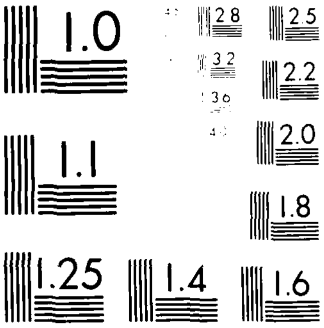
AD-A170 107 THE CHALLENGE OF CHARACTERIZING BRANCHING IN MOLECULAR SPECIES(U) GEORGIA UNIV ATHENS DEPT OF CHEMISTRY D R ROUVRAY 16 JUL 86 TR-24 N00014-84-K-0365 1/1

UNCLASSIFIED F/G 7/4 NL



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

OFFICE OF NAVAL RESEARCH

Contract N00014-84-K-0365

TECHNICAL REPORT NO. 24

The Challenge of Characterizing Branching in Molecular Species

by

D.H. Rouvray

Prepared for publication in

Discrete Applied Mathematics

University of Georgia
Department of Chemistry
Athens, Georgia 30602

July 16, 1986

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER Technical Report No. 24	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* THE CHALLENGE OF CHARACTERIZING BRANCHING IN MOLECULAR SPECIES		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) D.H. Rouvray		8. CONTRACT OR GRANT NUMBER(s) N00014-84-K-0365
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Georgia Department of Chemistry Athens, GA 30602		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR051-861
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Department of the Navy Arlington, VA 22217		12. REPORT DATE 7/16/86
		13. NUMBER OF PAGES 35
14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)*
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

To be published in Discrete Applied Mathematics

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Molecular Branching, Graph Theory

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

A precise definition of the concept of branching as applied to molecular species has been an objective of chemists for several decades and still remains a challenge today. Although the problem has been tackled in many ingenious ways, no completely satisfactory resolution can be said to have been achieved. The reason for this is that there are two fundamental difficulties confronting the researcher in this field. These are that (i) a universally agreed definition of branching appears to be unattainable because ultimately the concept can be defined only in intuitive

(continued)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

#20

terms; and (ii) any mathematical measure used to characterize the branching in molecular species must also accurately reflect the physicochemical properties of those species. It also has to be borne in mind that different properties are known to be associated with different orderings. In this first comprehensive survey of the field, the current state of the art in characterizing molecular branching is reviewed and the prospects for future advances are assessed.

The Challenge of Characterizing Branching in Molecular Species

D.H. Rouvray

Chemistry Department, University of Georgia, Athens, Georgia 30602, U.S.A.

Abstract

A precise definition of the concept of branching as applied to molecular species has been an objective of chemists for several decades and still remains a challenge today. Although the problem has been tackled in many ingenious ways, no completely satisfactory resolution can be said to have been achieved. The reason for this is that there are two fundamental difficulties confronting the researcher in this field. These are that (i) a universally agreed definition of branching appears to be unattainable because ultimately the concept can be defined only in intuitive terms; and (ii) any mathematical measure used to characterize the branching in molecular species must also accurately reflect the physicochemical properties of those species. It also has to be borne in mind that different properties are known to be associated with different orderings. In this first comprehensive survey of the field, the current state of the art in characterizing molecular branching is reviewed and the prospects for future advances are assessed.

In physical science ... the most important and most fruitful concepts are those to which it is impossible to attach a well-defined meaning.

H.A. Kramers (1947)

Introduction

In chemistry. frequent use is made of a number of concepts which, in a strictly mathematical sense, are ill-defined. Examples include the concepts of aromaticity, complexity, shape and structure, all of which have been widely used to describe molecular species, yet none of which has been precisely defined. Although this lack of precision on the part of chemists appears not to have seriously impeded the progress of chemistry to date, there are signs that precise definitions of several commonly employed concepts could make an important contribution to the future development of the subject. Accordingly, we shall focus here on one such concept, namely the concept of branching in molecular species, and explore the ways in which it has been approached by both chemists and mathematicians. Graph-theoretical ideas would appear to be highly relevant in this context, for the problem has already been tackled by several workers in the mathematical literature [17,32,83,90]. In the molecular graphs used by chemists to represent chemical species, branching has traditionally been considered to occur whenever the graphs contained at least one vertex having a valence greater than two. Moreover, the higher the valence of the vertices, the greater the extent of branching in the species was said to be. This notion has been formalized in terms of the valence partitioning of the vertices of molecular graphs. Nonisomorphic graphs having identical extents of branching were described [78] as differing in their 'branching content'. Before pursuing such ideas further here, we now pause to introduce some necessary chemical terminology.

For well over a century it has been known [26] that two chemical compounds which have the same chemical formula may differ in the internal arrangement of their atoms. Two such compounds are referred to as chemical isomers; isomers always differ from one another in at least one of their physicochemical properties. Overall, isomers have been classified into two broad categories designated as constitutional isomers and stereoisomers. Several schemes for the detailed classification of isomers have been developed in recent years [8,15,83], and at least 30 different subclasses of isomers are now recognized by chemists [72]. Computer programs for the enumeration of most of these subclasses are also available [43]. Our interest here will focus only on the former category of isomers, i.e. the constitutional isomers, which have also been widely referred to in the past as structural isomers [35]. We elect not to use this latter term, however, since, as mentioned above, the term structure is ill-defined in the chemical context, and in certain of its various meanings the adjective 'structural' has therefore become somewhat ambiguous. Constitutional isomers may be regarded as discrete molecular entities whose atoms are bonded together and held at approximately fixed positions in space relative to one another as a result of the constraints imposed upon their mutual motions by the bonding interactions [51]. A pair of constitutional isomers must differ in both the sequence and the nature of the bonding interactions occurring between their respective atoms [43].

