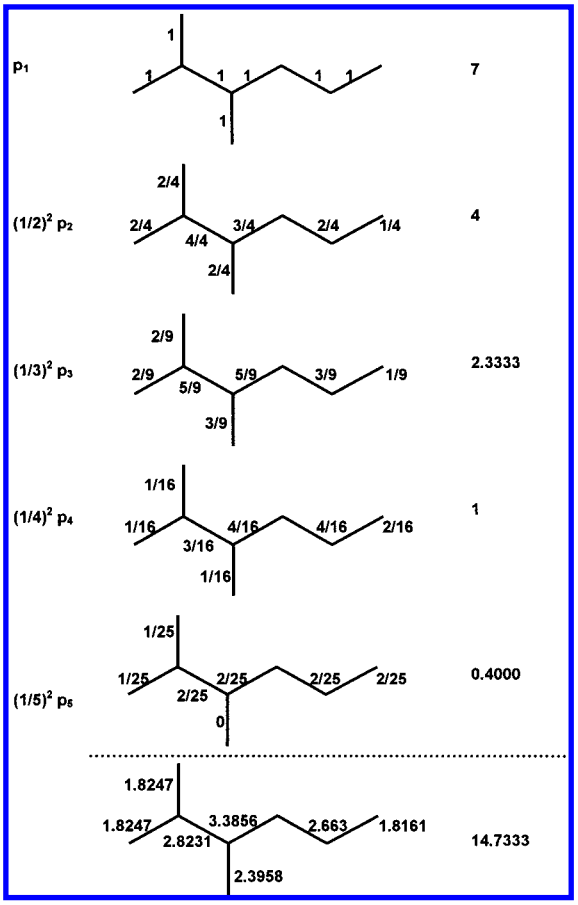


**Figure 6.** Participation of individual CC bonds of 2,3-dimethylhexane in paths of different length for Harary index H.

bonds. This happens despite the fact that in the reversed and the reciprocal distance matrices one gives greater weight to the nearest neighbors.

## NOVEL DISTANCE RELATED TOPOLOGICAL INDEX

Because the connectivity index gives very good regressions for numerous bond additive molecular properties, it seems desirable to construct an index derived from the distance considerations that will show the same pattern after partition into bond contributions as the connectivity index. Can this be accomplished?

The idea of Plavšić, Balaban, and co-workers[93,94] of considering reciprocal distance matrix appears attractive, but as we have seen it does not produce the desired result. However, we can modify the Wiener index W using the idea of the reciprocal distances by applying it directly to the algorithm for calculating W as originally stated by Wiener and not considering the distance matrix. According to Wiener one obtains W as follows: "For each bond multiply the number of carbon atoms on each side of the bond, and add all bond contributions". If we change "the number of atoms" by "the reciprocal of the number of atoms" we will reverse the role of the near and the distant neighbors. Hence, a novel Wiener-type topological index, for which we will use label W*, is obtained from the following algorithm:

"For each bond multiply the *reciprocal* of the number of carbon atoms on each side of the bond, and add all bond contributions."

**Table 4.** Values of the Novel Wiener-type Index W* Obtained by Summing the Bond Contributions Obtained by Multiplying the Reciprocals of Number Atoms on Each Side of the Bond Calculated for Octane Isomers

| isomer | W* | isomer | W* | isomer | W* |
|---|---|---|---|---|---|
| n | 0.64821 | $2,3M_2$ | 0.78810 | 3M3E | 0.82143 |
| 2M | 0.70774 | $2,4M_2$ | 0.78393 | $2,2,3M_3$ | 0.86012 |
| 3M | 0.72440 | $2,5M_2$ | 0.76726 | $2,2,4M_3$ | 0.84345 |
| 4M | 0.72857 | $3,3M_2$ | 0.80476 | $2,3,3M_3$ | 0.86429 |
| 3E | 0.74524 | $3,4M_2$ | 0.80060 | $2,3,4M_3$ | 0.84762 |
| $2,2M_2$ | 0.78393 | 2M3E | 0.80476 | $2233M_4$ | 0.91964 |

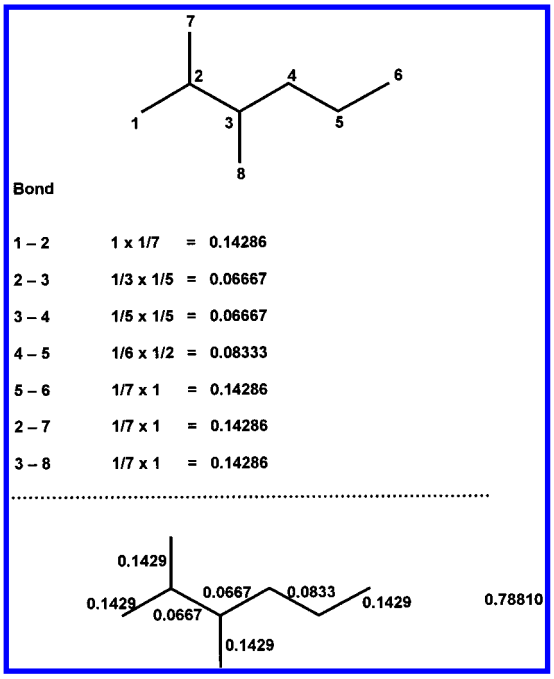


**Figure 7.** Participation of individual CC bonds of 2,3-dimethylhexane in paths of different length for novel Wiener-type index W*.

**Table 5.** Comparison of the Two Variable Regressions of the Boiling Points of Octanes for W, H, H′, and W* Each Combined With P[a]

| descriptors | coeff *r* | std error *s* | Fisher ratio *F* |
|---|---|---|---|
| W, P | 0.8887 | 2.987 | 28.2 |
| H, P | 0.8892 | 2.982 | 28.3 |
| H′, P | 0.8899 | 2.973 | 28.6 |
| W*, P | 0.9394 | 2.234 | 56.3 |

[a] The paths of length three.

