

Random Walks and Their Diagnostic Value for Characterization of Atomic Environment*

Milan Randić

Ames Laboratory, U.S. Department of Energy, and the Chemistry Department, Iowa State University, Ames, Iowa 50011
Received September 10, 1979; accepted June 9, 1980

A characterization of atomic environments based on counting random walks in a molecular skeleton is outlined. To each atom in a molecule a sequence of integers $w_1, w_2, w_3, \dots, w_n$ is assigned, where w_i represents the number of self-returning walks of length k , the length being defined by the number of bonds traversed. Properties of the derived atom codes are discussed. The codes display an impressive diversity and are superior to atomic codes based on enumeration of self-avoiding walks (or paths) in discriminating atomic environments. In certain cases the codes of individual atoms are not unique and the same codes appear in different molecules or even within the same molecule. The occurrence of the nonunique codes can be related to special structural situations, associated with the occurrence of isospectral graphs. These isospectral graphs which have atoms with identical codes can generate additional isospectral structures by attaching any arbitrary group to such points. If nonequivalent atoms of a single molecule have identical random walk codes, substitution at the singular points alternatively will produce isospectral graphs. Examples of such situations are given.

INTRODUCTION

Identification and recognition of identical structures remains one of the central problems in many mathematical and chemical studies involving manipulation with structures or graphs representing chemical systems. In more abstract form this is the problem of graph isomorphism, which has received considerable attention in the literature.¹ Basically there are two avenues for resolving the problem, either (1) an effort is made to arrive at some canonical labeling of vertices (atoms in a molecule), or (2) we search for a set of graph invariants hoping that the collection adopted will be adaptable to unforeseen circumstances. Both approaches have their strong and weak sides: canonical numbering is conceptually simple, rules can be formulated, but the main difficulty is in implementation, there being no assurance that a

case may not be found in which an otherwise practical approach may fail. When searching for graph invariants even greater uncertainty exists concerning extending such schemes to yet uninvestigated cases or types of structure. If we find that two graphs are not isomorphic, both approaches are equally useful; but if we find that two graphs show the same characteristics using selected graph invariants, we still cannot be quite certain whether the result may not change if additional discriminators are used. The essential difficulty here is due to our lack of knowledge of *completeness* of a collection of graph invariants. On the other hand, if we succeed in applying the canonical labeling test and we find two structures to be isomorphic, we may be certain that this is the case. A particular canonical labeling system has been proposed based on the concept of the smallest binary code for a graph.² The code is derived by interpreting the rows of the adjacency matrix as binary numbers and a labeling is sought which will

* Dedicated to Professor A. T. Balaban for his pioneering work in reviving interest in chemical graph theory.

produce for the matrix the smallest binary number if the rows are read from left to right and from top to bottom. Besides the use of such canonical labelings for graph isomorphism investigations, they have been found useful in a number of closely related problems: graph symmetry,³ construction of graphs,⁴ graph ordering.⁵ Although canonical numbering of atoms or vertices in a graph gives a unique (up to the automorphism group) discrimination of atomic environments, and occasionally the labels show some structural dependence,* such an approach, based on a convention, is not the most adequate for comparison of atomic environments, discussing differences, etc. In this article we will concentrate on the problem of structural characterization of atomic environments. A useful set of graph invariants that may serve in an isomorphism test, at least under some restricted conditions, will also be obtained.

It is desirable to have a *unique* characterization for atomic environments with the following properties: (1) relative ease of derivation, (2) the possibility for verification of the legitimacy of the collection, and (3) the availability of a reconstruction algorithm. In short, one would like to have a *complete* set of graph invariants that characterize a structure and do not require inordinate computational efforts. We do not require in advance that such a scheme provides insight into atomic and molecular properties, but if it does yield some results in that direction, it is even more useful. The first of the three mentioned desirable features for a characterization is self-evident, the second—legitimacy—is required in order to ensure that an error in one or more codes can be detected and that no time is wasted with codes that are inaccurate. Finally, the last condition, if it can be demonstrated for a general case, would provide the answer on uniqueness and completeness of the adopted graph invariants for characterization of a structure. From the literature, in particular the recent applications of the concept of paths or self-avoiding walks in composing atomic and molecular codes,⁶ one sees considerable progress in

the efforts directed toward characterization of atoms and molecules and discussion of their properties using exclusively a structural approach.⁷ Paths and walks are closely related (*vide infra*), but the count of walks gives generally much larger numbers and hence has a potential to lead to different sequences typifying atomic environments. For path counts, frequently constitutionally nonequivalent atoms in the same molecule or atoms in different molecules will give the same path sequence numbers. Hence, not the individual codes but the collection of codes for a molecule could form a basis for discriminating among structures. However, the discriminating power of path codes for atomic environments is limited, though still of use and of possible interest when discussing atomic properties. After all, many different atoms and atomic environments can be associated with similar atomic properties. Here we will be primarily interested in these atomic environments and will seek a way to assign to different atomic environments different structural codes. Our basic graph invariants to be examined are *walks*, random walks, self-returning walks, and generally walks of different length.

RANDOM WALKS AND THEIR ENUMERATION

For formal definitions of a graph and various graph invariants, such as a walk, a path, the topological distance, and other graph theoretical definitions, the reader should consult available textbooks.⁸ Unfortunately, there is no generally agreed upon nomenclature for many graph theoretical concepts and even the term *walk* has been used differently by different authors, which necessitates defining the term as is used in this work. Let us distinguish the following:

Random walk in a graph is a sequence of edges which can be continuously traversed, starting from any vertex and ending on any vertex. The repeated use of the same edge or edges is permitted.

Self-returning walk is a random walk starting and finishing at the same vertex.

Self-avoiding walk or path is a sequence of edges which can be continuously traversed, start-

* This topic has not been fully explored although a few illustrations are found in refs. 2–5.

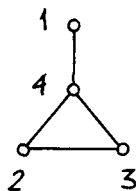


Figure 1. Molecular graph of the carbon skeleton of methylcyclopropane.

ing from any vertex and ending on a different vertex, and in which no vertex is visited more than once.

The *distance* of a walk or a path is defined by the number of edges involved in the walk or path.*

For a graph having few vertices and edges it is not difficult from visual inspection to enumerate all walks of desired lengths.⁹ The walks, but not the paths, are easy to enumerate indirectly, however, by considering different powers of the adjacency matrix \mathbf{A} of a graph.⁸ One can interpret the elements of the matrix \mathbf{A}^2 as walks of length 2, more precisely, the product $a_{ij}a_{jk}$ which enters the matrix \mathbf{A}^2 is not zero only if vertices i and k are connected by a walk of length 2 passing through their common neighbor j . In such a way we derive all walks of length 2. The same argument can be extended to triple products $a_{ij}a_{jk}a_{kl}$ for deriving the number of walks of length 3 and so on. Hence, instead of direct tracing of walks and their enumeration, all we have to do is to calculate powers \mathbf{A}^n of interest. In this respect, the enumeration of walks of different length is simple in comparison with counting paths.[†] Moreover, since the adjacency matrix \mathbf{A} satisfies the Cayley–Hamilton theorem (i.e., the characteristic polynomial is zero for substituting x^k terms with a_{ij}^k values for any i, j) the first n powers of \mathbf{A} (n being the number of

vertices in the graph considered) suffice to generate any higher values of \mathbf{A}^k .