In the mid-1850s Cayley [16] first depicted the constitutional isomers of the members of certain homologous series, namely the alkanes, $C_{\underline{n}}H_{2\underline{n}+2}$, and monosubstituted alkanes, $C_{\underline{n}}H_{2\underline{n}+1}X$, using graphs. He clearly established that a pair of constitutional isomers will always be possessed of two nonisomorphic graphs and that there is a 1:1 correspondence between the alkane isomers having $\underline{n}$ atoms and the relevant tree graphs on $\underline{n}$ vertices. The relevant tree graphs in this case are allowed to have a maximum vertex degree of four. Cayley also enumerated the isomers for the first several members of each series; later workers have subsequently corrected (where necessary) and substantially extended these early results [73,43]. We shall consider only alkane species here, for these species conveniently exemplify the nature of the problems we propose to discuss in this paper. The numbers of isomers for several different members of the alkane series are presented in Table 1. Given that the valence of the carbon atom is four and that of the hydrogen atom is one, it is easy to demonstrate that alkane species contain the maximum ratio of hydrogen to carbon of all the hydrocarbons [75]. For our purposes it will be sufficient to represent the alkanes by their carbon backbones and to ignore the hydrogen atoms, which can usually be inferred without difficulty and which in any case are nonessential in that they are not structure-determining. Graphs depicting only the carbon skeleton of hydrocarbon species are widely used in mathematical chemistry and are referred to as hydrogen-suppressed graphs.

TABLE 1

In this first comprehensive review on the mathematical description of molecular branching, we shall highlight the problem of characterizing in a chemically meaningful way the hydrogen-suppressed graphs of members of the alkane series, C_nH_{2n+2}. Ideally, such characterizations should satisfy two criteria, viz. (i) they should be unique in purely graph-theoretical terms, and (ii) they should accurately reflect the physicochemical properties of the species being characterized. It is fair to point out that it is not possible to satisfy both of these criteria simultaneously at present. Although it is certainly feasible to characterize species uniquely, e.g. by means of their adjacency matrix or by some appropriate code [82], characterizations of this kind are not only unwieldy but, more importantly, they usually fail to provide a sufficiently reliable description of the physicochemical and other properties. On the other hand, all of the simple numerical descriptors of species which have been employed to date have subsequently been shown to be nonunique. The problem chemists are confronted with is thus a challenging one and no completely satisfactory solution appears to be in sight. Over the past decade, however, steady progress has been made and some important new insights have been gained. It is our purpose now to review the current state of the art in the characterization of molecular branching though, in an effort to keep the number of literature citations down to manageable proportions, only key references will be given. We set the scene by first exploring the question whether it is feasible to attempt to characterize branching in purely mathematical terms.

The Measurement of Branching

Virtually all of the physicochemical properties of alkane species are either greatly influenced by or substantially dependent upon the degree of branching present in their constituent molecules. One notable example of such a property, which has very important commercial implications, is the octane rating assigned to fuels used in automobiles and other vehicles. In effect, the octane rating of a fuel determines its quality, for the higher the rating the less likely the fuel will be to self-ignite upon sudden compression in air. The octane rating of an alkane fuel is directly dependent upon the amount of branching present in its component molecules [3]. Even from this isolated example, the crucial importance of the concept of branching to chemistry should be evident. What chemists lack, however, is some effective means of measuring the amount of branching present in molecules based on some universally agreed definition. As indicated above, the notion of branching has traditionally been described in purely intuitive terms [24], such as the number of vertices of degree greater than two in the chemical graph. We discuss now whether it is possible to improve upon this seemingly unsatisfactory method of interpreting a highly important molecular property.

One approach to the problem favored by chemists in recent years has been to attempt to order chemical graphs according to some set of well-defined mathematical criteria. Once such an ordering has been achieved, the second issue of whether the ordering matches any of the orderings based on the various physicochemical properties of molecules can then be addressed. Let us start with a simple Gedankenexperiment. If we consider two tree graphs, one in the form of a path and the other in the form of a star, it is immediately obvious which of the two is more branched. Thus, any scheme we may devise to order branched molecular species must always give precedence to the star graph over the path graph. When comparisons of certain other pairs of tree graphs are made, however, intuition is no longer sufficient. For instance, it is by no means obvious which of the three graphs illustrated in Figure 1 is more branched, even though the molecules they represent can certainly be ordered hierarchically in terms of their physicochemical properties. Numerous other equally indeterminate examples might be cited. We now explore the contribution which ordering can make to the solution of problems of this type, bearing in mind that most ordering procedures merely define a hierarchy but do not assign absolute values to the degree of branching present in molecular species.

