

Wiener Matrix: Source of Novel Graph Invariants

M. Randić,* Xiaofeng Guo,† T. Oxley, and H. Krishnapriyan

Department of Mathematics and Computer Science, Drake University, Des Moines, Iowa 50311

Received March 12, 1993*

We report some properties of new graph matrices which appear to offer novel graph invariants of potential interest in structure–property studies. The matrices are constructed by generalizing Wiener's procedure for evaluation of Wiener numbers in alkanes. Among the invariants considered we particularly examined the sequences generated by summing the entries in the matrix for vertices at the same distance from one another. These numbers may be viewed as "higher" Wiener numbers in analogy with "higher" connectivity indices. We have listed the higher Wiener numbers of alkanes up to $n = 9$ carbon atoms and also report several recursions for the construction of these invariants for selected families of acyclic graphs. Briefly, we have outlined how the Wiener matrix can be extended to cyclic systems, while in the concluding comments we have outlined an extension of the Wiener matrix to molecules having heteroatoms. The significance of the matrices as a source of graph invariants is precisely in this possibility to go beyond simple models of molecular graphs and extend graph invariants of interest to molecules having different kinds of atoms.

INTRODUCTION

Modeling of structure–property relationships using graphs invariants to represent structures of chemical interest has received considerable attention in recent literature.¹ Hundreds of distinct graph invariants, often referred to as topological indices, have been suggested.² Recently, more critical studies, however, have shown that a relatively small number of these appear to lead to practical correlations.³ Nevertheless, a search for novel graph invariants continues, because, even in the case of alkanes, there are numerous properties of molecules for which one would like to derive better regression equations. Introduction of novel invariants is justified if one can demonstrate that such invariants have application in structure–property studies. Hence, it is not enough just to propose a novel graph invariant and expect that it may be of interest in chemistry. We feel that it is the burden of the investigators who introduce a novel structural invariant to illustrate its use and advantages over the existing molecular descriptors. A list of desirable conditions on graph invariants of interest in chemistry,⁴ based on a similar list of the desirable requirements for the chemical codes proposed by Read,⁵ should be consulted in order to curb an unnecessary proliferation of topological indices.

PRELIMINARIES

Besides the obvious conditions on structural and graph theoretical invariants outlined in ref 4, such as a desirability for uniqueness, demonstration of their use on a selected property for a class of compounds, and above all a possibility for a direct structural interpretation, we would like to add another desirable quality which was recognized as such only recently: the possibility of extending the descriptor and constructing a series of structurally related descriptors. Such a set of structurally related descriptors can be used as a *basis* (*vide infra*) for a discussion of structure–property relationships. Initially topological indices were used as *ad hoc* descriptors in multiple regression analysis and other statistical approaches, such as the principal components analysis (PCA).⁶ The connectivity index⁷ combined with the "higher" connectivity

indices⁸ was for the first time used as a family of structurally related topological indices. Such a set of descriptors offers a more homogeneous and systematic representation of the molecular structure. When such descriptors are ordered they, as recently pointed out,⁹ lead to a notion of the basis descriptors. An ordered set of descriptors makes comparison between different regressions convenient, and if the basis is well-chosen, it need not diminish the overall quality of the regression analysis. In contrast, when one uses statistical procedures to select an optimal set of descriptors, one will single out the dominant descriptors. Selection of dominant descriptors will depend on the properties considered and may change when different properties of the same molecules are examined, or when the same property in a different set of molecules is considered. Use of descriptors that are less than optimal, however, may compensate for the disadvantages of the use of basis descriptors, as it helps the comparative studies of structure–property relationships.

While it is generally not difficult to introduce novel structural invariants (graph indices) if one disregards the question of their relevance in structure–property studies, it is not so easy to construct a "family" of structurally related topological indices. For example, among the widely used topological indices, we find the Wiener number¹⁰ and the Hosoya's Z index,¹¹ both of which individually play an important role in many structure–property correlations.¹² However, these two indices have not yet been simply "extended" to generate higher members of a family of structurally related indices. In contrast, such an extension of an index was possible and was natural for the connectivity index.

With this paper we continue our efforts to arrive at a family of structural indices, rather than focusing attention on a single novel topological index. We will here discuss one particular generalization of the Wiener numbers W , based on matrices that we associate with acyclic graphs, which, at least in appearance, may lead to an additional basis of molecular descriptors.

KNOWN GRAPH MATRICES

One route to a family of structurally related molecular descriptors is the construction of novel graph matrices. Once one assigns a matrix to a graph, different invariants can be

† On leave from the Department of Mathematics, Xinjiang University, Wulumuqi Xinjiang 830046, People's Republic China.

* Abstract published in *Advance ACS Abstracts*, September 1, 1993.

Table I

basis	
path numbers	$^1p, ^2p, ^3p, ^4p, ^5p, ^6p, \dots$
connectivity indices	$^1\chi, ^2\chi, ^3\chi, ^4\chi, ^5\chi, ^6\chi, \dots$
weighted paths	$^1\pi, ^2\pi, ^3\pi, ^4\pi, ^5\pi, ^6\pi, \dots$
extended connectivities	$^1\epsilon, ^2\epsilon, ^3\epsilon, ^4\epsilon, ^5\epsilon, ^6\epsilon, \dots$
connectivity powers	$^1\alpha, ^2\alpha, ^3\alpha, ^4\alpha, ^5\alpha, ^6\alpha, \dots$

extracted from such a matrix. For example, one can count paths of different lengths in such matrices, or one can count weighted paths in such a matrix in analogy with the counts of paths and weighted paths in the adjacency matrix.¹³ Alternatively, matrices can be used in combination with vectors to produce novel vectors, as outlined by Balaban and collaborators.¹⁴ In this way one can generate many additional variables to be used in multiple regression analysis or the principal components analysis. Novel descriptors derived thus, whether based on novel matrices or matrix-vector multiplications, of course, still have to pass the test of "desirability".