Let us illustrate the derivation of the number of walks of different length on a simple graph depicting the carbon skeleton of methylcyclopropane (Fig. 1). There are four vertices and the first four powers of the adjacency matrix are

$$\begin{array}{c} \mathbf{A} \qquad \mathbf{A}^2 \qquad \mathbf{A}^3 \qquad \mathbf{A}^4 \\ \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 0 & 1 & 1 & 3 \end{pmatrix} \begin{pmatrix} 0 & 1 & 1 & 3 \\ 1 & 2 & 3 & 4 \\ 1 & 3 & 2 & 4 \\ 3 & 4 & 4 & 2 \end{pmatrix} \begin{pmatrix} 3 & 4 & 4 & 2 \\ 4 & 7 & 6 & 6 \\ 4 & 6 & 7 & 6 \\ 2 & 6 & 6 & 11 \end{pmatrix} \end{array}$$

The characteristic polynomial $\text{Ch}(x)$ can be derived conveniently using graph theoretical schemes.^{10,11} For the above example, one obtains¹³

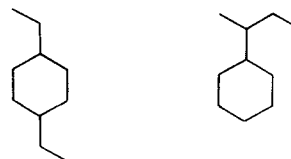
$$\text{Ch}(x) = x^4 - 4x^2 - 2x + 1$$

One can easily verify that the Cayley–Hamilton theorem, which leads to

$$\mathbf{A}^4 - 4\mathbf{A}^2 - 2\mathbf{A} + (1) = (0)$$

is satisfied, where (1) and (0) are the unit matrix and the zero matrix, respectively. The sequence $\mathbf{A}, \mathbf{A}^2, \mathbf{A}^3, \mathbf{A}^4$, with the knowledge of the characteristic polynomial, contains all the pertinent information on atomic walks in the carbon skeleton. Since matrix \mathbf{A} already contains the full information on a graph, the sequence of powers of \mathbf{A} has redundant information. The question is now in selecting the most convenient data for a graph by extracting hopefully complete information from the available collection of \mathbf{A}^n matrices to be used instead of the initial data supplied by \mathbf{A} .

The disadvantage of the adjacency matrix is that it is not *invariant*, its form depends on the adopted numbering of the vertices. The characteristic polynomial is invariant but is not unique to the structure. There are many graphs, such as the pair corresponding to carbon skeletons of 1,4-divinylbenzene and 2-phenylbutadiene¹³:



* In graph theory Cayley in 1875 had defined metrics, the distance (frequently referred to as topological distance) is measured by the numbers of bonds or edges traversed and as such is clearly graph invariant.

† Counting of the path involves $n!$ character for the search and is therefore limited in application to graphs of not too complex structure or size.

which have identical characteristic polynomials and are referred to as isospectral.¹⁵ The characteristic polynomial, which represents an expansion of the secular determinant in the well-known Hückel MO method, is related to random walks.¹⁶ The coefficients of the x^{n-j} term of the characteristic polynomial are easily derived if all self-returning walks of length j are known for the corresponding molecular graph. This relationship is reflected in the equivalence of the information given by the characteristic polynomial and spectral moments of a graph.¹⁷

Spectral moments have rarely been evoked in discussions of isospectral molecules (or any molecule) and their properties, although spectral moments, defined as the trace of \mathbf{A}^k matrices ($k = 1, 2, 3, \dots$), are readily derived and because they are integers they are suited for simple manipulations and comparisons. For the graph of methylcyclopropane (Fig. 1) one immediately finds the following sequence of spectral moments: 0, 8, 6, 28, One can continue the sequence with still higher spectral moments, but one also has to be aware that the sequence of spectral moments, despite being infinite, need not be unique to a graph. It is strictly equivalent to either the collection of associated eigenvalues (or spectrum) or to the information contained in the coefficients of the characteristic polynomial, which are already known to be nonunique. The deficient character of the spectrum, spectral moments, or the characteristic polynomial for representing a structure is not surprising. The adjacency matrix which defines a graph contains n^2 pieces of information, or in view of its symmetric character and zero diagonals $n(n-1)/2$ pieces of information. Hence, one should not be surprised that n coefficients or n roots may already have lost some essential initial structural information for a graph. There is no guarantee that a selection of n^2 pieces of information will necessarily provide a complete characterization of a graph, but at least one might expect to reduce the number of coincidences of equal sets of parameters belonging to different structures. The task is then to select n^2 invariants (or whatever the needs may dictate) rather than use

n^2 variables contained in an arbitrarily labeled matrix \mathbf{A} .

CHARACTERIZATION OF ATOMIC ENVIRONMENT BY ENUMERATION OF SELF-RETURNING WALKS OF DIFFERENT DISTANCE

We propose the use of n^2 diagonal elements of the first n powers of the adjacency matrix \mathbf{A} of a graph for characterization of the structure. As will be seen the derived codes will also represent a useful characterization of the individual atomic environments. An evaluation of the proposed parameters is straightforward requiring standard matrix multiplication. The selection of *individual* n diagonal elements a_{ii} of increasing powers of \mathbf{A}^k ($k = 1, 2, 3, \dots$) allows one to associate such sequences of self-returning walks with *individual* atoms. If desired one can continue to evaluate even higher powers of \mathbf{A}^k and extend the sequences indefinitely. For carbon atoms 1–4 of methylcyclopropane, one obtains the following sequences:

atom	\mathbf{A}	diagonal elements in			
		\mathbf{A}^2	\mathbf{A}^3	\mathbf{A}^4	...
1	0	1	0	3	...
2	0	2	2	7	...
3	0	2	2	7	...
4	0	3	2	11	...

It is important here—at the beginning of the exploration of the new codes—to point to the following important features of the approach: first, observe that only part of the available information is used. We selected the diagonal elements of \mathbf{A}^k matrices, but off-diagonal elements also contain useful structural information. Therefore, in case of ambiguities that may arise when different atomic environments have the same count of self-returning walks, there is the possibility of using these additional pieces of information in an attempt to resolve the problem and derive a unique characterization. Second, although we may find in many trials that nonequivalent atoms have different sequences for the count of self-returning walks of increasing length, there are no firm the-

oretical grounds to buttress such an expectation for the general case.