In Table 4 we have listed the W* values for the 18 isomers of octane. In Figure 7 we show partitioning of W* into bond contributions for 2,3-dimethylhexane. As we see, the terminal CC bonds now have a greater weight than the internal CC bonds, a property of the partitioning that we wanted. Hence, W* and the connectivity index $^1\chi$ belong to the same class of descriptors that give preference to the role of the terminal bonds. We anticipate therefore that W* will perform better as a descriptor of a bond additive property than W, M, or H such as the boiling points of alkanes. That indeed this is the case is illustrated in Table 5 in which we compare two-variable correlation for W, H, and H′ against W*. H′ is the modified Harary index. In all cases the second descriptor was P, the path of length three. We do not show the correlation for the pair of descriptors (M, P) because their regression is collinear with that of the pair (W, P). As we

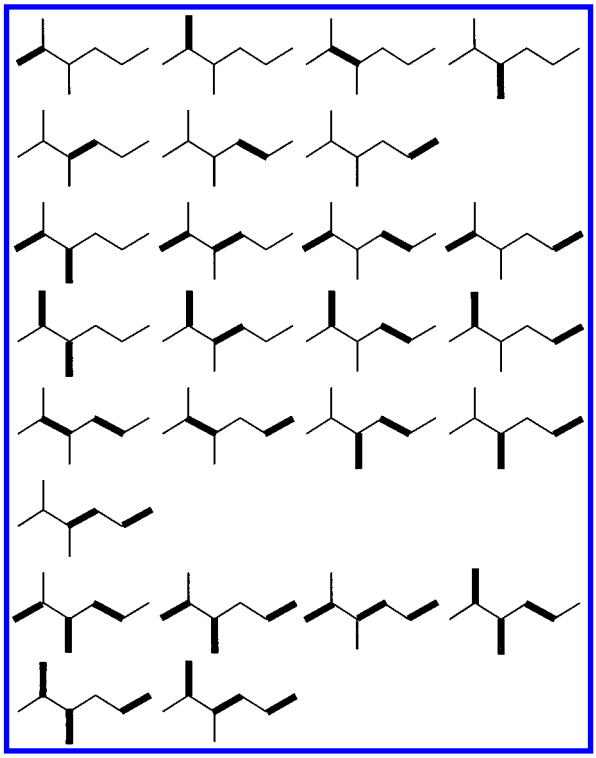

**Figure 8.** The patterns involving one, two, and three disjoint bonds for 2,3-dimethylhexane.

see W, H, and the modified Harary index H′ all give approximately equal quality regressions, a fact which is reflected in approximately equal statistical parameters (*r*, *s*, and *F*). The only visible improvement occurs when W* replaces W, H, or H′. This clearly shows the importance of terminal bonds in bond additivities for certain selected physicochemical properties, such as the boiling points of octane isomers (and alkanes in general).

## THE HOSOYA TOPOLOGICAL INDEX Z

In 1971 Hosoya introduced the topological index Z defined by the count of all possible patterns of considering k disjoint bonds in a molecule:[19]

$$Z = 1 + Z_1 + Z_2 + \ldots Z_k$$

Here $Z_1$ counts the number of bonds in a graph, $Z_2$ counts the pairs of disjoint bonds, $Z_3$ counts triplets of disjoint bonds, etc. In Figure 8 we have illustrated the patterns involving one, two, and three disjoint CC bonds for 2,3-dimethylhexane. Alternatively, the results can be presented also by a polynomial:

$$Z(x) = 1 + Z_1x + Z_2x^2 + \ldots + Z_kx^k$$

We can refer to the leading +1 as $Z_0$, formally representing the number of ways in which no bond is selected. In the case of 2,3-dimethylhexane we have

$$Z = 1 + 7 + 13 + 6 = 27$$

or

$$Z(x) = 1 + 7x + 13x^2 + 6x^3$$

The Z(x) counting polynomial has the apparent property that when $x = 1$, one obtains Z, i.e., $Z = Z(x = 1)$. In Table 6 we give *Z* values for 18 isomers of octane and their partitioning into $Z_1$, $Z_2$, $Z_3$, and $Z_4$ ($Z_1 = 7$, is constant for all octane isomers, and is not included in the table).

**Table 6.** Partitioning of the Hosoya Z Topological Index Into Contributions Arising From Two, Three, and Four Disjoint Bonds[a]

| isomer | $Z_2$ | $Z_3$ | $Z_4$ | Z | Z* |
|---|---|---|---|---|---|
| n | 15 | 10 | 1 | 34 | 72 |
| 2M | 14 | 7 | 0 | 29 | 57 |
| 3M | 14 | 8 | 1 | 31 | 64 |
| 4M | 14 | 8 | 0 | 30 | 60 |
| 3E | 14 | 9 | 1 | 32 | 67 |
| $2,2M_2$ | 12 | 3 | 0 | 23 | 41 |
| $2,3M_2$ | 13 | 6 | 0 | 27 | 52 |
| $2,4M_2$ | 13 | 5 | 0 | 26 | 49 |
| $2,5M_2$ | 13 | 4 | 0 | 25 | 46 |
| $3,3M_2$ | 12 | 5 | 0 | 25 | 47 |
| $3,4M_2$ | 13 | 7 | 1 | 29 | 59 |
| 2M3E | 13 | 7 | 0 | 28 | 55 |
| 3M3E | 12 | 7 | 1 | 28 | 57 |
| $2,2,3M_3$ | 11 | 3 | 0 | 22 | 39 |
| $2,2,4M_3$ | 11 | 0 | 0 | 19 | 30 |
| $2,3,3M_3$ | 11 | 4 | 0 | 23 | 42 |
| $2,3,4M_3$ | 12 | 4 | 0 | 24 | 44 |
| $2233M_4$ | 9 | 0 | 0 | 17 | 26 |

[a] For all isomers $Z_1 = 7$.