It is somewhat surprising that there are only a few matrices that can be directly associated with molecular graphs and graphs in general. In the case of trees, i.e., acyclic graphs, a few of the known matrices are illustrated in Table II for 2-methylpentane. We started with the well-known matrices: *adjacency matrix* A, the entries of which are 0, 1 depending on adjacency relation,^{15,16} and *distance matrix* D, with entries that count the length of the path between vertices (i, j), the length being given by the number of edges between i and j .¹⁷

In addition we listed several more recently introduced matrices: *extended distance matrix* E (of Tratch, Stankevitch and Zefirov¹⁸) the entries (i, j) of which are given as the product of the length of the path i, j and the multiplicity number $m_{i,j}$, which indicates the number of paths in the graph of which path i, j is a subgraph, *connectivity matrix* C (of Kier and Hall¹⁹), which is based on the connectivity index and higher connectivity index weights (an entry (i, j) is given by the contribution of the path i, j to the corresponding connectivity index), and *path matrix* P, which is analogous to matrix C but the entries are based on weighted path contributions.²⁰

When one extends these considerations to graphs embedded in 3-dimensional space, one obtains additional matrices: *topographic matrix* T,²¹ which reports distances (measured in units of a standard CC bond) between vertices for a graph embedded on a graphite or diamond lattice, *distance-distance matrix* DD,²² in which the individual entries are given as a ratio between the geometric distance (i.e., measured through space) to graph distance (i.e., measured through bonds), *geometric matrix* G,²³ in which genuine distances between atoms represent the entries in the matrix (and consequently no direct information on bonding is supplied), and *inverse G matrix* G^{-1} of Wilson²⁴ (and M. Eliashevich), well-known from normal coordinate calculations in vibrational spectroscopy.

This brief list almost exhausts the known matrices which have been mentioned in the chemical literature. One can construct additional matrices, particularly if one considers molecular properties as entries in such matrices. Thus, the G^{-1} matrix is accompanied with the F matrix for the force constants. Balandin already in 1940 (in what is the earliest review article on chemical graph theory²⁵) at some length discussed the use of properties as entries of matrices. We may also mention here alphanumeric matrices of Spialter²⁶ who was interested in the possibility of using matrices with nonnumerical entries along the main diagonal to represent a structure for the needs of chemical documentation. The work of Spialter is also significant as it hinted at how one can, by

Table II. Graph Matrices for 2-Methylpentane

Adjacency Matrix						Distance Matrix					
0	1	0	0	0	0	0	1	2	3	4	2
1	0	1	0	0	1	1	0	1	2	3	1
0	1	0	1	0	0	2	1	0	1	2	2
0	0	1	0	1	0	3	2	1	0	1	3
0	0	0	1	0	1	4	3	2	1	0	4
0	1	0	0	1	0	2	1	2	3	4	0
Expanded Distance Matrix						Wiener Matrix					
0	5	6	6	4	2	0	5	3	2	1	1
5	0	9	12	9	5	5	0	9	6	3	5
6	9	0	8	8	6	3	9	0	8	4	3
6	12	8	0	5	6	2	6	8	0	5	2
4	9	8	5	0	4	1	3	4	5	0	1
2	5	6	6	4	0	1	5	3	2	1	0
Connectivity Matrix						Weighted Path Matrix					
0	0.57735	0.40824	0.28867	0.28867	0.57735	0	0.57735	0.23571	0.11785	0.08333	0.33333
0.57735	0	0.40824	0.28867	0.28867	0.57735	0.57735	0	0.40824	0.20412	0.14434	0.57735
0.40824	0.40824	0	0.50000	0.50000	0.40824	0.23571	0.40824	0	0.50000	0.35355	0.23571
0.28867	0.28867	0.50000	0	0.70711	0.28867	0.11785	0.20412	0.50000	0	0.70711	0.11785
0.28867	0.28867	0.50000	0.70711	0	0.28867	0.08333	0.14434	0.35355	0.70711	0	0.08333
0.57735	0.57735	0.40824	0.28867	0.28867	0	0.33333	0.57735	0.23571	0.11785	0.08333	0
Topographic Matrix						Distance-Distance Matrix					
0	1	1.633	2.517	3.266	1.633	0	1	0.815	0.839	0.817	0.815
1	0	1	1.633	2.517	1	1	0	1	0.815	0.839	1
1.633	1	0	1	1.633	1.633	0.815	1	0	1	0.815	0.815
2.517	1.633	1	0	1	1.633	0.839	0.815	1	0	1	0.844
3.266	2.517	1.633	1	0	1	0.817	0.839	0.815	1	0	0.250
1.633	1	1.633	1.633	1	0	0.815	1	0.815	0.844	0.250	0
Restricted Random Walk											
0	1	$1/3$	$1/4$	$1/11$	$1/3$	0	1	$1/3$	$1/4$	$1/11$	$1/3$
$1/3$	0	$1/3$	$1/4$	$1/11$	$1/3$	$1/5$	$1/2$	0	$1/2$	$1/5$	$1/3$
$1/5$	$1/2$	0	$1/2$	$1/5$	$1/3$	$1/7$	$1/3$	$1/2$	0	$1/7$	$1/7$
$1/7$	$1/3$	$1/2$	0	$1/7$	$1/7$	$1/3$	1	$1/2$	1	0	$1/7$
$1/3$	1	$1/3$	$1/4$	$1/11$	0	$1/3$	$1/4$	$1/11$	$1/4$	$1/11$	0

introducing some parameters along the diagonal entries of the adjacency matrix, extend the concept of graphs which are mathematical objects, to represent molecules, objects of chemistry. The search for optimal parameters for heteroatoms in alcohols has recently been outlined in ref 27, while the search for representation of other heteroatoms continues.²⁸

Once a matrix is selected, one may consider powers of such matrices. In the case of the adjacency matrix the powers lead to matrices that count random walks.^{15,16,29} From the powers of the adjacency matrix one can also extract the count of restricted random walks (used in the construction of a novel nonsymmetrical matrix that can be assigned to graphs³⁰). However, in view of the Cayley-Hamilton theorem¹⁵ power matrices with exponents exceeding n , the number of vertices in a graph, do not introduce structural novelty. A way to circumvent this limitation was outlined with the concept of α matrices.³¹ The greek letter α alludes to a beginning of a possibly useful novel algorithmic route for generating infinite sequences of matrices and associated infinite sequences of molecular descriptors. The particular illustration in ref 31 is

based on the partitioning of the connectivity index. For all adjacent vertices, i, j the entry in the α matrix is given by $(m_i m_j)^{-1/2}$ where m_i, m_j are the valencies of the vertices i, j , respectively. All other elements in the α matrix are zero. Hence the matrix α may be viewed as a weighted adjacency matrix.