In the next section, we examine a number of acyclic structures, a selection of polycyclic structures, and finally several isospectral graphs in order to demonstrate the versatility of the walk-based atomic codes and also to obtain some insight into the sensitivity of the codes to changes in distant surroundings. As will be seen the derived atomic codes show a remarkable individuality. The important question to be resolved is that of the uniqueness of the walk-based atomic codes, and we will approach this question by screening a large number of graphs in a search for counterexamples, i.e., graphs in which constitutionally nonequivalent vertices have the same list of self-returning walks of different length. It will be seen that indeed there are instances of this rather unusual situation, but as the discussion will show, they can be analyzed. Hence, the characterization of atomic environments by counting self-returning walks may, despite occasional apparent failures, be useful and of diagnostic value. Walks, self-returning walks, and random walks appear to reflect subtleties of molecular connectivity in a way which seems worth further examination.

ACYCLIC STRUCTURES

In Table I, we list the atomic codes based on enumeration of self-returning walks for the carbon skeletons of all 32 isomers of hexane (C_6H_{14}), heptane (C_7H_{16}), and octane (C_8H_{18}). The examples illustrate the degree of variations in walk counts found for the selected atomic environments. In acyclic structures, all odd-length self-returning walks are necessarily zero, so for brevity we listed only even-length walks in the sequence of walk counts. Many of the structures have several equivalent carbon atoms which are shown only once. In addition, we include, as the last line for each structure, the molecular spectral moments which can be derived by adding all atomic sequences and which correspond to the count of self-returning walks for a molecule as a whole.

The walk-based atomic codes show a number of

interesting numerical properties. First, we observe that there is no single case of identical codes among hexane and heptane isomers. This also applies to octane isomers, except that since we have truncated the octane atomic codes to a radius of four bond lengths (i.e., matrices A^8), all carbon atomic codes which have the same *local* environment of this radius will show the same sequence. However, when higher powers of A are considered the apparent coincidences disappear. Thus, the sequence 1, 2, 5, 14 which arises in *n*-octane, 2-methylheptane, and 3-methylheptane becomes: 1, 2, 5, 14, 42, . . .; 1, 2, 5, 14, 42, . . ., and 1, 2, 5, 14, 43, . . ., respectively, by considering A^{10} matrices. The first two sequences become 1, 2, 5, 14, 42, 132, . . . and 1, 2, 5, 14, 42, 175, . . ., respectively, when considering A^{12} . Eventually the same will happen with other sequences which appear the same upon truncation with powers of A^8 since the characteristic polynomials which serve as the recursion expressions are *different* for different isomers of hexane, heptane, and octane.

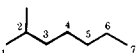
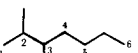
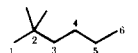
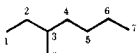
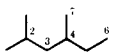
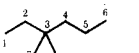
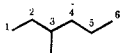
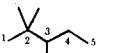
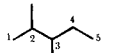
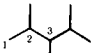
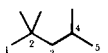
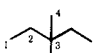
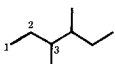
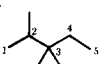

One may compare the individuality of the walk-based atomic codes with the results for path-based atomic codes for the same molecules. Within the group of hexane isomers there are instances of the occurrence of the same path code for atomic environments in different molecules. Among heptanes and octanes the number of coincidences is quite large, and for 2,3-dimethylpentane, for example, none of the seven path-based atomic codes is unique to the structure. Thus, a relatively simple structure such as 2,3-dimethylpentane can be characterized uniquely by the *collection* of data belonging to all atoms. Even this statement, strictly speaking, is only a conjecture, although it appears that at least for acyclic structures the conjecture may hold.* In contrast, the nonequivalent environments appearing in hexanes, heptanes, and octanes are uniquely described by the corresponding atomic

* The conjecture has been verified, by C. A. Shelley and M. Trulson (private communication, 1978) for alkane isomers having 14 or less carbon atoms, in all over 3000 structures. No serious attempt has yet been made to see if the conjecture is valid for polycyclic systems.

Table I. Listing of the number of self-returning walks of different length for nonequivalent atoms of carbon skeletons of hexane, heptane, and octane isomers.

Hexane isomers					
	1 2 3	1, 2, 5, 14 2, 5, 14, 42 <u>2, 6, 19, 61</u> <u>10, 26, 76, 234</u>		1 2 3 6	1, 2, 6, 21, 77 2, 6, 21, 77, 286 3, 11, 41, 153, 571 <u>1, 3, 11, 41, 153</u> <u>10, 30, 106, 390, 1450</u>
	1 2 3 4 5	1, 3, 10, 35, 125 3, 10, 35, 125, 450 2, 7, 25, 90, 325 2, 5, 15, 50, 175 <u>1, 2, 5, 15, 50</u> <u>10, 30, 100, 350, 1375</u>		1 2 3 4	1, 3, 11, 43, 3, 11, 43, 171, <u>10, 34, 130, 514</u> 1, 4, 17, 73 4, 17, 73, 314 2, 7, 29, 124 <u>1, 2, 7, 29</u> <u>10, 38, 160, 686</u>
Heptane isomers					
	1 2 3 4	1, 2, 5, 14, 42 2, 5, 14, 42, 132 2, 6, 19, 62, 206 <u>2, 6, 20, 68, 232</u> <u>12, 32, 58, 304, 992</u>		1 2 3	1, 3, 10, 36, 128 3, 10, 36, 136, 528 <u>2, 8, 32, 128, 512</u> <u>12, 40, 148, 556, 2080</u>
	1 2 4 5 6 7	1, 3, 10, 35, 126 3, 10, 35, 126, 461 2, 7, 26, 97, 362 2, 6, 20, 70, 251 2, 5, 14, 43, 142 <u>1, 2, 5, 14, 43</u> <u>12, 36, 120, 410, 1511</u>		1 2 4 5 6 7	1, 3, 11, 44, 182 3, 11, 44, 182, 752 3, 12, 50, 210, 884 1, 3, 12, 50, 210 2, 6, 22, 88, 364 <u>1, 2, 6, 22, 88</u> <u>12, 40, 156, 640, 2662</u>
	1 2 3 4 5 6 7	1, 2, 6, 21, 78 2, 6, 21, 78, 297 3, 11, 42, 162, 672 1, 3, 11, 42, 162 2, 7, 26, 99, 381 2, 5, 15, 51, 186 <u>1, 2, 5, 15, 51</u> <u>12, 36, 126, 468, 1827</u>		1 2 3 4 5	1, 4, 17, 74, 352 4, 17, 74, 325, 1432 2, 8, 34, 148, 650 2, 5, 16, 61, 254 <u>1, 2, 5, 16, 61</u> <u>12, 44, 180, 772, 3453</u>
	1 2 3	1, 2, 6, 22, 86 2, 6, 22, 86, 342 3, 12, 48, 192, 768 <u>12, 36, 132, 516, 2052</u>		1 2 3 4	1, 2, 7, 30, 135 2, 7, 30, 135, 614 4, 18, 82, 374, 1706 <u>1, 4, 18, 82, 374</u> <u>12, 44, 192, 868, 3952</u>
	1 2 3 4	1, 4, 18, 84, 396 4, 18, 84, 396, 1872 3, 12, 54, 252, 1188 <u>1, 3, 12, 54, 252</u> <u>12, 48, 216, 1008, 4752</u>		1 2 3 4	1, 4, 18, 84, 396 4, 18, 84, 396, 1872 3, 12, 54, 252, 1188 <u>1, 3, 12, 54, 252</u> <u>12, 48, 216, 1008, 4752</u>
Octane isomers					
	1 2 3 4	1, 2, 5, 14, 2, 5, 14, 42, 2, 6, 19, 62, 2, 6, 20, 69,		1 2 3	1, 3, 10, 35, 3, 10, 35, 127, 2, 7, 27, 106,