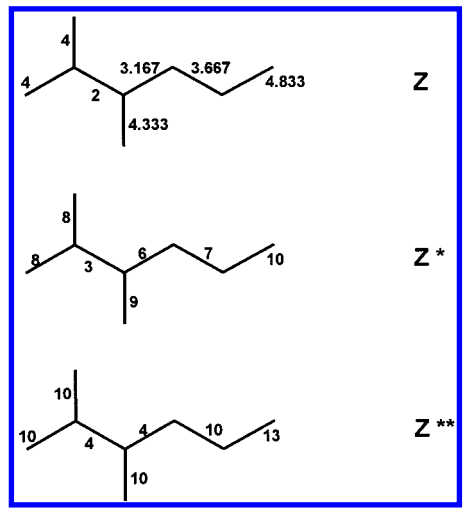


**Figure 9.** Participation of individual CC bonds of 2,3-dimethylhexane in paths of different length for Hosoya Z topological index.

The Hosoya index is a global molecular index, which has no apparent bond partition. To arrive at bond decomposition of Z we will superimpose all the diagrams of Figure 8. We obtain a weighted molecular graph shown at the middle of Figure 9. The numbers assigned to each bond give the total number each bond occurs as a disjoint component in all patterns of Figure 8. For example, the weight for bond 1−2 is obtained as a sum of 1 + 4 + 3 = 8, the bond 1−2 contributing once to $Z_1$, appearing four times in the patterns for $Z_2$, and appearing three times in the pattern involving three disjoint bonds ($Z_3$). Observe however, if we now add all the bond weights (to which +1 has to be added as contribution of $Z_0$), we obtain 1 + 51, not 1 + 26 (Z = 27 is the value for 2,3-dimethylhexane). Hence, the above superposition does not represent the partitioning of Z but a partitioning of another topological index, to be labeled Z*, which is defined precisely by the outlined superposition of all $Z_k$ patterns.

To obtain a partitioning of Z we have to modify the contributions of $Z_1$, $Z_2$, and $Z_3$, by dividing the respective counts by 1, 2, and 3 in order to obtain the participation of individual bonds in each type of patterns. Indeed, when this modification is introduced, we obtain as the weight for the bond 1−2 instead of 1 + 4 + 3 the sum: 1 + 4/2 + 3/3 = 4. As we can see from Figure 9 (the upper part) the new contributions now add to 26, hence they represent the partitioning of Z, which only have to be augmented by 1, or 1/7 per bond, to account for the presence of $Z_0$. The obtained relative contribution of different bonds to Z agree with the partitioning of the connectivity index $^1\chi$: the dominant role to the terminal CC bonds and assigning a lesser contributions to the interior bonds.

## MODIFIED HOSOYA INDEX Z*

The Hosoya polynomial

$$Z(x) = 1 + Z_1x + Z_2x^2 + ... + Z_kx^k$$

can be generalized to a linear function of several variables

$$Z(x) = 1 + Z_1x_1 + Z_2x_2 + ... + Z_kx^k$$

in which powers of x are treated as independent variables. For 2,3-dimethylhexane this gives

$$Z(x, y, z) = 1 + 7x + 13y + 6z$$

If we assume x = 1, y = 2, and z = 3, we obtain the special case of the generalized Hosoya index, the Z(1, 2, 3) = 1 + 7 × 1+ 13 × 2 + 6 × 3, or briefly Z*, already mentioned in Figure 9. This new index, which gives the frequency of occurrence of individual CC bonds in the superposition of the patterns of disjoint bonds, has already been illustrated in the middle of Figure 9. The integer weights give the number of times each bond has appeared in *all* disjoint bond patterns to which again we have to add 1, or 1/7 per bond, to account for $Z_0$. In comparison of the partitioning of Z and Z* we see that for the latter terminal CC bonds have even somewhat greater weight than for Z. Which are the optimal weights may depend on the property considered, as has been recently found with the use of variable paths and variable connectivity indices.[77−80] If we consider the boiling points of octane isomers, as can be seen below, Z* shows a somewhat better statistic than Z:

| | | | |
|---|---|---|---|
| Z | $r = 0.888$ | $s = 2.90$ °C | $F = 60$ |
| Z* | $r = 0.908$ | $s = 2.65$ °C | $F = 75$ |

Observe that both descriptors in a one-variable regression give a better regression than the regressions using two descriptors based on the distance matrix including W, H, and H′ combined with P. The regression using W* and P gives a better standard error than the regression using Z but a lesser *F* value (because *F* is more sensitive to the number of descriptors used in a regression than the standard error *s*).

## MORE GENERAL Z-TYPE TOPOLOGICAL INDICES

Z* is just a special case of Z(x, y, z) index in which the weights have taken simple integer values $x = 1$; y = 2; and z = 3. As we have seen, these particular integer values offer a straightforward interpretation for the bond contributions to Z* as a superposition of all patterns of disjoint bonds in a molecular graph. In most general situations, however, one can view x, y, z as variables and write $Z = 1 + xZ_1 + yZ_2 + zZ_3$. In view that we have relatively small set of data points ($n = 18$) we will consider only Z as a linear function of two variables: $Z = 1 + xZ_1 + yZ_2$. In Table 7 we give the results of a regression of BP = f ($Z_1$, $Z_2$) and for comparison also BP = f(Z, Z*). As we can see, the optimal solution yields a standard error below 2 °C.