If one adds entries above the main diagonal in the α matrix, one obtains the connectivity index $^1\chi$. Powers $\alpha^2, \alpha^3, \alpha^4, \alpha^5, \alpha^6, \dots$ can be easily computed, from which one can extract the contributions associated with adjacent vertices (i.e., bonds). By adding all bond contributions in each of the matrices corresponding to various powers of α , one generates a sequence of novel graph invariants.

THE WIENER MATRIX

Can we introduce other novel matrices associated with graphs that are of potential interest in structure-property studies? Here we would like to consider in some detail one such graph matrix as a source of structural invariants of interest in chemistry. A matrix, the elements of which have relationship to the Wiener number, was recently suggested as being of potential interest in structure-property studies.³² Preliminary studies have shown that an invariant derived from this particular matrix, to be referred to as the Wiener matrix, gives the best single variable regression for the isomeric variations of the critical volumes and the mean radius of octanes.³²

Wiener's work in the late 1940s¹⁰ represents a novel approach to structure-property studies that had escaped the attention of most chemists for well over 2 decades. This occurred despite the fact that Platt³³ explicitly pointed to the significance and the potential of graph theoretical descriptors for studies of physicochemical molecular properties. Wiener successfully accounted for the variations in many properties of alkanes, alcohols, amines, and fatty acids by using two "strange" descriptors: W which counts distances between all pairs of carbon atoms in a molecule and P which counts all "heavy" atoms three bonds apart. According to Platt W is related to the molecular volume and P reflects the overcrowding (i.e., the presence of close hydrogen atoms in 3-dimensional space). Wiener himself interpreted W as a measure of molecular compactness:

"The path number W was calculated as the total distance between all carbon atoms. The smaller this total distance, the larger is the compactness of the molecule. The boiling point (in paraffins) is seen to vary inversely with compactness."¹⁰

Subsequent studies showed that W is a fundamental descriptor, in that it appears in numerous applications. There is considerable literature on the evaluation of W in molecules.³⁴ Recent reports consider extension of the Wiener number to polycyclic structures, as well as to the 3-dimensional structures.³⁵

We want to generalize the Wiener number W in a different sense: We want to arrive at a matrix that is structurally related to W . To design a matrix for a graph, one has to find a procedure that will assign to each pair of vertices (i, j) a numerical value. This may not be difficult for adjacent vertices, in which case we produce a sparse matrix, that can be viewed as a weighted adjacency matrix. However, it may be difficult to generalize many such schemes to non-adjacent vertices. We will outline a scheme which generalizes the

computation of the Wiener number to a "full" matrix, i.e., a matrix with all elements being nonzero except the diagonal elements.

We start by recalling the construction of W as outlined by Wiener:¹⁰

"The path number W is defined as the sum of the distances between any two carbon atoms in the molecule, in terms of carbon-carbon bonds. Brief method of calculation: Multiply the number of carbon atoms on one side of any bond by those on the other side; W is the sum of those values for all bonds."

The first part of the outline could have been simply stated by saying that W is given as a sum of elements above the diagonal of the distance matrix D if the distance matrix had been known at that time! Hosoya was the first to express W in terms of the distance matrix.¹¹ The second part of the above outline suggests a scheme for the construction of a *weighted* adjacency matrix: assign the contributions to the Wiener number W associated with individual bonds (adjacent vertices) to those bonds (instead of adding them into a single parameter W). In this way each pair of adjacent vertices has a nonzero entry. It remains now to see how such a sparse matrix can be completed with nonzero entries for the non-adjacent vertices. We have to try to generalize the scheme which assigned nonzero values to paths of length 1 (bonds) to paths of any length. The following definition accomplishes our goal.

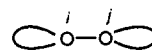
Definition: For each path P_{ij} , i.e., for each pair of vertices i, j (adjacent or not) the corresponding element W_{ij} in the matrix $\langle W \rangle$ is given as the product of the numbers P_i and P_j , where P_i and P_j signify the number of vertices on each side of the path i, j , including i and j , respectively.

Observe that our generalization of the initial approach of Wiener is based on extending to paths a definition limited initially to edges. The above definition applies only to trees, since paths are unique and their deletion separates the graph into disjoint parts only in trees. How to extend such consideration to cyclic structures is yet to be considered. Such a generalization need neither be simple nor possible, which, however, does not diminish the significance and the importance that the novel matrices may have for acyclic structures.

ILLUSTRATIONS

In Table III we have listed the novel matrices for the nine isomers of heptane. Because the $\langle W \rangle$ matrices are symmetric (by definition), only the entries above the main diagonal are shown. A glance at the individual matrices shows some of their main features: The entries belonging to adjacent and close neighbors are, as a rule, large, while the entries belonging to more distant neighbors tend to be smaller. Entries belonging to terminal vertices are always equal to 1; hence the number of 1's that appear in such a matrix (above the main diagonal) indicates the number of terminal paths (paths between terminal vertices).

It is not difficult to see that the count of all paths involving a path i, j is equivalent to multiplying "the number of vertices on one side of any path by those on the other side". We will represent a general tree schematically as



by emphasizing an edge i, j , which partitions the vertices into

Table III. Wiener Matrices for Heptane Isomers (Matrices Are Symmetrical)