FIGURE 1

The Ordering of Structures

Any ordering of structures necessarily implies that comparisons have to be made. In the chemical context, the comparisons are frequently made between sequences of numbers which are used to identify the structures they represent. The numbers chosen might be integers; one convenient way of obtaining these is to take the vertex degrees of the hydrogen-suppressed graphs arranged as a nonascending sequence, i.e. $v_i \geq v_{i+1}$ for all i = 1,2,..., (n - 1). Two sequences of numbers of the same length are said to be comparable if there exists an inequality between them for all intervals defined by the values of the variables. Comparability can be tested for by constructing sequences of partial sums. To illustrate this, let us suppose that the two sequences are $V = \{v_i\}$ and $V' = \{v_i'\}$. Now, for all the v_i and v_i' these sequences will be comparable only if $V \geq V'$ or $V \leq V'$ for all the intervals. Muirhead [14,15] defined a relative ordering for such sequences by imposing the conditions:

$$\sum_{i=1}^{k} v_i \geq \sum_{i=1}^{k} v_i' \text{ , where } 1 \leq k \leq n \qquad (1)$$

and

$$\sum_{i=1}^{n} v_i = \sum_{i=1}^{n} v_i'. \qquad (2)$$

Whenever these conditions are satisfied, sequence V is said to precede sequence V'.

Such criteria were first introduced into the chemical literature by Gutman and Randić [30], who applied them to the ordering of alkane isomers. They were able to show that a complete ordering is possible for all such isomers having $n \leq 7$, whereas for $n \geq 8$ only a partial ordering can be achieved. It is thus not legitimate to compare certain pairs of isomeric structures having $n \geq 8$ since Muirhead's conditions [50] are not fulfilled in all cases. The three structures illustrated in Figure 1 are not comparable, for instance, since for all i we have $v_i = v_i'$, that is to say the sequence to be compared equals 32222111 for each of these isomers. Later refinements of these conditions have not brought any significant improvement. Thus, the generalized conditions of Karamata [40,7], which removed the restriction that only integers be used in the sequences, certainly made the comparison of sequences of real, nonintegral numbers possible, though the drawback of having a number of noncomparable pairs of structures in the set still remained. Randić [59] indicated how this difficulty might be alleviated to some extent by the use of additional information in the form of several new partial sums derived for the sequences. This expedient, however, did not satisfactorily resolve the problem.

From a different vantage point, an equivalent approach to that of Randić [59] has emerged in recent years. In a fundamental study of the phenomena of chirality in molecules, Ruch [79,77] made use of Young diagrams [89], which were subsequently shown to have relevance not only in the interpretation of chirality but in several other areas as well, including the study of molecular branching [80,34]. When used for this latter purpose, Young diagrams are constructed by ordering the vertex degrees of graphs in a nonascending sequence as described above. The graphs are then depicted by arrays of square boxes in which each of the rows represents a single vertex and the number of boxes in a given row is determined by the degree of the relevant vertex. The Young diagrams for the three isomers shown in Figure 1 will all be based on the vertex sequence 32222111 and are thus all identical, as is apparent from Figure 2. The fact that these three isomers correspond to the same diagram makes the limitations of the approach manifest. Clearly, only a partial ordering will be possible by this means, for the ordering which results is precisely the same as that attained by the use of Muirhead's criteria [78]. Accordingly, there is no special advantage to be gained by adopting this particular approach to ordering; we shall therefore not discuss it further here.

FIGURE 2

A more promising approach to the ordering of graphs was put forward by Randić and Wilkins [68], who used paths of differing lengths as the basis for their procedure rather than vertex degrees. In the tree graphs of alkane species, the enumeration of the various paths present in the graphs is straightforward. In the case of the 18 octane isomers, the result of the enumeration is presented in tabular form in Table 2. For reasons of convenience, Randić and Wilkins [68] ordered these isomers in terms of the pair of numbers (p_2, p_3), representing respectively paths of lengths two and three. Strictly speaking, a septuple rather than a pair should have been used to account for all the paths present, though even their simplistic approach produced a surprisingly good ordering. The various isomers were positioned on a grid according to their (p_2, p_3) values as illustrated in Figure 3. The conditions invoked for the actual ordering were that two structures were comparable only if $p_2 \geq p_2'$ and $p_3 \leq p_3'$; whenever these conditions were satisfied, the points on the grid corresponding to the two structures were connected. The ordering attained by this method is again only a partial one, and two of the 16 points on the grid correspond to structures having identical (p_2, p_3) values. One of the identical pairs is the 3-methylheptane and 4-methylheptane pair, illustrated in Figure 1. It should be pointed out, however, that if all paths had been used in the ordering process, a complete ordering of all the 18 isomers would have been possible since no two isomers have all their path length sequences identical. The approach was later extended to the sets of alkane isomers having $\underline{n}$ = 9 (the 35 nonanes) [69] and $\underline{n}$ = 10 (the 75 decanes) [64] with similar results.

TABLE 2

FIGURE 3

The Encoding of Molecules

All of the methods discussed so far for discriminating among isomers have depended upon the use of numerical codes, namely upon sequences of nonascending vertex degrees or upon sequences of path numbers. In this section we shall briefly examine codes which provide a unique characterization of species. We shall again confine the discussion to alkane molecules, and say little about various other, nonunique codes which have been put forward for species characterization. In fact, we can only touch upon the subject here, for the study of codes covers so vast an area that it deserves a separate review by itself. Now, it is widely recognized [57] that the use of some standard numbering procedure for the vertices of graphs would render the problem of establishing the isomorphism of a pair of graphs an essentially trivial one. Once such a procedure has been devised, each of the graphs may then be represented by a so-called canonical matrix and this permits an ordering of those graphs e.g. by lexicographical ordering of the matrices.