**Table 7.** Correlations of Z and Various Combinations of Z(x, y, z) in One- and Two-Variable Regressions of the Boiling Points of Alkanes

| descriptor | coeff *r* | std error *s* | Fisher ratio *F* | eq |
|---|---|---|---|---|
| Single Variable | | | | |
| $Z = Z_1 + Z_2 + Z_3$ | 0.888 | 2.90 °C | 60 | 1 |
| $Z^* = Z_1 + 2Z_2 + 3Z_3$ | 0.908 | 2.65 °C | 75 | 2 |
| $Z^{**} = -2Z_1 + 3Z_2$ | 0.955 | 1.88 °C | 163 | 3 |
| Two Variables | | | | |
| Z, Z* | 0.921 | 2.54 °C | 42 | 4 |
| $Z_1$, $Z_2$ | 0.954 | 1.94 °C | 77 | 5 |

| the regression equations | eq |
|---|---|
| $BP = 1.2211\ Z + 81.6939$ | 1 |
| $BP = 0.7956\ Z^* + 96.6615$ | 2 |
| $BP = -0.9762\ Z^{**} + 122.4462$ | 3 |
| $BP = -2.1515\ Z + 1.9162\ Z^* + 106.6110$ | 4 |
| $BP = -2.0395\ Z_1 + 2.9641\ Z_2 + 123.3487$ | 5 |

It is interesting to notice that the coefficients of the optimal regression equation are so close to integers that a linear combination $-2Z_1 + 3Z_2$ offers an outstanding simple regression with a standard error of 1.88 °C. Some caution is called for because this regression was constructed *a posteriori*, being based on the optimal two-variable regression. Nevertheless, because it does produce an outstanding regression it is of interest to see how the molecular descriptor obtained as a combination Z(−2, 3, 0) is partitioned into bond contributions. At the bottom of Figure 9 we show the outcome: Again the terminal bonds have greater weights than the internal bonds. Besides terminal methyl groups also a terminal ethyl group plays a dominant role in bond additive properties.

The question can be raised how is it possible that Z**, which is a particular combination of $Z_1$ and $Z_2$, is associated with a smaller standard error ($s = 1.88$ °C) and larger *F* value ($F = 163$) than the optimal linear combination of $Z_1$ and $Z_2$ ($s = 1.94$ °C and $F = 77$). The reason lies in the definition of the standard error (and which is also reflected in the Fisher ratio), which includes the number of free variables as a parameter. Thus by increasing the number of variables the standard error may increase in comparison with an outstanding regression using a single variable.

## ON CLASSIFICATION OF TOPOLOGICAL INDICES

As we have seen, different topological indices after partitioning into bond contributions imply different behavior. The connectivity index, the Hosoya index Z, the here introduced Z*, and the index W* also introduced here (which is based on Wiener algorithm using reciprocal function dependence) all give to terminal bonds greater weights than to the interior CC bonds. In contrast, the Wiener number W

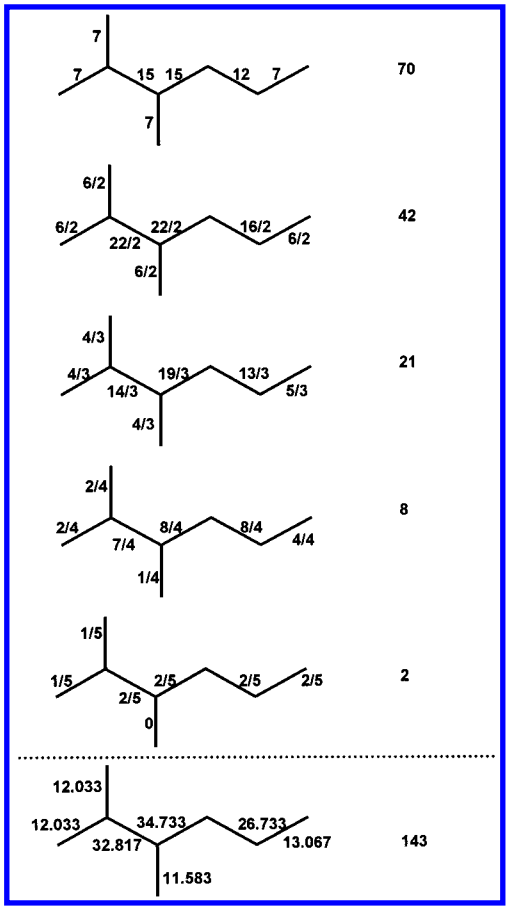


**Figure 10.** Participation of individual CC bonds of 2,3-dimethylhexane in paths of different length for the hyper-Wiener-type index WW.

**Table 8.** Classification of the Topological Indices Into Two Groups Depending on the Relative Magnitudes of the Terminal and Internal CC Bond Contributions in the Partitioning of the Indices into Bond Contributions

| class I | class II |
|---|---|
| the connectivity index $^1\chi$ | the Wiener index W |
| the Hosoya topological index Z | the reversed Wiener index M |
| the modified Hosoya index Z* | the Harary index H |
| the modified Wiener index W* | the modified Harary index H′ |
| | the hyper-Wiener index WW |

and the Harary indices H and H′ show the opposite behavior: They give more weights to internal CC bonds and lesser weights to the terminal CC bonds. To this group we can also add the hyper-Wiener index,[95−97] the partition of which into bond contributions is illustrated in Figure 10. As we have seen, for bond additive properties such as boiling points the former class of descriptors offers regressions characterized by better statistics. For some other properties the opposite may be the case—a task outside the scope of the present work but definitely worthy of attention. In Table 8 we give a list of molecular descriptors discussed in this paper classified into the two groups: those giving higher weights to terminal bonds (including terminal ethyl group) and those giving higher weights to internal CC bonds.