Table I (continued from previous page)

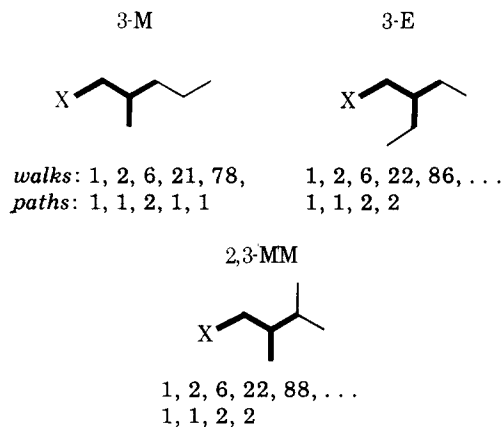
Octane isomers											
											
1	1, 3, 10, 30,		1	1, 3, 11, 44,		1	1, 4, 17, 74,				
2	3, 10, 35, 126,		2	3, 11, 44, 183,		2	4, 17, 74, 326,				
3	2, 7, 26, 98,		3	3, 12, 51, 219,		3	2, 8, 35, 155,				
4	2, 6, 21, 77,		4	2, 7, 27, 110,		4	2, 6, 21, 81,				
5	2, 6, 19, 63,		5	2, 5, 15, 52,		5	2, 5, 14, 44,				
6	2, 5, 14, 42,		6	1, 2, 5, 15,		6	1, 2, 5, 14,				
7	1, 2, 5, 14,		7	1, 3, 12, 51,							
											
1	1, 2, 6, 21,		1	1, 3, 10, 36,		1	1, 2, 7, 30,				
2	2, 6, 21, 78,		2	3, 10, 36, 137,		2	2, 7, 30, 136,				
3	2, 11, 42, 163,		3	2, 8, 33, 137,		3	4, 18, 83, 385,				
4	2, 7, 27, 106,		4	3, 11, 43, 173,		4	2, 8, 35, 159,				
5	2, 6, 20, 71,		5	2, 6, 21, 79,		5	2, 5, 16, 62,				
6	2, 5, 14, 43,		6	1, 2, 6, 21,		6	1, 2, 5, 16,				
7	1, 2, 5, 14,		7	1, 3, 11, 43,		7	1, 4, 18, 83,				
8	1, 3, 11, 42,										
											
1	1, 2, 6, 22,		1	1, 4, 18, 85,		1	1, 3, 11, 45,				
2	2, 6, 22, 87,		2	4, 18, 85, 409,		2	3, 11, 45, 193,				
3	3, 12, 49, 201,		3	3, 13, 61, 293,		3	3, 13, 57, 251,				
4	2, 7, 27, 108,		4	2, 6, 23, 101,		4	2, 6, 23, 97,				
5	2, 5, 15, 52,		5	1, 2, 6, 23,		5	1, 2, 6, 23,				
6	1, 2, 5, 15,		6	1, 3, 13, 61,							
											
1	1, 3, 11, 45,		1	1, 4, 17, 75,		1	1, 2, 7, 31,				
2	3, 11, 45, 195,		2	4, 17, 75, 338,		2	2, 7, 31, 146,				
3	3, 13, 59, 269,		3	2, 9, 41, 188,		3	4, 19, 91, 436,				
4	1, 3, 13, 59,		4	3, 10, 37, 149,		4	1, 4, 19, 91,				
			5	1, 3, 10, 37,							
											
1	1, 2, 6, 22,		1	1, 3, 12, 55,		1	1, 4, 19, 97,				
2	2, 6, 22, 89,		2	3, 12, 55, 265,		2	4, 19, 97, 508,				
3	3, 12, 51, 221,		3	4, 19, 93, 458,							
4	1, 3, 12, 51,		4	2, 7, 31, 148,							
			5	1, 2, 7, 31,							
			6	1, 4, 19, 93,							

codes which count self-returning walks of different length rather than self-avoiding walks of different length.

The unique atomic codes of lower alkenes allow

an ordering of atomic environments. As a rule for ordering we compare the first entries in the sequences, and if they are different they define the relative order. If they are equal, one proceeds to

the second entries and compares them in order to decide on the relative order. If the second entries are also equal, the process is continued until one arrives at the first entries in which the two sequences differ. In Table II we show the list of ordered atomic environments for all nonequivalent *terminal* environments in heptanes. One could order in a similar way the eight nonequivalent terminal environments of hexanes or 39 nonequivalent environments of terminal carbon atoms in octanes, and one could combine these with the 17 terminal environments of heptanes in a single list. In this way one arrives at a systematic cataloging of atomic environments for acyclic structures, regardless of the size of the structure, and, of course, the nonterminal atoms can be similarly arranged. Then similarities in sequences would reflect the similarities in the corresponding atomic environments. The sequences with the same initial entries correspond to atomic environments which are identical up to a critical radius. For instance the initial values for walks of 1, 2, 6, ... appear in the following heptanes (for atoms marked with X):



The common parts of three skeletons which give rise to the same count of walks of length 6 (which happens to be 6 in this case) are indicated by thick lines. Some parallelism between paths and walks persists despite the lower discriminatory power of the former and their finite character: the walk-based codes for 3-E and 2,3-MM heptanes are identical for the first four entries and the corresponding path-based codes are also identical, while

Table II. Lexical ordering of the walk-based atomic codes for terminal carbon atoms of various heptane isomers.