Since the adjacency matrix is known [18] to characterize any graph it represents up to isomorphism, many workers have focused attention on this particular matrix. The adjacency matrix, A(G), which may be defined as follows:

$$A(G) = \begin{cases} a_{\underline{i}\underline{i}} = 0 \\ a_{\underline{i}\underline{j}} = 0 \quad (\underline{i}, \underline{j} \notin e(G) \\ a_{\underline{i}\underline{j}} = 1 \quad (\underline{i}, \underline{j} \in e(G) \ , \end{cases} \qquad (3)$$

where e(G) is the edge set of G, can be written out in the form of a binary number by reading the rows sequentially from left to right and from top to bottom. Standard forms of presenting A(G) have been sought which would yield either the maximum or the minimum binary number using this representation. The problem has been examined from a variety of different standpoints, including those of Nagle [52], who proposed a general linear ordering relation for graphs to derive the canonical matrix; Randić [57,62], who devised canonical labeling schemes for graphs based upon A(G) and who went on to apply these notions to the study of topological symmetry [62,53]; El-Basil and coworkers [23,22], who utilized codes based on the traces of $\underline{A}(G)^{\underline{k}}$, where $1 \underline{<} \underline{k} \underline{<} \underline{n}$, to characterize both cyclic and noncyclic organic molecules; and Herndon and Leonard [36], who extended the concepts of canonical labeling and unique linear notation to organic and inorganic polyhedral cluster compounds.

To illustrate the types of code which can be derived from canonical labeling, we now consider the approach of Randic [57,62] in some detail. Since any graph on $\underline{n}$ vertices will have a total of $\underline{n}!$ possible labelings, the three tree graphs in Figure 1 will have 8! possible labelings. To reduce this large number, some algorithm is necessary to devise a labeling which will yield a binary number of minimum value without screening all the $\underline{n}!$ possibilities. Initially, Randić [62] suggested that the labeling be obtained simply by permuting the rows and columns of A(G) two at a time, starting with a graph having arbitrary labeling. It was later demonstrated by MacKay [45], however, that such a procedure can result in trapping in a local minimum, and is thus not foolproof. A more satisfactory procedure, also developed by Randić [57], involved carrying out operations on A(G) to ensure that its first row would have the maximum number of zeros in it and that these would precede ones whenever possible. In terms of graph labelings, this implies that the smallest label (1) should have as its immediate neighbors vertices bearing the largest labels ($\underline{n}$, $\underline{n}$ - 1, etc.). After treating the first row of $\underline{A}$(G) in this way, the second and subsequent rows are then dealt with in the same manner. In general, this can be accomplished without difficulty, for the procedure is a very efficient one [57]. Examples of the canonical labelings and resulting codes for the three isomers of pentane (C_5H_{12}) are depicted in Figure 4.

FIGURE 4

We conclude this section by making brief mention of a newly developed unique code, known as a <u>compact</u> code. The evolution of this type of code can be traced back at least two decades. Hiz [39] introduced the idea of linearizing chemical graphs in the form of codes called ciphers, which omitted all extraneous chemical information pertaining to the species represented. Knop et al. [42] developed those ciphers for the purpose of enumerating the classes of molecules originally studied by Cayley [16], namely the alkanes, C_nH_{2n+2}, and the substituted alkanes, $C_nH_{2n+1}X$. Recently, Randić [55] demonstrated how these ciphers could be adapted to the labeling of various molecules having tree graphs. The compact code is constructed by locating the vertex (vertices) of highest degree (degrees) and then writing nonascending vertex degree sequences for all the paths emanating from such vertices. The various sequences are concatenated into one code according to their lengths, with the longest being written down first. For the three alkane isomers in Figure 1, the codes now differ and assume the forms 32222111, 32221211, and 32212211. Not only is the code useful for ordering species, but direct reconstruction of the chemical species represented is also possible, for a 1 can be interpreted as a primary carbon atom or methyl group (CH_3); a 2 as a secondary carbon atom or methylene group (CH_2); a 3 as a tertiary carbon atom or a methyne group (CH); and a 4 as a quarternary carbon atom (C).

The Use of Polynomials and Eigenvalues

An important graph invariant now being increasingly used in the characterization of molecular branching is the characteristic polynomial, $P_G(x)$, which is defined as $(-1)^n$ det $|A(G) - xE(G)|$, where E(G) is the unit matrix for the graph G. Various methods for the evaluation of $P_G(x)$ have recently been discussed by Randić [61]. Although this polynomial has long been known not to provide a unique characterization of graphs [17], it has remained of interest to chemists because the coefficients of $P_G(x)$ may be obtained from certain combinations of subgraphs comprised of disjoint edges or cycles [80]. These subgraphs are clearly related to the numbers of random and self-returning walks in G, and also to the nonadjacent number and cycle counts. This fact led Randić [56] to explore the idea of representing $P_G(x)$ in terms of summations of the polynomials of paths on n vertices, $L_n(x)$, as defined in equation (4). In the case of the three isomers in Figure 1, the $P_G(x)$ assume the forms (a) $L_9 - L_5$; (b) $L_9 - L_5 - L_3$; and (c) $L_9 - L_5 - L_3 - L_1$. The coefficient of L_5 was found to reflect the number and type of substitutions occurring on the main chain: for a methyl (CH_3) substitution it takes the value -1; for methyl substitutions at two different atoms -2; for dimethyl substitution on the same atom -3; for disubstitution on one atom and monosubstitution on another atom -4; for tetramethyl substitution -5; and for one tetramethyl substitution and two other single substitutions -6.