To classify an index as "the connectivity type" or "the Wiener type" one needs first to partition an index into its bond additive contributions. If an index is bond additive, as is the case with the connectivity index, the Balaban J index,[98,99] and indices derived from row sums of graph matrix

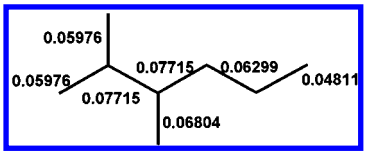


**Figure 11.** Participation of individual CC bonds of 2,3-dimethylhexane in paths of different length for Balaban index J.

by using the algorithm $\Sigma\, 1/\sqrt{(R_i\, R_j)}$, where $R_i$ and $R_j$ are the row sums corresponding to adjacent vertices i, j, and summation is over all bonds, the partitioning in bond contributions is trivial. If an index is defined as an average matrix element, the average row sum, or sum of all matrix elements above the main diagonal of a matrix, as is the case with the Wiener index, partitioning can be accomplished by considering contributions from paths of different length. It is possible that an index does not belong to either of the two classes, as is the case with the Balaban index J, which is bond additive (Figure 11). Here the contributions of the terminal bonds, which are $1/\sqrt{280}$, $1/\sqrt{280}$, $1/\sqrt{432}$, and $1/\sqrt{216}$ for bonds 1−2, 2−7, 5−6, and 3−8, respectively, almost overlap in magnitude with contributions of the interior bonds 2−3, 3−4, and 4−5 which are respectively $1/\sqrt{168}$, $1/\sqrt{168}$, and $1/\sqrt{252}$.

For correlation of the boiling points of alkanes it is desirable that terminal bonds would make a greater contribution to the construction of molecular descriptors. If descriptors are constructed from row sums of a graph matrix, then we may have three cases: (1) the row sums belonging to terminal bonds are smaller than the row sums belonging to internal CC bonds (as is the case with the adjacency matrix); (2) the row sums belonging to terminal bonds are greater than the row sums belonging to internal CC bonds (as is the case with the distance matrix); and (3) the row sums belonging to terminal bonds and the row sums belonging to internal CC bonds have "mixed" weights (as is the case with Balaban's J index). To obtain descriptors of desired quality if terminal bond weights are larger than the weights of internal bonds one should use the algorithm $\Sigma\, 1/\sqrt{(R_i\, R_j)}$, the summation to be taken over all bonds. To obtain descriptors of desired quality of terminal row sums that are smaller than the weights of internal bonds one should use the algorithm $\Sigma\, \sqrt{(R_i\, R_j)}$, where the summation is again over all bonds.

We may add that one can consider a classification based not on the dominance on terminal or interior bonds but on the relative magnitudes of the average row sums for terminal bonds and average row sums for interior bonds. In that case Balaban's J index, which according to the above approach does not offer a clear-cut answer, would belong to class II, together with the Wiener index. As an additional flexibility of the approach if partitioning of an invariant introduces apparently too large of differences in the contributions of the terminal and the internal bonds we may consider suitable modifications. This may be the case with WW and WW* (Figure 10 and Figure 12, respectively). In that case one can use the square root function and obtain a modified topological index $\sqrt{WW^*}$ which will show a lesser variation in bond contribution than WW*. In a similar fashion one can reconsider W* and instead of using reciprocals of products one can take reciprocals of the squares of products, etc., in an effort to find an optimal form of W* type topological index.

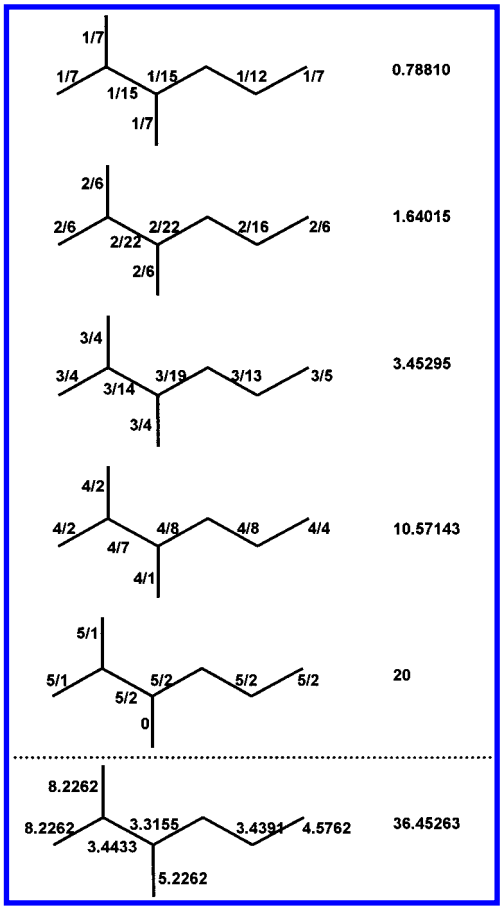


**Figure 12.** Participation of individual CC bonds of 2,3-dimethylhexane in paths of different length for novel hyper-Wiener index WW*.

Finally, we may add that the outlined procedure for construction of the index W* equally applies to the construction of modified Szeged index,[98−100] which represents a generalization of Wiener index to cyclic and polycyclic systems. The Szeged index is obtained as a bond additive quantity where bond contributions are given as the product of the number of atoms closer to each of the two end points of each bond. To obtain modified index Sz*, instead of multiplying the number of atoms closer to each end atom of a bond, we have to take products of the reciprocal number of atoms closer to each of the end atoms of a bond.

## ACKNOWLEDGMENT

The authors would like to thank Professor Ch. Rücker and Professor A. T. Balaban for critical reading and numerous suggestions that resulted in improvement of the manuscript.