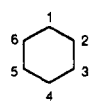
n-Heptane							2-Methylhexane						
0	6	5	4	3	2	1	0	6	4	3	2	1	1
	0	10	8	6	4	2		0	12	9	6	3	6
		0	12	9	6	3			0	12	8	4	4
			0	12	8	4				0	10	5	3
				0	10	5					0	6	2
					0	6						0	1
						0							0
3-Methylhexane							3-Ethylpentane						
0	6	5	3	2	1	1	0	6	5	2	1	2	1
	0	10	6	4	2	2		0	10	4	2	4	2
		0	12	8	4	6			0	10	5	10	5
			0	10	5	3				0	6	4	2
				0	6	2					0	2	1
					0	1						0	6
						0							0
2,2-Dimethylpentane							2,3-Dimethylpentane						
0	6	3	2	1	1	1	0	6	4	2	1	1	1
	0	12	8	4	6	6		0	12	6	3	6	3
		0	10	5	3	3			0	10	5	4	6
			0	6	2	2				0	6	2	2
				0	1	1					0	1	1
					0	1						0	1
						0							0
2,4-Dimethylpentane							3,3-Dimethylpentane						
0	6	4	3	1	1	1	0	6	5	2	1	1	1
	0	12	9	3	6	3		0	10	4	2	2	2
		0	12	4	4	4			0	10	5	6	6
			0	6	3	6				0	6	2	2
				0	1	1					0	1	1
					0	1						0	1
						0							0
2,2,3-Trimethylbutane													
0	6	3	1	1	1	1							
	0	12	4	6	6	4							
		0	6	3	3	6							
			0	1	1	1							
				0	1	1							
					0	1							
						0							

m vertices on the left and n vertices on the right side of the bond. According to Wiener the contribution of the bond i,j to the Wiener to the Wiener number W is simply mn . On the other hand the count of paths which incorporate the edge i,j is as follows:

1	for the edge i,j
$m-1$	paths starting at the left and ending at vertex j
$n-1$	paths starting at the right and ending at vertex j
$(m-1)(n-1)$	paths starting at the left and ending at the right

The sum of all the above contributions again gives mn .

There is, however, an important advantage to viewing $\langle W \rangle$ to be derived by the count of paths involving the path between vertices i,j . Such viewing allows the construction of the Wiener matrix for cyclic graphs. For example, in the case of benzene the matrix is particularly simple (being a Toeplitz matrix, with entries on all diagonals being equal):



$$\begin{pmatrix} 0 & 9 & 7 & 5 & 7 & 9 \\ 9 & 0 & 9 & 7 & 5 & 7 \\ 7 & 9 & 0 & 9 & 7 & 5 \\ 5 & 7 & 9 & 0 & 9 & 7 \\ 7 & 5 & 7 & 9 & 0 & 9 \\ 9 & 7 & 5 & 7 & 9 & 0 \end{pmatrix}$$

By inspection of the molecular graph one can immediately

obtain for the symmetry non-equivalent elements of $\langle W \rangle$

$$p_{1,2} = 1 + 4 + 4; \quad p_{1,3} = 1 + 3 + 3; \quad p_{1,4} = 1 + 2 + 2$$

The Wiener matrix $\langle W \rangle$ is related to the matrix introduced by Tratch, Stankevitch, and Zefirov,¹⁸ the elements of which were defined by the count of weighted paths involving a path i,j . The weight was determined by the length of the path considered. Hence, the extension of the Wiener procedure for calculating the contributions arising from different paths and the procedure of Tratch *et al.* give the same answer for the matrix elements corresponding to adjacent vertices (when the multiplicative factor is 1).

PROPERTIES OF THE WIENER MATRICES

Matrices in Table III will be referred to as Wiener matrices to honor Harry Wiener, the pioneer of early structure-property studies based on use of mathematical descriptors. These matrices have been briefly mentioned in a review of a search for graph theoretical invariants.³⁶ They were initially referred to as vicinal matrices, in view of the fact that entries in such matrices corresponding to vertices in the vicinity tend to be large. In contrast the opposite is true for the distance matrix.

If one adds all entries in such a matrix (above the main diagonal) belonging to adjacent vertices, one obtains Wiener number W . However, we can generate a sequence of Wiener-like numbers by adding entries corresponding to paths of the same length. Adjacent vertices (which define bonds) correspond in a matrix to paths of length 1. If we add all such contributions, we obtain, as already mentioned, the Wiener number. However, we can now extend this by adding contributions associated with paths of length 2, then paths of length 3, and so on, making similar summations for paths of increasing length. A molecule (graph) is then represented by a sequence: ${}^1W, {}^2W, {}^3W, {}^4W, {}^5W, {}^6W, \dots$. The first entry 1W is the Wiener number W . In analogy with the connectivity index and higher connectivity indices we may call the remaining contributions the "higher" Wiener numbers.

If we add all the entries in a sequence ${}^1W, {}^2W, {}^3W, {}^4W, {}^5W, {}^6W$, we obtain yet another graph invariant, which we call WW . This invariant represents the sum of all entries above the main diagonal in $\langle W \rangle$. In other words WW is the sum of all paths in $\langle W \rangle$ matrix. In Table IV we list the Wiener sequences and WW values for smaller alkanes, and in Table V we extend the list to octane and nonane isomers.

REGULARITIES FOR HIGHER WIENER NUMBERS

Some regularities for the members of Wiener sequences shown in Table IV and Table V are immediately visible. The W -sequences for the chains (n -alkanes) have particularly simple structures: a sequence for a n -alkane contains all the entries found in smaller alkanes:

ethane	1
propane	4, 1
butane	10, 4, 1
pentane	20, 10, 4, 1
hexane	35, 20, 10, 4, 1
heptane	56, 35, 20, 10, 4, 1
octane	84, 56, 35, 20, 10, 4, 1

Hence, when we know the Wiener number for any n -alkane, which is the leading term in the sequence, we can recursively construct the W -sequence without resorting to the Wiener matrix $\langle W \rangle$. The numbers in the sequences are sums of triangular numbers 1, 3, 6, 10, 15, etc.

Table IV. Wiener Sequences for Smaller Alkanes, the Sequence Sum (WW), and the First Eigenvalue (M = Methyl; E = Ethyl)^a

isomer	<i>k_W</i> numbers	WW	x_1
ethane	1	1	1
propane	4, 1	5	3.372
butane	10, 4, 1	15	7.720
2-M	9, 3	12	6.292
pentane	20, 10, 4, 1	35	14.661
2-M	18, 8, 2	28	12.053
2,2-MM	16, 6	22	9.639
hexane	35, 20, 10, 4, 1	70	24.805
2-M	32, 17, 7, 2	58	21.339
3-M	31, 16, 6, 1	54	19.703
2,3-MM	29, 14, 4	47	17.705
2,2-MM	28, 13, 3	44	16.559
heptane	56, 35, 20, 10, 4, 1	126	38.776
2-M	52, 31, 16, 7, 2	108	34.680
3-M	50, 29, 14, 5, 1	99	31.621
2,4-MM	48, 27, 12, 4	91	30.281
3-E	48, 27, 12, 3	90	28.396
2,2-MM	46, 25, 10, 3	84	28.013
2,3-MM	46, 25, 10, 2	83	27.336
3,3-MM	44, 23, 8, 1	76	24.902
2,2,3-MMM	42, 21, 6	69	23.429

^a Each group of isomers is ordered by decreasing values of WW.