An explicit closed form for $L_n(x)$ polynomials was originally presented by Collatz and Sinogowitz [17] as follows:

$$L_n(x) = \sum_{k=0}^{[n/2]} (-1)^k \binom{n-k}{k} x^{n-2k} . \tag{4}$$

From this expression, it is evident that the $L_n(x)$ may be written as Chebyschev polynomials in $x/2$. Several prescriptions for discerning the general form assumed by the $L_n(x)$ for various families of structures based on the graphs of alkane isomers were put forward by Randić [56]. These prescriptions were generalized by Hosoya and Randić [37], who derived a number of closed expressions and who pointed out, for instance, that x^n can be formulated as:

$$x^n = \sum_{k=0}^{[n/2]} \frac{n-2k+1}{n+1} \binom{n+1}{k} L_{n-2k} . \tag{5}$$

The regrouping of the terms in these expansions of $P_G(x)$ renders the patterns for the individual coefficients obvious in many cases. Thus, by focusing attention on families of structurally related graphs, it is possible to utilize $P_G(x)$ for the purpose of characterizing such graphs. The expansion based on Chebyshev polynomials can also be used in those cases in which $P_G(x)$ exhibits sets of identical spectra for pairs of nonisomorphic graphs.

Isospectral graphs have been investigated for many years and numerous references can be cited in the mathematical [17,32,83,90] and chemical [61,91,65] literature. Such graphs have received attention from chemists principally because the eigenvalues of chemical graphs correspond to the quantum-mechanically allowed energy levels within the species represented [46]. Our interest here, however, stems from the observation of several authors [17,44,31,11] that the degree of branching in a graph is closely related to its maximum eigenvalue, λ_1, frequently referred to as the <u>spectral radius</u>. Cvetković and Gutman [19] were the first to demonstrate that λ_1 may be expressed in terms of the total number, $w(k)$, of walks of length k in a graph G by means of the following approximate formula:

$$w(k) \approx n\,(\lambda_1)^k \, . \tag{6}$$

The approximation becomes an equality only in the case of regular graphs.

The result in equation (6) represents an interesting relationship between a spectral property (λ_1) and a combinatorial property ($w(k)$) of a graph, and thereby confirms the empirical finding that λ_1 provides a reliable measure of branching in molecular graphs. Moreover, since λ_1 satisfies the inequality:

$$v_{min} \leq \lambda_1 \leq v_{max} \, , \tag{7}$$

it may be interpreted [19] as a kind of mean vertex degree for the graph G. Lovász and Pelikán [44] were able to prove that if all trees on _n_ vertices are ordered according to their λ_1 value, the path will occupy the first position (minimal λ_1) whereas the star will occupy the final position (maximal λ_1) in the sequence. The three molecules illustrated in our Figure 1 are differentiated in terms of their respective λ_1 values, having the values (a) 1.950, (b) 1.989, and (c) 2.000. It should be reiterated, however, that neither λ_1 nor the complete set of eigenvalues $\{\lambda_{\underline{n}}\}$ offers a unique characterization of G. Thus, the graphs of 3-ethylpentane and 2,4-dimethylpentane have identical λ_1 values, and the two graphs illustrated at the top of Figure 5 have identical sets of eigenvalues, $\{\lambda_{\underline{n}}\}$.

FIGURE 5

Graph Invariants as Branching Descriptors

In the discussion of graph invariants many of the diverse lines of thought introduced above find their natural intersection. We are concerned here only with those invariants which have been used specifically for the correlation of molecular structures with physicochemical properties. Although a wide variety of invariants has been employed for this purpose over the past four decades, it is only during the last few years that their great importance to chemistry has been fully appreciated [54]. Nowadays, graph invariants are usually referred to in the chemical literature as _topological indices_; for convenience, we shall refer to them here simply as _indices_. In recent years a steady stream of indices has emerged, allegedly providing an increasingly reliable characterization of molecular branching. We shall focus especially on the newer indices and the claims made for them, for it is neither feasible nor appropriate here to review comprehensively the vast field of topological indices; interested readers are referred to several detailed reviews on the subject [11,49,74,2].

The first graph invariant to be used in chemistry was introduced in 1947 by Wiener [88], and is commonly referred to nowadays as the Wiener index, W(G). The index was originally defined as the sum of the chemical bonds existing between all pairs of carbon atoms in a molecule, and later shown [38] to be equal to one half the sum of the entries in the relevant distance matrix, i.e.:

$$W(G) = \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}(G) . \qquad (8)$$

W(G) has been widely used to model the physicochemical properties of chemical species, such as boiling point and refractive index [76]. Although the index gives good correlations for species having unbranched graphs, when branched species are included the results are not nearly as satisfactory. This is well illustrated by the plot in Figure 6, which reveals the wide scatter in the points for the 75 decanes ($C_{10}H_{22}$) when the boiling point is plotted against W(G). In fact, the correlation coefficient for linear regression turns out to be only 0.0035! Moreover, W(G) is associated with a fairly high level of degeneracy; a pair of trees having identical W(G) values are shown in Figure 5.

FIGURE 6

Since the time of Wiener, strenuous endeavors have been made to devise better indices than W(G). The first major advance came in 1971 when Hosoya [38] introduced an index of the form:

$$Z(G) = \sum_{k=0}^{[n/2]} p(G,k) , \qquad (9)$$

where $p(G,k)$ is the number of ways in which k edges can be chosen from G such that no two of them are adjacent; by definition $p(G,0) = 1$ and $p(G,1) = n_e$, the number of edges in G. For trees, the characteristic polynomial and Z(G) are interrelated, and this polynomial can be expressed in terms of the $P(G, k)$ as follows:

$$P_{G=T}(x) = \sum_{k=0}^{[n/2]} (-1)^k \, p(G, k) \, x^{n-2k} \, . \qquad (10)$$

Gutman [29] has shown that Z(G) is particularly well suited to reflect the alternations in boiling point in monomethyl alkanes as the methyl group is displaced along the main carbon chain. The index suffers from the drawback, however, that it too displays a high level of degeneracy i.e. it is far from being one-to-one, for the classes of graphs of interest here.