## REFERENCES AND NOTES

(1) Hoffmann, R. Qualitative thinking in the age of modern computational chemistry − or what Lionel Salem knows. *J. Mol. Struct. (THEOCHEM)* **1998**, *424*, p 3.

(2) CODESSA (Comprehensive Descriptors for Structures and Statistical Analysis) reports most of the descriptors designed by 1995. Katritzky, A. R., Lobanov, V., Karelson, M., Eds.; CODESSA (Comprehensive Descriptors for structural and Statistical Analysis), University of Florida: Gainesville, FL.

(3) Trinajstić, N. *Chemical Graph Theory*; CRC Press: Boca Raton, FL, 1992; Chapter 10, pp 225−273.

(4) Balaban, A. T. *Historical developments of topological indices*, in: *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach: Amsterdam, The Netherlands, 1999; pp 403−453.

(5) Randić, M. Topological Indices. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, 1998; pp 3018−3032.

(6) Basak, S. C. *Information theoretic indices of neighborhood complexity and their applications*. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T, Eds.; Gordon and Breach: Amsterdam, The Netherlands, 1999; pp 563−593.

(7) Randić, M.; Novič, M.; Vračko, M. *Molecular Descriptors, New and Old*; Lecture Notes in Chemistry, Submitted for publication.

(8) Randić, M. On characterization of chemical structure. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 672−687.

(9) Wiener, H. Structural determination of paraffin boiling points. *J. Am. Chem. Soc.* **1947**, *69*, 17−20.

(10) Platt, J. R. Prediction of isomeric differences in paraffin properties. *J. Phys. Chem.* **1952**, *56*, 328−336.

(11) Bytautas, L.; Klein, D. J. Mean Wiener numbers and other mean extensions for alkane trees. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 471−481−196.

(12) Motoc, I.; Balaban, A. T. Topological indices: Intercorrelations, physical meaning, correlation ability. *Rev. Roum. Chim.* **1981**, 26*5*, 593−600.

(13) Kier, L. B.; Hall, L. H. Derivation and significance of valence molecular connectivity. *J. Pharm. Sci.* **1981**, *70*, 583−589.

(14) Labanowski, J. K.; Motoc, I.; Dammkoehler, R. A. The physical meaning of topological indices. *Computers Chem.* **1991**, *15*, 47−53.

(15) Hosoya, H.; Goth, M.; Murakami, M.; Ikeda, S. Topological index and thermodynamic properties. 5. How can we explain the topological dependency of thermodynamic properties of alkanes with the topology of graphs. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 192−196.

(16) Bögel, H.; Dettmann, J.; Randic, M. Why is the topological approach in chemistry so successful. *Croat. Chem. Acta*, **1997**, *70*, 827−840.

(17) Kier, L. B.; Hall. L. H. Derivation and significance of valence molecular connectivity. *J. Pharm. Sci.* **1981**, *70*, 583−589.

(18) Randić, M. On the characterization of molecular branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609−6615.

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

(20) Coulson, C. A. The electronic structure of some polyenes and aromatic molecules VII: Bonds of fractional order by the molecular orbital method. *Proc. R. Soc. (London)* **1939**, *A169*, 413−428.

(21) Fukui, K. The path oc chemical reactions − The IRC approach. *Acc. Chem. Res.* **1981**, *14*, 363−368.

(22) Coulson, S. A.; Streitwieser, A., Jr. (with extensive help from Poole, M. D.; Brauman, J. I.). *Dictionary of π-Electron Calculations*; Freeman: San Francisco, 1965.

(23) Pauling, L.; Brockway, L. O.; Beach, J. Y. The dependence of interatomic distance on single bond − double bond resonance, *J. Am. Chem. Soc.* **1935**, *57*, 2705

(24) Kekulé, A. Sur la constitution des substances aromatiques. *Bull. Soc. Chim. Paris, Nouv. Ser.* **1865**, *3*, 98.

(25) Randić, M. Conjugated Circuits and Resonance Energies of Benzenoid Hydrocarbons. *Chem. Phys. Lett.* **1976**, *38*, 68−70.

(26) Randić, M. Aromaticity and Conjugation. *J. Am. Chem. Soc.* **1977**, *99*, 444−450.

(27) Platt, J. R. Influence of neighbor bonds on additive bond properties in paraffins, *J. Chem. Phys.* **1947**, *15*, 419.

(28) Randić, M. Characterization of atoms, molecules, and classes of molecules based on path enumeration. *MATCH* **1979**, *7*, 5−60.

(29) Randić, M.; C. L. Wilkins, C. L. On Graph Theoretical Basis for Ordering of Structures. *Chem. Phys. Lett.* **1979**, *63*, 332−336.

(30) Randić, M.; C. L. Wilkins, C. L. Graph-Based Fragment Search in Polycyclic Structures. *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 23−31.

(31) Randić, M.; C. L. Wilkins, C. L. Graph Theoretical Approach to Recognition of Structural Similarity in Molecules. *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 31−37.

(32) Randić, M.; C. L. Wilkins, C. L. Graph Theoretical Study of Structural Similarity in Benzomorphans. *Int. J. Quantum Chem: Quantum Biol. Symp.* **1979**, *6*, 55−71.

(33) Randić, M.; C. L. Wilkins, C. L. Graph Theoretical Ordering of Structures as a Basis for Systematic Searches for Regularities ion Molecular Data. *J. Phys. Chem.* **1979**, *83*, 1525−1540.

(34) Randić, M.; C. L. Wilkins, C. L. Graph Theoretical Analysis of Molecular Properties. Isomeric Variations in Nonanes. *Int. J. Quantum Chem.* **1980**, *18*, 1005−1027.