Isomer	Self-returning walks
<i>n</i>	1, 2, 5, 14, 42
2-M	1, 2, 5, 14, 43
3-M	1, 2, 5, 15, 51
2,2-MM	1, 2, 5, 16, 61
3-M	1, 2, 6, 21, 78
3-E	1, 2, 6, 22, 86
2,3-MM	1, 2, 6, 22, 88
3,3-MM	1, 2, 7, 30, 135
2-M	1, 3, 10, 35, 126
2,4-MM	1, 3, 10, 36, 128
3-M	1, 3, 11, 42, 162
2,3-MM	1, 3, 11, 44, 182
2,3-MM	1, 3, 12, 50, 210
2,2,3-MMM	1, 3, 12, 54, 252
2,2-MM	1, 4, 17, 74, 352
3,3-MM	1, 4, 18, 82, 374
2,2,3-MM	1, 4, 18, 84, 396

both codes differ slightly for the 3-M isomer when compared with those of 3-E and 2,3-MM molecules. The two environments 3-E and 2,3-MM are discriminated by higher walks, i.e., with the count of walks of length 10. The above example illustrates that the place of truncation of the codes of necessity may be beyond *n*, the number of vertices. In the above case we have seven vertices, but in order to differentiate among the two terminal environments of 3-E and 2,3-MM, we need powers of the adjacency matrix up to A^{10} . It is only when we confine the comparison of the codes for atoms of the *same* molecule that we need not go beyond the *n*th power of *A*, since if there are (nonequivalent) vertices having the same walk counts, the Cayley–Hamilton theorem ensures that *all* the higher counts will also be identical. Different characteristic polynomials ensure that initially equal lower walk numbers will produce different higher walks. Therefore, the only situations of further concern are (1) an appearance of the same walk count for nonequivalent atoms *within* a same structure, and (2) the appearance of the same walk counts for atoms in a pair of isospectral molecules. In both situations the initial equality in the walk counts will be transmitted to higher walks too, since such situations are characterized by the same recursive expression. Hence, two topics deserve

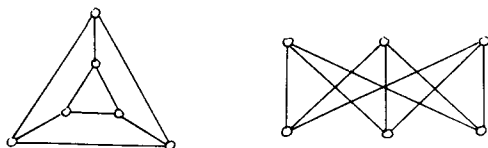


Figure 2. Graph depicting trigonal prism and the Kuratowski $K_{3,3}$ graph, the only regular trivalent graphs on six vertices.

more attention: (1) a systematic search for structures in which nonequivalent atoms may have the same walk-based atomic codes, and (2) a detailed study of isospectral graphs in search for atomic environments in different molecules having the same atomic codes.

The walk-based atomic codes of acyclic structures show some regularities also exhibited by path codes. The code of a terminal atom and its adjacent neighbor differ by a shift of the latter to the left by one place (accompanied by erasure of the first entry). Since the collection of path codes in acyclic structures has been conjectured to give a unique characterization of a structure,⁶ a similar conjecture for the collection of walk-based codes is even more likely to hold. Some walk sequences are related to others by a common factor, and in some other sequences the walks of higher rank are simply related to walks of lower length. Such properties are of considerable interest, particularly for the problem of graph reconstruction from a given list of atomic codes. A successful reconstruction is tantamount to the assertion that the collection of codes is unique to the structure—an important topic which is beyond the scope of the present work. Here, we can mention recent interesting work aimed at establishing a connection between the enumeration of walks and paths.¹⁷

POLYCYCLIC STRUCTURES

It is well known that algorithms that apply to acyclic structures need not apply to cyclic and polycyclic structures. It is therefore of interest to see how the properties of atomic codes based on a count of self-returning walks hold for cyclic structures. The proliferation of cyclic graphs, as

one increases the number of atoms and bonds, is so fast that an exhaustive analysis, which was possible for acyclic structures, is no longer practical even for systems having a half-dozen atoms. Instead, one usually selects a few classes of graphs, such as monocyclic, bicyclic, planar, maximally planar, bipartite, regular, etc., and looks for counterexamples which have the same code for nonequivalent atoms. The class that is most likely to exhibit such cases is the case of *regular* graphs. Such graphs, defined by an equal valency for all vertices, are more elusive in graph isomorphism tests.⁷ Among the simplest of such graphs are the Kuratowski $K_{3,3}$ graph, which is important for the characterization of the planarity of graphs, and a graph depicting a trigonal prism (Fig. 2). These are the only trivalent regular graphs on six vertices, and in both graphs all the vertices are equivalent. Therefore, it is sufficient to give the sequence of walk numbers for a single vertex alone in each case. Since we now encounter odd-membered cycles as well, walks of odd length are possible, hence we have to list all even and odd walks. The following is the result for the two graphs:

Kuratowski $K_{3,3}$: 0, 3, 0, 27, 0, 243, 0, 2187, 0, . . .

Trigonal prism: 0, 3, 2, 19, 40, 143, 332, . . .

The sequences have been initiated by zero, the number of self-returning walks of length 1; thus, we are reminded that both even and odd powers of the A matrix are included in the codes. Self-returning walks of length 1 can appear in more general graphs in which loops are allowed. Since the Kuratowski graph is bipartite, it has only even-membered self-returning walks, and because of its internal structure all walks are given by odd powers: $3, 3^3, 3^5, 3^7, \dots$

Next in size are the five trivalent graphs on eight vertices (Fig. 3). In Table III are listed the atomic codes for nonequivalent vertices. We also give the corresponding spectral moments, which tell us immediately that there are no isospectral pairs among the five graphs. Nevertheless, we find the same code in the structures B and D. By examining the corresponding matrices A^k and the rows corresponding to the same codes we also find that

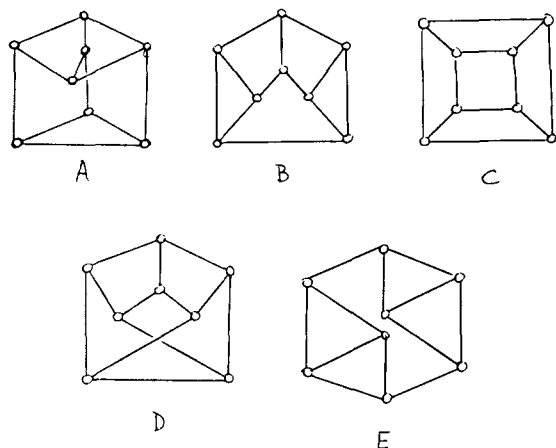


Figure 3. Regular graphs of valency three and having eight vertices.

associated random walks (i.e., the off-diagonal elements in \mathbf{A}^k matrices) are also equal. Intrigued by this peculiar persistence of equal numbers of walks for the respective vertices, we continued to evaluate still larger powers of \mathbf{A} , up to the 20th power.¹⁹ We find not only the same number of self-returning walks for the two different environments, but also that this is not an isolated case, and several such instances occur among trivalent graphs on ten vertices. If we view the two graphs

from the position of atoms 1 (which have identical walk counts), we see that the difference in connectivity comes from a crossover of the linking of the second neighbors. Apparently, as far as the environment of atom 1 is concerned, the enumeration of walks is highly insensitive if we switch the bridging of the more distant neighbors as implied in graphs B and D. It should be noted that if the same residual is attached to sites 1 in both graphs, the count of walks will not be different for the two atoms. Hence, only the smallest graphs showing these peculiarities are of special interest.