The first index specifically designed to be of low degeneracy was the molecular connectivity index of Randić [58]. This has proved to be a highly successful index in that it is the most widely used of all indices propounded so far; moreover it is the only index to have had a whole book devoted to it [41]. The index was designed with the intention of characterizing branching in chemical species and is based on the notion of edge types in molecular graphs. An edge is said to be of type (v_1, v_2) if the two end vertices of the edge have degrees v_1 and v_2 respectively. In formal terms, the index may be defined by the relationship:

$$x = \sum_{e=1}^{n_e} (v_i v_j)^{-\frac{1}{2}} \, , \qquad (11)$$

where the summation extends over all e edges, and n_e is the total number of edges in G. To date, the index has been used in a vast number of correlations ranging from the prediction of physicochemical properties to the design of drugs

[41]. Its degeneracy is moderately low [72].

Because of the low degeneracy of $\chi(G)$ and its great value in correlational studies, several attempts have been made to extend its range of usefulness. The first proposal put forward [41] envisaged summing over paths of different lengths instead of choosing paths of length one. This idea led to the introduction of a whole range of $\chi(G)$ indices designated as $^{0}\chi(G)$, $^{1}\chi(G)$, $^{2}\chi(G)$, $^{3}\chi(G)$, etc. for paths of length zero, one, two, three, etc. The index defined above should thus more correctly be referred to as $^{1}\chi(G)$. The generalized index, $^{h}\chi(G)$ may be defined by the equation:

$$^{h}\chi(G) = \sum_{\pi} [v_1(\pi) v_2(\pi) \ \ldots \ v_{h+1}(\pi)]^{-\frac{1}{2}} \ , \qquad (12)$$

where π extends over all paths of length h and $v_i(\pi)$ denotes the valence of the *i*th vertex on path π, with $1 \le i \le h+1$. The $^{h}\chi(G)$ index is, of course, also derivable from the *h*th power of A(G). Each of the $^{h}\chi(G)$ will give a different weighting for the contributions made by primary (CH_3), secondary (CH_2), tertiary (CH), and quarternary (C) carbon atoms. The basic purpose of such indices is to give prominence to the contributions from adjacent and nearby atoms (vertices) and to deemphasize those which are further away, in accord with chemical intuition. The $^{h}\chi(G)$ indices have moderately low degeneracies, and $^{1}\chi(G)$ correlates highly (0.98) with W(G) [58].

Further means of elaborating such indices have also been examined in the chemical literature. Thus, Balaban [68] put forward an index known as the distance sum connectivity index, J(G), which is defined as follows:

$$J(G) = \frac{n_e}{n_e - n - 2} \sum (\delta_i \ \delta_j)^{-\frac{1}{2}} \qquad (13)$$

where n_e is the number of edges in G, and δ_i represents the sum of the entries in the ith row of the distance matrix D(G) for G. The degeneracy of J(G) has been shown to be very low; in alkane graphs the first degenerate pair encountered has n = 12 vertices [5]. Following several earlier studies on the characterization of graph vertices in terms of their path numbers [54,58,9], Randić [60] proposed combining the $\chi(G)$ with path numbers. This resulted in an index having a very low degeneracy known as the molecular identity number, MID. The first pair of alkane trees with identical MID numbers has 16 vertices [63].

Approaching the Ultimate Goal

The success in developing ever more discriminating indices with lower and lower degeneracies, has prompted several researchers to pose the question: can a simple, graph-theoretical, numerical descriptor be derived which will be unique, at least for the classes of graphs of interest to chemists? Although much progress has been made on the difficult task of characterizing alkane trees uniquely by means of such an index, this ultimate goal seems to be a very elusive one. Numerous conjectures put forward over the years postulating that certain indices -- including the Randić MID number -- were unique have subsequently been proved to be invalid [85,87]. In spite of this, new conjectures continue to be made. For instance, it has recently been conjectured [1] that if distance sums and path numbers were used in the MID number instead of vertex degrees and path numbers, the degeneracy would vanish. The search for unique indices will almost certainly be continued for many years to come. Below we touch upon some of the more novel approaches which have been explored recently and which are claimed to lead to highly discriminating, if not unique, descriptors for alkane tree graphs.

The use of random walks on trees has been investigated by Randić et al. [70] and Barysz and Trinajstić [6]. The former workers used random walks to characterize graphs by enumerating all the walks for every individual vertex. Attempts were then made to decide which factors were critical in determining the walk counts, and to locate isospectral vertices in graphs. Unusual walks, i.e. walks for nonequivalent sites which have the same counts, are of fundamental importance in the study of isospectral graphs. These facts were exploited by the latter workers to establish a 1-1 correspondence between trees and a code called the ordered structural code [6]. This code distinguishes even isospectral graphs. The code, which is claimed to be unique, can be used for calculating the coefficients of the characteristic polynomials of trees and for demonstrating the dependence of the spectral moments on the various tree structures. Spectral moments are obtained by summing the diagonal elements of $(A(G))^k$ for each k, and correspond directly to the count of all self-returning walks of length k in a given molecular graph.