(35) C. L. Wilkins, C. L.; Randić, M. A Graph Theoretical Approach to Structure−Property and Structure−Activity Correlations. *Theor. Chim. Acta* **1980**, *58*, 45−68.

(36) Randić, M. Random Walks and Their Diagnostic Value for Characterization of Atomic Environment. *J. Comput. Chem.* **1980**, *1*, 386-399.
(37) Razinger, M. Discrimination and ordering of chemical structures by the number of walks. *Theor. Chim. Acta* **1986**, *70*, 365−378.
(38) Knop, J. V.; Müller, W. R.; Szymanski, K.; Randić, M.; Trinajstić, N. Note on acyclic structures and their self-returning walks. *Croat. Chem. Acta* **1983**, *56*, 405−409.
(39) Randić, M.; Woodworth, W. L.; Graovac, A. Unusual Random Walks *Int. J. Quantum Chem.* **1983**, *24*, 435−452.
(40) Shalabi, A. S. Random walks: computations and applications to chemistry. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 483−491.
(41) Müller, W.; Szymanski, K.; Knop, J. V.; Mihalić, Z.; Trinajstić, N. The walk ID number revisited. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 231−233.
(42) Randić, M.; Brissey, G. M.; Spencer, R. G.; Wilkins, C. L. Search for all self-avoiding paths for molecular graphs. *Computers Chem.* **1979**, *3*, 5−13.
(43) Rücker, G.; Rücker, C. On using the adjacency matrix power method for perception of symmetry and isomorphism testing of highly intricate graphs. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 123−126.
(44) Rücker, G.; Rücker, C. Counts of all walks as atomic and molecular descriptors, C. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 683−695.
(45) Razinger, M. Extended connectivity in chemical graphs. *Theor. Chim. Acta* **1982**, *61*, 581−586.
(46) Collatz, L.; Sinogowitz, U. Spectra of finite graphs. *Abh. Math. Sem. University Hamburg* **1957**, *21*, 63−77.
(47) Harary, F.; King. C.; Mowshowitz, A.; Read, R. C. Cospectral graphs and digraphs. *Bull. London Math. Soc.* **1971**, *3*, 321−328.
(48) Balaban, A. T.; Harary, F. The characteristic polynomial does not uniquely determines the topology of a molecule. *J. Chem. Doc.* **1971**, *11*, 258−259.
(49) Živković, T.; Trinajstić, N.; Randić, M. On conjugated molecules with identical topological spectra. *Mol. Phys.* **1975**, *30*, 517−532.
(50) Herndon, W. C. Isospectral molecules. *Tetrahedron Lett.* **1974**, 671.
(51) Randić, M.; Trinajstić, N.; Živković, T. Molecular graphs having identical spectra. *J. Chem. Soc., Faraday Trans. 2* **1976**, *72*, 244−256.
(52) Živković, T.; Trinajstić, N.; Randić, M. On topological spectra of composite molecular systems. *Croat. Chem. Acta* **1977**, *49*, 89−100.
(53) D'Amato, S. S.; Gimarc, B. M.; Trinajstić, N. Isospectral and subspectral molecules. *Croat. Chem. Acta* **1981**, *54*, 1−52.
(54) Heilbronner, E. Some comments on cospectral graphs. *MATCH* **1979**, *5*, 105−133.
(55) Lowe, J. P.; Davis, M. V. Isospectral points and edges in graph theory. *J. Math. Chem.* **1990**, *5*, 27−38.
(56) Jiang, Y. Problems on isospectral molecules. *Sci. Sin.* **1984**, *27*, 236.
(57) Randić, M.; Barysz, M.; Nowakowski, J.; Nikolic, S.; Trinajstić, N. Isospectral Graphs Revisited. *J. Mol. Struct. (THEOCHEM)* **1989**, *185*, 249−274.
(58) Randić, M.; Baker, B. Isospectral Multitrees. *J. Math. Chem.* **1988**, *2*, 249−265.
(59) Rücker, G.; Rücker, Ch. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 422−427.
(60) Rücker, Ch.; Rücker, G. Understanding the properties of isospectral points and pairs in graphs: The concept of orthogonal relation. *J. Math. Chem.* **1992**, *9*, 207−238.
(61) Randić, M. Novel shape descriptors for molecular graphs. *SAR QSAR Environ. Sci.* Submitted for publication.
(62) Randić, M. On Characterization of shape of molecular graphs. *J. Chem. Inf. Comput. Sci.*, in press.
(63) Kier, L. B.; Murray, W. J.; Randić, M.; Hall, L. H. Molecular Connectivity V: Connectivity series applied to density. *J. Pharm. Sci.* **1975**, *65*, 1226−1230.
(64) Kier, L. B.; Hall. H. L. Molecular connectivity VII. Specific treatment of heteroatoms. *J. Pharm. Sci.* **1976**, *65*, 1806−1809.
(65) Kier, L. B.; Hall. L. H. Molecular Connectivity in Chemistry and Drug Research; Academic Press: New York: 1976; p 102.
(66) Kier, L. B.; Hall. L. H. Molecular Connectivity in Chemistry and Drug Research; Academic Press: New York: 1976;p 115.
(67) Kier, L. B.; Hall. L. H. Molecular Connectivity in Chemistry and Drug Research; Academic Press: New York: 1976; p 120.
(68) Kier, L. B.; Hall. L. H. Molecular Connectivity in Chemistry and Drug Research; Academic Press: New York: 1976; p 131.
(69) Kier, L. B.; Hall. L. H. Molecular Connectivity in Chemistry and Drug Research; Academic Press: New York: 1976; p 135.
(70) Kier, L. B.; Hall. L. H. Molecular Connectivity in Chemistry and Drug Research; Academic Press: New York: 1976; p 136.
(71) Kier, L. B.; Hall. L. H. Molecular Connectivity in Chemistry and Drug Research; Academic Press: New York: 1976; p 137.