There are other similar recursions for the Wiener sequences of other families of structurally related molecules. Consider the family of 2-methyl substituted alkanes:

2-methylpropane (isobutane)	9, 3
2-methylbutane	18, 8, 2
2-methylpentane	32, 17, 7, 2
2-methylhexane	52, 31, 16, 7, 2
2-methylheptane	79, 51, 30, 16, 7, 2
2-methyloctane	114, 78, 50, 30, 16, 7, 2
2-methylnonane	158, 113, 77, 50, 30, 16, 7, 2

We see the same regularity found in *n*-alkanes also here for the "tail" part of the sequences. As the size of the molecule increases and the "tail" of the molecular becomes longer and more prominent, we see that the tail part of the sequence is "constant" among isomers of different size. Just as in the case of normal alkanes, once we know the Wiener number *W* for a structure (the leading term) and the sequences for the smaller members of the family, the whole sequence can be constructed.

Additional illustration of the Wiener sequences for structurally closely related graphs is shown in Table VI. Regularities already illustrated on *n*-alkanes and 2-methylalkanes can be detected in other cases too. There are some differences for the leading members in each family, the differences among the initial coefficients being 1, 2, or more. Thus the critical step is to know the Wiener number of a graph, and the sequences for at most a few additional members on the family. Since there are schemes (and computer programs) that give *W*, the construction of the *W*-sequences, shown in Tables IV–VI, can be accomplished without construction of the corresponding matrices. This is important since while construction of *W*-matrices (for molecules of relatively small size) need not be tedious, the search for elements in the matrix that correspond to paths of a same length can be tedious and error-prone. All the sequences shown in Table VI have a same simple structure: The third difference between the coefficients of successive sequences is constant and equal to 1. Hence it suffices to know sequences for two successive molecules in a sequence to construct Wiener sequences for all the molecules of a family!

Another kind of relationship can be found among structures of the same size, i.e., among isomers. Let us illustrate this

Table V. Wiener Numbers and Higher Wiener Numbers for Octane and Nonane Isomers (M = Methyl; E = Ethyl)

isomer	<i>k_W</i> numbers	WW
octane	84, 56, 35, 20, 10, 4, 1	210
2-M	79, 51, 30, 16, 7, 2	185
3-M	76, 48, 27, 11, 5, 1	170
4-M	75, 47, 26, 12, 4, 1	165
2,5-MM	74, 46, 25, 12, 4	161
3-E	72, 44, 23, 9, 2	150
2,2-MM	71, 43, 22, 10, 3	149
2,4-MM	71, 43, 22, 9, 2	147
2,3-MM	70, 42, 21, 8, 2	143
3,4-MM	68, 40, 19, 6, 1	134
3,3-MM	67, 39, 18, 6, 1	131
2-M, 3-E	67, 39, 18, 5	129
2,2,4-MMM	66, 38, 17, 6	127
2,3,4-MMM	65, 37, 16, 4	122
3-M, 3-E	64, 36, 15, 3	118
2,2,3-MMM	63, 35, 14, 3	115
2,3,3-MMM	62, 34, 13, 2	111
2,2,3,3-MMM	58, 30, 9	97
nonane	120, 84, 56, 35, 20, 10, 4, 1	330
2-M	114, 78, 50, 30, 16, 7, 2	297
3-M	110, 74, 46, 26, 13, 5, 1	275
2,6-MM	108, 72, 44, 25, 12, 4	265
4-M	108, 72, 44, 24, 11, 4, 1	264
2,2-MM	104, 68, 40, 22, 10, 3	247
2,5-MM	104, 68, 40, 21, 9, 2	244
3-E	104, 68, 40, 20, 8, 2	242
2,3-MM	102, 66, 38, 19, 8, 2	235
2,4-MM	102, 66, 38, 19, 7, 2	234
4-E	102, 66, 38, 18, 6, 1	231
3,5-MM	100, 64, 36, 17, 6, 1	224
2,2,5-MMM	98, 62, 34, 17, 6	217
3,3-MM	98, 62, 34, 16, 6, 1	217
3,4-MM	98, 62, 34, 15, 5, 1	215
2,4-ME	98, 62, 34, 15, 4	213
4,4-MM	96, 60, 32, 14, 4, 1	207
2,3,5-MMM	96, 60, 34, 12, 4	206
2,3-ME	96, 60, 32, 13, 3	204
2,2,4-MMM	94, 58, 230, 13, 3	198
3,4-ME	94, 58, 30, 11, 2	195
2,2,3-MMM	92, 56, 28, 11, 3	190
2,4,4-MMM	92, 56, 28, 11, 2	189
3,3-ME	92, 56, 28, 10, 2	188
2,3,4-MMM	92, 56, 28, 10, 2	188
2,3,3-MMM	90, 54, 26, 9, 2	181
2,3,4-MEM	90, 54, 26, 8	178
2,2,4,4-MMMM	88, 52, 24, 9	173
3,3,4-MMM	88, 52, 24, 7, 1	172
2,2,3-MME	88, 52, 24, 7	171
3,4-EE	88, 52, 24, 6	170
2,2,3,4-MMMM	86, 50, 22, 6	164
2,3,3-MME	86, 50, 22, 5	163
2,3,3,4-MMMM	84, 48, 20, 4	156
2,2,3,3-MMMM	82, 346, 18, 3	149

on smaller *n*-alkanes and 2-methylalkanes for which we already listed the Wiener sequences. Since construction of Wiener sequences for *n*-alkanes are particularly straightforward, we will consider how to construct the sequences for 2-methylalkanes using the information on *n*-alkanes. We start by considering the differences among the corresponding terms in the two Wiener sequences:

isomer pair	difference sequence
<i>n</i> -butane–isobutane	1, 1, 1
<i>n</i> -pentane–isopentane	2, 2, 2, 1
<i>n</i> -hexane–isohexane	3, 3, 3, 2, 1
<i>n</i> -heptane–isohexane	4, 4, 4, 3, 2, 1
<i>n</i> -octane–isooctane	5, 5, 5, 4, 3, 2, 1
and so on	