The important question that remains to be settled is the uniqueness of atomic codes for the above singular situations. It is not apparent how far one should go in raising powers of matrices \mathbf{A} with the expectation of eventually seeing a different number of walks for atoms in different structures. The graphs B and D are not isospectral, hence they use *different* recurrence relations. It is hard to imagine that different recurrence expressions can continue to yield the same higher walk counts indefinitely.*

* According to Professor A. Fink (Iowa State University, private communication) one can expect the differences to appear with $3n$ th power, and for common eigenvalues, even sooner.

Table III. The self-returning walks for vertices of the five regular trivalent graphs having eight vertices. The spectral moments (i.e., the sum of the walk-based atomic codes) are also shown.

Structure	Vertices	Self-returning walks (and spectral moments)
A	1,2	0, 3, 0, 21, 6, 159, 108, ...
	3,4,5	0, 3, 2, 15, 26, 105, 258, ...
	6,7,8	0, 3, 0, 19, 10, 143,
	Moments:	0, 24, 6, 144, 120, 1062, ...
B	1,8	0, 3, 0, 19, 10, 141, 154, 1109, 1764,
	2,3,6,7	0, 3, 2, 17, 22, 123, 220,
	4,5	0, 3, 2, 15, 26, 105, 260,
	Moments:	0, 24, 12, 136, 160, 984, 1708,
C	All	0, 3, 0, 21, 0, 183, 0, 1641, 0, 14763,
	Moments:	0, 24, 0, 168, 0, 1464, 0, 13128, 0, 118104,
D	All	0, 3, 0, 19, 10, 141, 154, 1109, 1764,
	Moments:	0, 24, 0, 152, 80, 1128, 1232, 8872, 14112,
E	1,2,5,6	0, 3, 2, 17, 24, 123,
	3,4,7,8	0, 3, 4, 17, 36, 123,
	Moments:	0, 24, 24, 136, 480, 984,

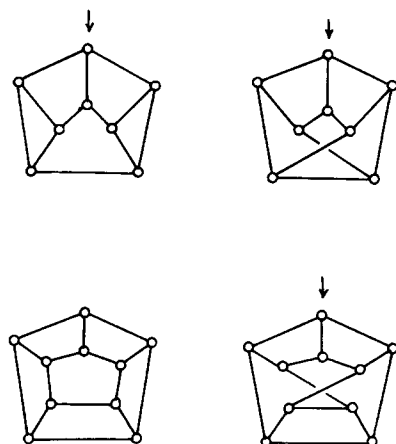
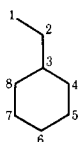


Figure 4. Regular graphs showing a large number of initial identical counts of self-returning walks. Arrows indicate the cases the sites of common initial count of walks.

ISOSPECTRAL GRAPHS REVISITED

As is well known,^{13,14} the skeleton of vinylbenzene

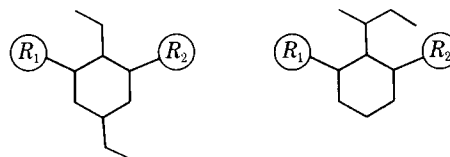


gives rise to many isospectral molecules of chemical interest. By substituting any residual at atom 2 or atom 6 we arrive at a pair of isospectral structures, such as 1,4-divinylbenzene and 2-phenylbutadiene. This suggests that the sites 2 and 6 have the same number of self-returning walks of different length (see Fig. 4). From the powers of the associated adjacency matrix one can easily find for vinylbenzene graph the following count of self-returning walks for all carbon atoms:

atom	self-returning walks
1	1, 2, 6, 22,
2	2, 6, 22, 90,
3	3, 12, 52, 235,
4, 8	2, 7, 30, 135,
5, 7	2, 6, 23, 98,
6	2, 6, 22, 90,

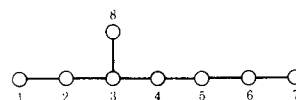
Because the graph is bipartite we have listed only walks of even length, with the odd-length walks

being zero. Atoms 2 and 6, which are not constitutionally equivalent, have the same code. Clearly an attachment of a fragment in turn at these sites will not change the molecular spectral moments for the two generated structures. If we examine all walks (i.e., random walks) for positions 2 and 6, we find that *even*-length random walks are identical for the two nonequivalent sites, but walks of *odd* lengths differ for the two sites. Consequently, at an *even* distance from sites 2 and 6 we can introduce a substitution without disturbing the isospectral character of vinylbenzene skeleton. A similar tampering with the sites at an odd distance from positions 2 and 6 will destroy the coincidental count of walks for the two sites and will produce different spectral moments. It follows from the above argument, without further necessity of proofs, that



are isospectral regardless the nature of the residuals R_1 and R_2 . This has been demonstrated previously,¹⁵ but now we see how it could have been anticipated.

Another graph with the above special property has been studied by Schwenk¹⁹:



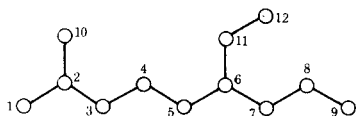
The positions 2 and 5 have the same list of self-returning walks: 2, 6, 21, 78, 298, 1157, ... (only even-length walks are given). Again we have within the *same* skeleton two nonequivalent sites having the same walk-based codes. Hence, substitutions at 2 and 5 alternatively will produce a pair of isospectral graphs. In contrast to the graph of vinylbenzene, in this case there are no sites that would preserve the isocodal character of atoms 2 and 5 by a substitution.

It is of considerable interest to find all graphs, such as those for vinylbenzene and 2-methyloc-

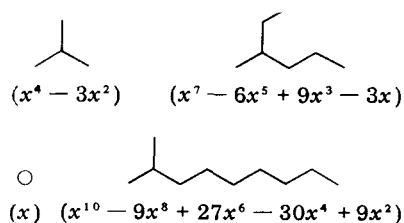
Table IV. The self-returning walks for acyclic graphs with 12 vertices in which two nonequivalent vertices 4 and 11 have an identical list of walk counts.