(72) Kier, L. B.; Hall. L. H. Molecular Connectivity in Chemistry and Drug Research; Academic Press: New York: 1976; p 151.
(73) Kier, L. B.; Hall. L. H. Molecular Connectivity in Chemistry and Drug Research; Academic Press: New York: 1976; p 165.
(74) Kupchik, E. J. Structure-Molar refraction relationship of alkylgermanes using molecular connectivity. *Quant. Struct. Act. Relat.* **1988**, *7*, 57−59.
(75) Randić, M.; Sabljic, A.; Nikolić, S.; Trinajstić, N. A rational selection of graph theoretical indices in the QSAR. *Int. J. Quantum Chem: Quantum Biol. Symp.* **1988**, *15*, 267−285.
(76) Basak, S. C. Information theoretic indices of neighborhood complexity and their applications. In *Topological Indices and Related descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon & Breach Sci. Publ.: Amsterdam, 1999; pp 563−593 (Table 9 p 587).
(77) Randić, M. Novel graph theoretical approach to heteroatoms in QSAR. *Chemometrics Intel. Lab. Syst.* **1991**, *10*, 213−227.
(78) Randić, M. On computation of optimal parameters for multivariate analysis of structure−property relationship. *J. Chem. Inf. Comput. Sci.* **1991**, *12*, 970−980.
(79) Randić, M.; Basak, S. C. Optimal molecular descriptors based on weighted path numbers. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 261−266.
(80) Randić, M.; Basak, S. C. Multiple regression analysis with optimal molecular descriptors. *SAR QSAR Environ. Res.* **2000**, *11*, 1−23.
(81) Randić, M.; Mills, D.; Basak, S. C.; Pogliani, L. On characterization of physical properties of amino acids. Work in progress.
(82) Randić, N.; Basak, S. C. Variable Molecular Descriptors. In *Some Aspects of Mathematical Chemistry*; Sinha, D. K., Basak, S. C., Mohanty, R. K., Busamallick, I. N., Eds.; Visva-Bharati University Press: Santiniketan, India, In press.
(83) Randić, M. Pompe, M. On characterization of CC double bond in alkenes. *SAR QSAR Environ. Sci.* **1999**, *10*, 451−411.
(84) Randić, M.; Dobrowolski, J. Cz. Optimal molecular connectivity descriptors for nitrogen-containing molecules. *Int. J. Quantum Chem.* **1998**, *70*, 1209−1215.
(85) Kier, L. B.; Hall. L. H. Molecular Connectivity in Chemistry and Drug Research; Academic Press: New York: 1976; Table 3, p 246.
(86) Wiener, H. Correlations of heats of isomerization, and differences in heats of vaporization of isomers among paraffin hydrocarbons. *J. Am. Chem. Soc.* **1947**, *69*, 2636−2638.
(87) Wiener, H. Vapor pressure − temperature relationship among the branched paraffin hydrocarbons. *J. Phys. Chem.* **1948**, *52*, 425−430.
(88) Wiener, H. Relation of the physical properties of the isomeric alkanes to molecular structure. *J. Phys. Chem.* **1948**, *52*, 1082−1089.
(89) Wiener, H. Influence of interatomic forces on paraffin properties. *J. Chem. Phys.* **1947**, *15*, 766−766.
(90) Randić, M. Linear combinations of path numbers as molecular descriptors. *New J. Chem.* **1997**, *21*, 945−951.
(91) Balaban, A. T.; Mills, D.; Ivanciuc, O.; Basak, S. C. Reverse Wiener Indices. *Croat. Chem Acta* **2000**, *73*, 923−941.
(92) Ivanciuc, O.; Ivanciuc, T.; Balaban, A. T. The complementary distance matrix, a new molecular graph metric. *Models Chem.* **2000**, *137*, 57−82.
(93) Plavšić, D.; Nikolić, S.; Trinajstić, N.; Mihalić, Z. On the Harary index for the characterization of chemical graphs. *J. Math. Chem.* **1993**, *12*, 235−250.
(94) Ivanciuc, O.; Balaban, T. S.; Balaban, A. T. Design of topological indices. Part. 4. Reciprocal distance matrix, related local vertex invariants and topological indices. *J. Math. Chem.* **1993**, *12*, 309−318.
(95) Harary, F. *Graph Theory*; Addison-Wesley: Reading MA, 1969.
(96) Randić, M. Novel Molecular Descriptor for Structure−Property Studies. *Chem. Phys. Lett.* **1993**, *211*, 478−483.
(97) Randić, M.; Guo, X.; Oxley, T.; Krishnapriyan, H. Wiener matrix: Source of novel graph invariants. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 709−716.
(98) Randić, M.; Guo, X.; Oxley, T.; Krishnapriyan, H.; Naylor, L. Wiener matrix invariants. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 361−367.
(99) Balaban, A. T. Highly discriminating distance-based topological index. *Chem. Phys. Lett.* **1982**, *89*, 399−404.
(100) Balaban, A. T. Topological index based on topological distances in molecular graphs. *Pure Appl. Chem.* **1983**, *55*, 199−206.
(101) Gutman, I. A formula for the Wiener number of trees and its extension to graphs containing cycles. *Graph Theory Notes NY Acad. Sci.* **1994**, *27*, 9−15.
(102) Klein, D. J.; Lukovits, I.; Gutman, I. On the definition of the hyper-Wiener index for cycle-containing structures. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 50−52.
(103) Zhu, H.-Y.; Klein, D. J.; Lukovits, I. Extensions of the Wiener number. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 420−428.

CI000095O