Information theory has played a role in the development of new topological indices for many years. Recently, a book devoted solely to this subject has appeared [10]. One of the most successful information-theoretical index in terms of its discriminating power is the so-called mean information on distance equality index, defined as follows [13]:

$$I_D^E(G) = \sum_{\ell=1}^{m} \frac{2k_\ell}{n(n-1)} \log_2 \left[\frac{2k_\ell}{n(n-1)}\right] , \qquad (14)$$

where the distance ℓ appears $2k_\ell$ times in the distance matrix $D(G)$ for the graph G and m is the greatest value of ℓ. Another very successful index is the so-called graph distance complexity index advanced by Raychaudhury et al. [71], which is based on an average information measure for G. Even though both display high discriminatory power for alkane trees, it was postulated by Bonchev et al. [13] that an effective practical solution to the problem of discrimination would be the introduction for a superindex. Such an index is simply a sum of several separate topological indices. Using a superindex based on six topological indices, Bonchev et al. [13] achieved complete discrimination of a set of 427 graphs of chemical interest.

An index which is of extremely low degeneracy, and which has the advantage of being easily obtainable, is based on the hierarchically ordered extended vertex connectivities in G. The algorithm used to calculate the index, commonly referred to as the HOC algorithm [47], starts with a partitioning of the vertices of G into equivalence classes according to their degrees. Additional discrimination is built into each of the classes by means of vertex extended connectivities, i.e. sums of vertex degrees of the nearest neighbors. For equal extended connectivities, further discrimination is introduced via the sequences of the degrees, arranged in ascending order. The newly formed equivalence classes are assigned ranks that increase with the extended connectivities and their ordered summations. These ranks are then used for the iterative recalculation of the extended connectivities; the whole procedure is terminated when the same ranks appear after two consecutive steps. The approach represents a natural extention of Morgan's [48] algorithm which forms the basis of Chemical Abstracts coding system, and is analogous to a procedure of Randić and Wilkins [67], based on sequences of path numbers, in that it can be employed for the recognition of structural similarity in molecular graphs.

The Chemical Ordering of Branched Species

Our earlier discussion has revealed that molecular graphs can in general be partially ordered in purely mathematical terms according to their degree of branching. Moreover, if the two graphs G_j and G_k within a given class (here alkane graphs) can be associated with the numbers Γ_j and Γ_k in such a way that $\Gamma_j > \Gamma_k$ whenever it is decided G_j is more branched than G_k, a measure of the branching is implied. We may now enquire whether such a measure accords with the ordering of these graphs based on the observed physicochemical properties of the molecules concerned. If the graphs in a particular class of graphs are associated with so many different values of a property over a given interval that they may be interpreted as representing a continuum of properties, the theorem of Karamata [40] can be applied. Karamata's theorem, which is valid for continuous and convex functions defined on a sequence of numbers, permits conclusions to be drawn concerning the relative magnitudes of the function if the relative magnitudes for the terms in the sequence are known. Using some well-selected subgraph structures to yield a numerical sequence, it thus becomes possible to predict the relative magnitudes of molecular properties of interest. This was accomplished by Randić et al. [69,64,68] for the alkane isomers up to $n = 10$. The selected graph invariants were paths of different lengths (especially of lengths two and three) and the physicochemical parameters ranged from boiling points through thermodynamic properties to refractive indices. In all cases the trends established by mathematical ordering corresponded with those based on physicochemical properties. This demonstrated that grid diagrams represent a convenient device for the ordering and prediction of properties, and established the significance of the underlying conceptual framework. Apparent inconsistencies or errors in the raw data are clearly revealed using this approach [68].

A different way of interpreting the behavior of branched alkane species is that based on the additive nature of most of their physicochemical properties. This way has been exhaustively investigated by Gordon and Kennedy [27,25], who postulated the idea of expressing all measurable parameters of a chemical system in terms of a linear combination of graph-theoretical invariants. Such a derived parameter, M, can be represented by the summation:

$$M = \sum_{i} \alpha_{i} N_{i} \tag{15}$$

where the α_i are coefficients, and the N_i are appropriate graph invariants. This simple formulation effectively summarizes all the manifold additivity schemes which have been proposed in the chemical literature over the past century [27]. It should be borne in mind, however, that the approach is a purely graph-theoretical one and that properties governed by stereospecificity or precise geometry will be beyond its scope. Even with this restriction, the value of equation (15) is beyond doubt, for it has been established [25] that each parameter analyzed in this way becomes stable to the introduction of further invariants beyond a certain point. The stable values can readily be calculated and used for comparisons of properties derived from mathematical ordering.

The N_i in equation (15) are, of course, topological indices and some of the indices mentioned above have been employed in this type of analysis. In particular, paths of different lengths have been widely featured [86]. Trends in more complex topological indices with branching have also been presented by several workers. Thus, for the Wiener index, W(G), Bonchev and Trinajstić [14] have given detailed mathematical expressions for the variation in the value of W(G) with the differing types of branching encountered in alkane species. In the case of the Hosoya index, Z(G), a composition principle was given [38] from which it was apparent that Z(G) depends on certain subgraphs of G for alkane isomers. Randić's molecular connectivity indices, ${}^{h}\chi(G)$, have also been investigated [41,72] with a view to interpeting their dependence on various graph invariants. In general, however, topological indices do not give good correlations with the physicochemical properties of branched species.