We have emphasized the leading terms which are the same. Similar patterns can be found for the differences in the coefficients between other pairs of structurally related families

Table VI. Wiener Sequences for Families of Structurally Related Trees

<i>n</i>	<i>kW</i> numbers
2-Methylalkanes	
4	9, 3
5	18, 8, 2
6	32, 17, 7, 2
7	52, 35, 19, 7, 2
8	79, 51, 30, 16, 7, 2
9	114, 78, 50, 30, 16, 7, 2
10	158, 113, 78, 50, 30, 16, 7, 2
3-Methylalkanes	
6	31, 16, 6, 1
7	50, 29, 14, 5, 1
8	76, 48, 27, 13, 5, 1
9	110, 74, 46, 26, 13, 5, 1
10	153, 108, 72, 45, 26, 13, 5, 1
11	206, 150, 106, 71, 45, 26, 13, 5, 1
2,2-Dimethylalkanes	
5	16, 6
6	28, 13, 3
7	46, 25, 10, 3
8	71, 43, 22, 10, 3
9	104, 68, 40, 22, 10, 3
10	146, 101, 65, 40, 32, 10, 3
2,(<i>n</i> - 1)-Dimethylalkanes	
6	29, 14, 4
7	48, 27, 12, 4
8	74, 46, 25, 12, 4
9	108, 72, 44, 25, 12, 4
10	151, 106, 70, 44, 25, 12, 4
3-Ethylalkanes	
7	48, 27, 12, 3
8	72, 44, 23, 9, 2
9	104, 68, 40, 20, 8, 2
10	145, 100, 64, 37, 19, 8, 2
11	196, 141, 132, 61, 36, 19, 8, 2
3,3-Dimethylalkanes	
7	44, 23, 8, 1
8	67, 39, 18, 6, 1
9	98, 62, 34, 16, 6, 1
10	138, 93, 57, 32, 16, 6, 1

of compounds. Thus, we obtain the following when comparing 2-methyl and 3-methyl derivatives:

isomer pair	difference sequence
2-methylpentane-3-methylpentane	1, 1, 1, 1
2-methylhexane-3-methylhexane	2, 2, 2, 2, 1
2-methylheptane-3-methylheptane	3, 3, 3, 3, 2, 1
2-methyloctane-3-methyloctane	4, 4, 4, 4, 3, 2, 1
etc.	

Again we emphasized the leading terms which are the same, this time repeating four times. Once we find recursions for the coefficients of the Wiener sequences, we can derive the invariant WW for such graphs by adding all members in the sequence, again without explicitly constructing the **W** matrix first.

There are a number of interesting questions that have yet to be considered. Of immediate interest is to find the smallest graphs having the same WW numbers and the smallest graphs having the same *W*-sequences. The size of such smallest graphs indicates the limitations of such invariants. The larger size of the graphs for which duplicates occur first implies that the particular invariant has captured more of the structural information. Thus the occurrence of coincidences in WW and *W* sequences is a measure of "deficiencies" of such descriptors as a basis for extensive structure-property studies.

From Tables IV and V we see that for octanes and smaller alkanes all *W* sequences and WW numbers are distinct,

although, as is known, among octanes we already have isomers that have same Wiener number *W*. However, when we extend the comparison to nonane isomers, we find a pair of structures (3-methyl-3-ethylhexane and 2,3,4-trimethylhexane) which have not only same WW (188) but also identical *W* sequences (92, 56, 28, 10, 2). Incidentally, the same pair of nonane isomers show identical generalized Wiener number derived from extended distance matrices of Tratch and collaborators.¹⁸

To arrive at isomers with the same WW for relatively small graphs is perhaps not so surprising. Isomers having the same *Z* Hosoya index¹¹ and the same connectivity χ index are already found among octanes. The smallest isomers having the same Balaban index *J*³⁷ have *n* = 12 vertices. A much higher discrimination among isomers is obtained using the ID numbers,³⁸ as found by Knop and co-workers. The smallest acyclic pairs of graphs with the same ID numbers have *n* = 15 vertices.³⁹ When the weights of bonds of different type are optimized (by using the consecutive prime numbers in the weighting algorithm⁴⁰), the smallest acyclic graphs having the same prime ID number have *n* = 20 vertices,⁴¹ an impressive discriminating power for a single invariant molecular representation. It may appear, however, somewhat surprising that when we use several invariants simultaneously, we find, among relatively smaller alkanes, isomers having identical characterization based on *W* sequences. This signifies that *W* sequences do not carry as much structural information as we may have hoped for. The "early" occurrence of the degeneracy among structural invariants signals limitations of such invariants in capturing distinct structural features. A close look at the *W* sequences reveals that the leading three members in all such sequences for isomers (structures with same number of vertices) are strictly collinear! The mutual correlation of such coefficients gives for *R*, the coefficient of the regression an exact value of 1. This interdependence of the leading coefficients is also reflected in the constancy of the leading entries in the difference sequences among isomers earlier mentioned. Hence, without loss of information one can eliminate two of the three leading members from such sequences.

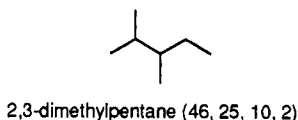
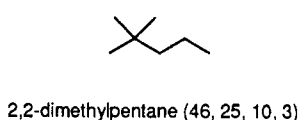
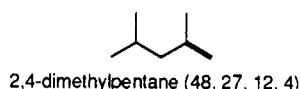
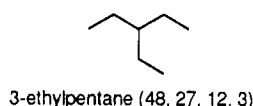
Advantage of matrices, over *ad hoc* constructed invariants that may have structural interpretation but are not related to a matrix representation of a graph, is that matrices offer many additional invariants. Besides the characteristic polynomial and matrix eigenvalues (spectrum), we can count paths of different length in **W** matrices. Alternatively, in analogy with Balaban's construction of the *J* index³⁷ from the distance matrix (which itself was constructed in an analogy to the connectivity index χ), one can construct first the row sums from which one can derive new descriptors.