Atom	Self-returning walks
1	1, 3, 10, 35, 126,
2	3, 10, 35, 126, 463,
3	2, 7, 26, 99, 384,
4	2, 6, 22, 88, 365,
5	2, 7, 28, 118, 507,
6	3, 12, 50, 109, 897,
7	2, 7, 27, 109, 451,
8	2, 5, 15, 52, 198,
9	1, 2, 5, 15, 52,
10	1, 3, 10, 35, 126,
11	2, 6, 22, 88, 365,
12	1, 4, 6, 22, 88,

tane, having nonequivalent sites with identical walk counts. Herndon and Ellzey¹⁴ have constructed several such polycyclic systems. A systematic search for such structures may be posed. Few acyclic graphs on 11 points are known among isospectral graphs.¹⁴ An example of a graph on 12 points is shown below (see Table IV):



Here atoms 4 and 11 have the same walk counts. When points 4 and 11 are erased (one at a time) we obtain fragments, the characteristic polynomials of which produce a same polynomial*:



The major conclusion to be made here is the existence of isocodal vertices in graphs, hence the walk-based characterization of atomic environ-

* For a list of the coefficients of the characteristic polynomials of acyclic graphs with $n = 10$ and fewer vertices see ref. 20.

Table V. The self-returning walks for two sets of isospectral pairs of graphs. In the second pair several vertices of the two graphs show the same count of walks.

Graph	Atoms	Self-returning walks
I	1	0, 1, 0, 5, 4, 29,
	2-5	0, 2, 2, 9, 18, 58,
	6	0, 5, 4, 29, 44, 185,
	Spectral moments:	0, 14, 12, 70, 120, 446,
II	1,2	0, 1, 0, 3, 2, 15,
	3,4	0, 3, 4, 17, 36, 115,
	5,6	0, 3, 2, 15, 22, 93,
	Spectral moments:	0, 14, 12, 70, 120, 446,
III	1,2	0, 2, 0, 8, 4, 38, 40,
	3,4	0, 2, 2, 8, 14, 42, 88,
	5,6	0, 2, 0, 6, 2, 22, 20,
	7	0, 4, 2, 20, 20, 106, 152,
	Spectral moments:	0, 16, 6, 64, 60, 310, 448,
IV	1	0, 1, 0, 3, 0, 11, 2,
	2,3	0, 2, 0, 8, 4, 38, 40,
	4	0, 2, 2, 8, 14, 42, 88,
	5,6	0, 3, 2, 13, 18, 67, 126,
	7	0, 3, 0, 11, 2, 47, 26,
	Spectral moments:	0, 16, 6, 64, 60, 310, 448,

ments is not unique. The counterexample found, however, shows that the appearance of identical walk codes for nonequivalent atoms in the same structure is associated with special structural features, the potential to produce isospectral graphs by allowing unlimited substitution at such points.

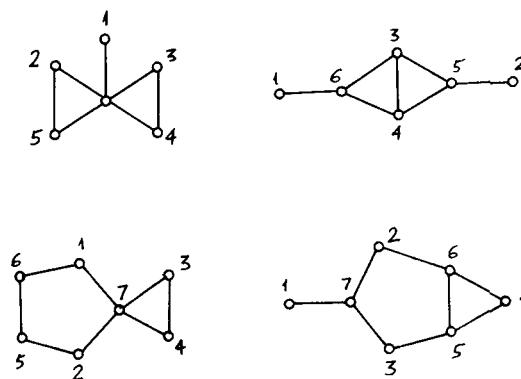
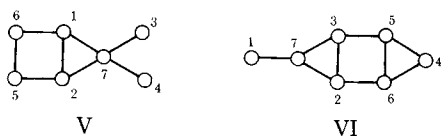


Figure 5. Isospectral graphs without (the first pair) and with (the second pair) common walk-based atomic codes.

ISOCODAL POINTS IN DIFFERENT GRAPHS

Isocodal points may also appear among vertices of isospectral graphs. Consider the graphs I and II of Figure 5. The self-returning walks for their vertices are given in Table V. By summing the atomic sequences, term by term, for each graph separately we derive identical corresponding molecular moments confirming that indeed I and II are isospectral. No instance of vertices of the isospectral pair having the same code is found. Hence, the isocodal vertices of isospectral graphs do not necessarily occur. Because no coincidences in walk counts have been found graphs I and II cannot generate other isospectral graphs similar to that of vinylbenzene.

In contrast isospectral graphs III and IV (Fig. 5) have several sites with identical walk counts, as one can see by inspecting Table V which lists all walk-based atomic codes in the two structures. Any substituent, for instance, placed at site 1 or 2 of graph III or at site 2 or 3 of graph IV will necessarily produce isospectral graphs. Examples of such graphs have been given in the literature.¹⁵ Observe an important difference between the above situation and that of vinylbenzene. Here the isocodal vertices are in *different* graphs. The codes clearly cannot characterize the atom environment, uniquely, but occurrence of identical codes appears to be indicative of a rather uncommon situation. For example, the removal of isocodal points from such graphs leads to isomorphically derived subgraphs. Similarly one finds that by insertion of a bond between the equivalent isocodal vertices in *both* graphs one does not change the situation.



New graphs have different new codes, but again vertices 1,2 in V and 2,3 in VI are isocodal (as are also the vertices 3,4 of V with the vertex 4 in VI).

CONCLUDING REMARKS

Atom codes based on a count of self-returning walks appear promising for a development of useful characterization of atomic environments. The discriminatory power of walk-based codes surpasses other available alternatives even though the codes are not unique. A closer look at occasional coincidences shows them to be the root for occurrence of simply related family of isospectral graphs. Hence, the walk count (i.e., algebraic consideration of A^k matrices) can diagnose rather uncommon structural features and could prove to be the basis for classification of isospectral environments. In particular, one can differentiate isolated cases of isospectral graphs, cases with substitution sites for unrestricted insertion of fragments, cases with multiple number of such vertices, cases with several inactive sites that do not affect the isospectral character of the graph, and finally cases of isospectral pairs, rather than a single skeleton, which can generate new families of isospectral graphs. Such classification focuses attention on parent structures and eliminates a need for cataloging numerous derivatives, once the isocodal points have been identified. On the other hand, one can compensate for the occurrence of nonunique codes by inspecting the codes for the neighboring atoms in order to distinguish constitutionally equivalent sites from isospectral points.

It remains to be seen how the codes based on self-returning walks, and even random walks, will be found useful in discussions of molecular properties and in searches for regularities in available data. In this respect, path-based codes have already been found very valuable. However, there are some indications of limitations of their application. For example, a review of ^{13}C chemical shifts for alkanes shows that path codes (which in acyclic structures are equivalent to enumeration of neighbors at increasing radius of separation) have a limited discrimination power. Carbon atoms with path code 2,2,1 are found to have chemical shifts of about 32.0 and 29.6 ppm²¹: 32.0, 32.1, 32.0, and 32.0 and 29.5, 29.7, 29.5, and 29.7. Inclusion of additional neighbors does not change the situation,

the difference of about 2.5 ppm is caused by differences in the environments at shorter radii. The walk-based codes, however, "recognize" the different environments and at least provide different descriptors.