We have included in Table IV values of the largest eigenvalue of Wiener matrices to illustrate the degree of variations in the first eigenvalue of these matrices. As one can see the first eigenvalue shows a greater variation than was the case with WW, since no two isomers have been found (yet) to have the same spectrum. In particular, for the case of 2,3,4-trimethylhexane and 3-methyl-3-ethylhexane, which have the same WW and the same *W* sequences we find distinct x_1 eigenvectors: 51.567 and 49.413, respectively. It is interesting to see that x_1 , the first (positive) eigenvalue gives the same relative ordering of isomers as does WW. The largest value belongs to *n*-alkane and the eigenvalue decreases generally with branching, though interestingly 2,4-dimethylpentane has a larger x_1 than 3-ethylpentane. Another interesting aspect of the spectra of $\langle \mathbf{W} \rangle$ matrices is a relative constant value for the largest negative eigenvalue. In the

case of heptane isomers this eigenvalue varies between -13.15 (for *n*-heptane) and -13.80 (for 3-ethylpentane). In contrast no apparent regularities follow for the coefficients of the characteristic polynomials, which vary widely. For example, the constant term in the polynomial for *n*-hexane is over 37 000 while in 2,3-dimethyl butane is less than -1000 . For other hexane isomers the coefficients have intermediate values.

SIMILARITY

The representation of the molecules by the Wiener sequences captures different aspects of molecular structure and will offer a different approach to molecular similarity. For example, if we focus attention on heptanes we find that the most similar pairs of structures, when structures are represented by the Wiener sequences, are the following:



Incidentally the above pairs are the smallest alkanes (trees) with the same Wiener number. Alternative schemes (using other invariants, like paths etc.) usually point to *n*-alkane and 2-methylalkanes as the most similar pair of alkanes. Clearly Wiener numbers and higher Wiener numbers capture a different structural relationship. This is also indirectly reflected in the order in which the molecules appear in Tables IV and V, which is different from the lexicographic ordering or the ordering based on *n*-tuples of Knop and collaborators.⁴² In the Wiener sequences the shorter path numbers have greater weights. Consequently, since the shorter paths, in particular paths of length 2 and paths of length 3, dominate the variations of most of molecular thermodynamical properties,⁴³ one would expect that the most similar pairs of structures found in this way will also show the most similar thermodynamic properties. This, however, is not the case. For example the boiling points of 3-ethylpentane and 2,4-dimethylpentane differ by more than 10 °C (93.48 and 80.50 °C, respectively), and a similar difference in the BP is found among 2,2-dimethylpentane and 2,3-dimethylpentane (79.20 and 89.78 °C, respectively). The reason for this discrepancy is the already mentioned loss of information in the leading members of the *W* sequences. The heptane isomers shown above have the same leading members in the *W* sequence, and their similarity is thus reduced to differences in the number of paths of length 4, which are not so critical in determining the variations in molecular properties.

In conclusion, let us comment on the prevailing "interest" among chemical graph theorists to molecular graphs rather than molecular structures. Of course, without relationship to molecular structure, graphs would be of limited interest for chemistry. We already mentioned extensions to heteroatomic structures and to structures embedded in the 3-dimensional space. The significance of associating with graphs matrices, like the Wiener matrix considered here, is that the transition to molecules having heteroatoms and molecules in 3-dimensional space are possible. Although Wiener applied his approach based on *W* and *P* parameters to compounds other than hydrocarbons (to alcohols, amines, and fatty acids), these have been compounds in which the heteroatom was in a "fixed" position, and was thus ignored when *W* was calculated. To

extend *W* to molecules with a heteroatom is by no means apparent—until we suggested *W* matrices. Now the diagonal entries in such matrices allow one to discriminate among atoms of different kinds. It is this aspect of the newly introduced Wiener matrices that perhaps represents the most significant advance by allowing a "natural" extension of useful molecular descriptors to molecules of wider interest in chemistry.

ACKNOWLEDGMENT

We thank Dr. G. Maggiora and the Upjohn Co. from Kalamazoo, MI, for partial support of this work.

REFERENCES AND NOTES

- (1) Randić, M. In search of structural invariants. *J. Math. Chem.* **1992**, *9*, 97–146.
- (2) Rouvray, D. H. Predicting chemistry from topology. *Sci. Am.* **1986**, *254*, 40–47.
- (3) Katritzky, A. R.; Gordeeva, E. V. Traditional topological indices vs electronic, geometrical, and combined molecular descriptors in QSAR/QSPR research. *J. Chem. Inf. Comput. Sci.*, submitted for publication.
- (4) Basak, S. C.; Magnuson, V. R.; Niemi, G. J.; Regal, R. R.; Veith, G. D. Topological indices: Their nature, mutual relatedness, and applications. *Math. Modell.* **1986**, *8*, 300–305.
- (5) Randić, M. Generalized molecular descriptors. *J. Math. Chem.* **1991**, *7*, 155–168.
- (6) Read, R. C. A new system for the designation of chemical compounds. 1. Theoretical preliminaries and the coding of acyclic compounds. *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 135–149.
- (7) Malinowski, E. R. *Factor Analysis in Chemistry*, (2nd ed.); Wiley-Interscience: New York, 1991.
- (8) Randić, M. On the characterization of the molecular branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609–6615.
- (9) Kier, L. B.; Murray, W. J.; Randić, M.; Hall, L. H. Molecular connectivity V: Connectivity series applied to density. *J. Pharm. Sci.* **1976**, *65*, 1226–1230.
- (10) Randić, M. Representation of molecular graphs by basic graphs. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 57–69.
- (11) Wiener, H. Structural determination of paraffin boiling points. *J. Am. Chem. Soc.* **1947**, *69*, 17–20.
- (12) 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.
- (13) Randić, M. Comparative regression analysis. Regressions based on a single descriptor. *Croat. Chem. Acta*, in press.
- (14) Randić, M.; Brisse, G. M.; Spencer, R. B.; Wilkins, C. L. Search for all self-avoiding paths for molecular graphs. *Comput. Chem.* **1979**, *3*, 5–13.
- (15) Filip, P. A.; Balaban, T. S.; Balaban, A. T. A new approach for devising local graph invariants: Derived topological indices with low degeneracy and good correlation ability. *J. Math. Chem.* **1987**, *1*, 61–83.
- (16) Harary, F. *Graph Theory*; Addison-Wesley: Reading, MA, 1969.
- (17) Trinajstić, N. *Chemical Graph Theory*, 2nd ed.; CRC Press: Boca Raton, FL, 1992.
- (18) Rouvray, D. H. In *Chemical Applications of Topology and Graph Theory*; King, R. B., Elsevier: Amsterdam, 1983; p 159.
- (19) Tratch, S. S.; Stankevitch, M. I.; Zefirov, N. S. Combinatorial models and algorithms in chemistry. The expanded Wiener number—a novel topological index. *J. Comput. Chem.* **1990**, *11*, 899–908.
- (20) Hall, L. H. Computational aspects of molecular connectivity and its role in structure-property modeling. *Computational Graph Theory*; Nova Publishing: New York, 1990; pp 202–233.
- (21) Randić, M. To be published.
- (22) Randić, M. Molecular topographic descriptors. *Studies Phys. Theor. Chem.* **1988**, *54*, 101–108. Randić, M. On characterization of three-dimensional structures. *Int. J. Quantum Chem: Quantum Biol. Symp.* **1988**, *15*, 201–208. Randić, M.; Jerman-Blazic, B.; Trinajstić, N. Development of 3-dimensional molecular descriptors. *Comput. Chem.* **1990**, *14*, 237–246. Balasubramanian, K. Geometry-dependent characteristic polynomial of molecular structures. *Chem. Phys. Lett.* **1990**, *169*, 224–228.
- (23) Randić, M.; Kleiner, A. F.; DeAlba, L. M. Distance-distance matrices for graphs embedded on 3-dimensional lattices. Proceedings of the 5th International Conference on Chemical Graph Theory, Kansas City, May 1993; to appear in *J. Chem. Inf. Comput. Sci.*
- (24) Crippen, G. M. A novel approach to calculation of conformation: Distance geometry. *J. Comput. Phys.* **1977**, *24*, 96–107.
- (25) Wilson, E. B.; Decius, J. C.; Cross, P. C. *Molecular Vibrations*; McGraw-Hill: New York, 1955.
- (26) Balandin, A. A. Strukturnaja algebra v khimii. *Usp. Khim.* **1940**, *9*, 390.
- (27) Spialter, L. The atom connectivity matrix (ACM) and its characteristic polynomial (ACMCP): A new computer oriented chemical nomenclature. *J. Am. Chem. Soc.* **1963**, *85*, 2012–2013.

- (27) Randić, M. On computation of optimal parameters for multivariate analysis of structure-property relationship. *J. Comput. Chem.* **1991**, *12*, 970-980. Randić, M. Novel graph theoretical approach to heteroatoms in quantitative structure-activity relationship. *Chemom. Intell. Lab. Syst.* **1991**, *10*, 213-227.
- (28) Randić, M.; Dobrowolski, J. Cz. Optimal molecular connectivity descriptors for nitrogen containing molecules. *J. Math. Chem.*, submitted for publication. Randić, M. *Acta Pharm.*, submitted for publication. Randić, M.; Basak, S. C. Work in progress.
- (29) Randić, M. Random walks and their diagnostic value for characterization of atomic environment. *J. Comput. Chem.* **1980**, *1*, 386-399.
- (30) Randić, M. Restricted random walks and molecular properties. *Theor. Chem. Acta*, submitted for publication.
- (31) Randić, M. Similarity based on extended basis descriptors. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 686-692.
- (32) Randić, M. Novel molecular descriptor for structure-property studies. *Chem. Phys. Lett.*, submitted for publication.
- (33) Platt, J. R. Prediction of isomeric differences in paraffin properties. *J. Phys. Chem.* **1952**, *56*, 328-336.
- (34) Muller, W. R.; Szymanski, K.; Knop, J. V.; Trinajstić, N. An algorithm for construction of the molecular distance matrix. *J. Comput. Chem.* **1987**, *8*, 170-173. Mohar, B. A novel definition of the Wiener index. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 153-154. Mohar, B.; Pisanski, T. How to compute the Wiener index of a graph. *J. Math. Chem.* **1988**, *2*, 267-277. Mohar, B.; Babic, D.; Trinajstić, N. A novel definition of the Wiener index for trees. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 153-154. Klein, D. J.; Mihalic, Z.; Plavšić, D.; Trinajstić, N. Molecular topological index: A relation to the Wiener index. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 304-305. Lukovits, I. The generalized Wiener index for molecules containing double bonds and the partition coefficients. *Rep. Mol. Theory* **1990**, *1*, 127-131. Lukovits, I. Wiener indices and partition coefficients of unsaturated hydrocarbons. *Quant. Struct.-Act. Relat.* **1990**, *9*, 227-231. Maier, B. J. Wiener and Randić topological indices for graphs. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 87-90.
- (35) Bogdanov, B.; Nikolic, S.; Trinajstić, N. On the three-dimensional Wiener number, A comment. *Math. Chem.* **1989**, *3*, 299-309.
- (36) Randić, M. In search of structural invariants. *J. Math. Chem.* **1992**, *9*, 97-146.
- (37) Balaban, A. T. Highly discriminating distance-based topological index. *Chem. Phys. Lett.* **1982**, *89*, 399-404.
- (38) Randić, M. On molecular identification numbers. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 164-175.
- (39) Szymanski, K.; Muller, W. R.; Knop, J. V.; Trinajstić, N. On M. Randić's identification numbers. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 413-415.
- (40) Randić, M. Molecular ID numbers: By design. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 134-136.
- (41) Szymanski, K.; Muller, W. R.; Knop, J. V.; Trinajstić, N. Molecular ID numbers. *Croat. Chem. Acta* **1986**, *59*, 719-723.
- (42) Knop, J. V.; Muller, W. R.; Jericevic, Z.; Trinajstić, N. Computer enumeration and generation of trees and rooted trees. *J. Chem. Inf. Comput. Sci.* **1993**, *21*, 91-99.
- (43) Randić, M.; Wilkins, C. L. Graph theoretical ordering of structures as a basis for systematic searches for regularities in molecular data. *J. Phys. Chem.* **1979**, *83*, 1525-1540.