Finally, we should mention that the concepts of self-avoiding walks (paths), self-returning walks, and random walks have over the past decade been found to be of interest in such diverse fields as lattice theory, statistical mechanics of macromolecules, the Ising problem, residual entropy of ice, the dimer problem, enumeration of conformations, etc.²² Lattices and log chains differ from molecular graphs only in size, hence it is plausible to expect that techniques already found useful in physics, and even biology, are likely to be useful in chemistry also!

This work was supported in part by a grant to Ames Laboratory from the U.S. Department of Energy, Division of Basic Energy Sciences, and was made possible by continuous understanding of Professor R. S. Hansen. Dr. A. Graovac and Professor W. C. Herndon examined the manuscript and suggested several improvements in the presentation of the material. Professor Fink was kind to engage in some algebraic considerations relating to the occurrence of apparently identical codes and made his results available prior to completion of his work.

References

1. R. C. Read and D. G. Corneil, *J. Graph Theory*, **1**, 339 (1977); C. J. Colburn, *Techn. Rep. No. 123/78*, Department of Computer Science, University of Toronto, 1978; B. Weisfeiler, "On Construction and Identification of Graphs," *Lecture Notes in Mathematics*, Vol. 558, Springer-Verlag, Berlin, 1976; G. Gati, *J. Graph Theory*, **3**, 95 (1979).
2. M. Randić, *J. Chem. Phys.*, **60**, 3920 (1974); *J. Chem. Inf. Comput. Sci.*, **17**, 171 (1977).
3. M. Randić, *Chem. Phys. Lett.*, **42**, 283 (1976); *Croat. Chem. Acta*, **49**, 643 (1977); *Int. J. Quantum Chem.*, **15**, 663 (1979); *Int. J. Quantum Chem. Quantum Biol. Symp.* (1980), to be published.
4. M. Randić, *Acta Crystallogr. Sect. A*, **34**, 275 (1978).
5. M. Randić, *J. Am. Chem. Soc.*, **97**, 6609 (1975).
6. M. Randić, *J. Chem. Inf. Comput. Sci.*, **18**, 101 (1978); M. Randić and C. L. Wilkins, *J. Chem. Inf. Comput. Sci.*, **19**, 31 (1979); M. Randić, G. M. Brissey, R. B. Spencer, and C. L. Wilkins, *Comput. Chem.*, **3**, 5 (1979); M. Randić, *MATCH Inf. Commun. Math. Chem.*, No. 7, 5-64 (1980); M. Randić and C. L. Wilkins, *J. Chem. Inf. Comput. Sci.*, **20**, 36 (1980).
7. M. Randić and C. L. Wilkins, *Chem. Phys. Lett.*, **58**, 181 (1978); M. Randić, *J. Chromatogr.*, **161**, 1 (1978); M. Randić, *Int. J. Quantum Chem. Quantum Biol. Symp.*, **5**, 245 (1978); M. Randić and C. L. Wilkins, *Chem. Phys. Lett.*, **63**, 332 (1979); *J. Phys. Chem.*, **83**, 1525 (1979); M. Randić, *Chem. Phys. Lett.*, **38**, 68 (1976); *Tetrahedron*, **33**, 1905 (1977); *J. Am. Chem. Soc.*, **99**, 444 (1977); *Int. J. Quantum Chem.*, **17**, 549 (1980).
8. R. J. Wilson, *Introduction to Graph Theory*, Academic, New York, 1972; F. Harary, *Graph Theory*, Addison-Wesley, Reading, MA, 1969; R. G. Bursacker and T. L. Saaty, *Finite Graphs and Networks*, McGraw-Hill, New York, 1965; N. L. Biggs, E. K. Lloyd, and R. F. Wilson, *Graph Theory, 1736-1936*, Clarendon, Oxford, 1976.
9. W. Brostow and A. Schnitzel, *J. Stat. Phys.*, **4**, 103 (1972).
10. C. A. Coulson, *Proc. Cambridge Philos. Soc.*, **46**, 202 (1949).
11. N. Trinajstić, *Croat. Chem. Acta*, **49**, 593 (1977).
12. A. Graovac, I. Gutman, N. Trinajstić, and T. Živković, *Theor. Chim. Acta*, **26**, 67 (1972).
13. T. Živković, N. Trinajstić, and M. Randić, *Mol. Phys.*, **30**, 517 (1975) (reported at the Quantum Chemistry School in Leningrad, December 1973); W. C. Herndon, *Tetrahedron Lett.*, 671 (1974).
14. L. A. Collatz and U. Sinogowitz, *Abh. Math. Sem. Univ. Hamburg*, **21**, 63 (1957); M. E. Fisher, *J. Comb. Theory*, **1**, 105 (1966); A. T. Balaban and F. Harary, *J. Chem. Doc.*, **11**, 258 (1971); F. Harary, C. King, A. Mowshowitz, and R. C. Read, *Bull. London Math. Soc.*, **3**, 321 (1971); F. Harary, *SIAM Rev.*, **4**, 202 (1962); W. C. Herndon and M. E. Ellzey, Jr., *Tetrahedron*, **31**, 99 (1975); M. Randić, N. Trinajstić, and T. Živković, *J. Chem. Soc. Faraday Trans. 2*, **72**, 244 (1976).
15. R. A. Marcus, *J. Chem. Phys.*, **43**, 2643 (1976); M. F. Sykes and M. E. Fisher, *Adv. Phys.*, **9**, 315 (1960).
16. G. A. Baker, Jr., *J. Math. Phys.*, **7**, 2238 (1966).
17. O. E. Polansky and I. Gutman, *MATCH Inf. Commun. Math. Chem.*, No. 5, 219 (1979).
18. A. Graovac and M. Randić, unpublished.
19. A. J. Schwenk, in *New Directions in the Theory of Graphs*, F. Harary, Ed., Academic, New York, 1973.
20. A. Mowshowitz, *J. Comb. Theory*, **12**, 177 (1972).
21. M. Randić, *Int. J. Quantum Chem.* (submitted).
22. M. E. Fisher, *J. Math. Phys.*, **7**, 1776 (1966); P. W. Kasteleyn, *J. Math. Phys.*, **4**, 287 (1963); S. Sherman, *J. Math. Phys.*, **1**, 202 (1960); G. S. Rushbrooke, *J. Math. Phys.*, **5**, 1106 (1964); M. E. Fisher and J. W. Essam, *J. Math. Phys.*, **2**, 609 (1961); J. W. Essam and M. F. Sykes, *J. Math. Phys.*, **7**, 1573 (1966); J. F. Nagle, *J. Math. Phys.*, **7**, 1484 (1966).