In an attempt to overcome this problem with topological indices, Bonchev and Mekenyan [12] introduced the concept of the comparability graph for the ordering of alkane and other isomers. The comparability graph is constructed for a complete set of isomers by making use of known rules of structural complexity, e.g. those put forward by Bonchev and Trinajstić [14]. Each rule serves to partially order the isomers by expressing trends which occur in various topological indices as systematic changes to the structure of the isomers are made. In such graphs, the vertices correspond to individual isomers and the directed edges to isomer interconversions. The paths in these oriented graphs specify the ordering of the vertices; isomers associated with different graph paths are taken to be noncomparable. Combined comparability graphs based on several different topological indices were set up for alkane isomers with $\underline{n} = 7$ (the heptanes) and $\underline{n} = 8$ (the octanes), including a total of 20 physicochemical properties. The majority of properties followed the predicted ordering; those showing the greatest deviations were the critical temperature, the Antoine equation coefficient, surface tension, molecular volume density, molecular refraction and refractive index. These properties may well depend on graph-theoretical factors not included in the invariants used in constructing the comparability graph, and also on stereochemical and geometrical effects. A similar approach based on the degree of structural similarity of pairs of isomers has recently been put forward by Grossman [28].

Conclusion

The problem of characterizing branching in a completely satisfactory way to the physical scientist is likely to remain unsolved for the forseeable future. The two main reasons for this are that (i) the notion of branching is an essentially intuitive one, and (ii) in general different physicochemical properties seem to require different orderings of sets of isomers. Thus, in spite of many highly ingenious approaches to the quantification of branching in molecular species, only a partial ordering can be attained in most cases. Such partial orderings are based on mathematical criteria such as those of Muirhead [50], and are appropriate for certain physicochemical properties, but by no means all of them. The latter properties are probably not dependent to the same extent on the molecular connectivity as the former, and in addition may also be strongly influenced by geometric or stereochemical factors. At present it is not possible to characterize molecular graphs uniquely in terms of graph invariants, but several invariants have been shown to possess high discrimination ability. Codes, however, based on the adjacency matrix, A(G), of the graph are able to provide unique characterizations of molecular graphs, although these are rather unwieldy and therefore unsuitable for most chemical correlations.

Acknowledgments

The author is indebted to the U.S. Office of Naval Research for partial support of this research project. Dr. E.R. Canfield of the Department of Computer Science at the University of Georgia is also thanked for reading an early version of this manuscript and making helpful comments.

References

1. A.T. Balaban, Applications of graph theory in chemistry, J. Chem. Inf. Comp. Sci. 25, (1985), 334-343.
2. A.T. Balaban, I. Motoc, D. Bonchev, and O. Mekenyan, Topological indices for structure-activity correlations, Topics Curr. Chem. 114, (1983), 21-55.
3. A.T. Balaban and I. Motoc, Chemical graphs XXXVI. Correlations between octane numbers and topological indices in alkanes, Math. Chem. 5, (1979), 197-218.
4. A.T. Balaban, Highly discriminating distance-based topological index, Chem. Phys. Lett. 89, (1982), 399-404.
5. A.T. Balaban and L.V. Quintas, The smallest graphs, trees, and 4-trees with degenerate topological index, J. Math. Chem. 14, (1983), 213-233.
6. M. Barysz and N. Trinajstić, A novel approach to the characterization of chemical structures, Int. J. Quant. Chem.: Quant. Chem. Symp. 18, (1984), 661-673.
7. E.F. Beckenbach and R. Bellman, Inequalities, (Springer, Berlin, 1961), pp. 30-32.
8. R.L. Bent, Aspects of isomerism and mesomerism: II. Structural isomerism, J. Chem. Educ. 30, (1953), 284-290.
9. G.S. Bloom, J.W. Kennedy, and L.V. Quintas, Some problems concerning distance and path degree sequence, Lect. Notes in Math. 1018, (1981), 179-190.
10. D. Bonchev, Information Theoretic Indices for Characterization of Chemical Structures, Research Studies Press, (Chichester, United Kingdom, 1983).
11. D. Bonchev, J.V. Knop, and N. Trinajstić, Mathematical models of branching, Math. Chem. 6, (1979), 21-47.
12. D. Bonchev and O. Mekenyan, Comparability graphs and molecular properties:

A novel approach to the ordering of isomers, J. Chem. Soc. Faraday II 80, (1984), 695-712.

13. D. Bonchev, O. Mekenyan, and N. Trinajstić, Isomer discrimination by topological information approach, J. Comp. Chem. 2, (1981), 127-148.
14. D. Bonchev and N. Trinajstić, Information theory, distance matrix, and molecular branching, J. Chem. Phys. 67, (1977), 4517-4533.
15. R.S. Cahn, C. Ingold, and V. Prelog, Specification of molecular chirality, Angew. Chem. Int. Edit. Engl. 5, (1966), 385-415.
16. A. Cayley, On the mathematical theory of isomers, Phil Mag. 47 [4], (1874), 444-447; A. Cayley, On the analytical forms called trees, with application to the theory of chemical combinations, Rept. Brit. Assoc. Advmt. Sci. (1875), 257-305.
17. L. Collatz and U. Sinogowitz, Spektren endlicher Graphen, Abhl. Math. Sem. Univ. Hamburg 21, (1957), 63-77.
18. D.M. Cvetković, M. Doob, and H. Sachs, Spectra of Graphs, (Academic Press, New York, 1980), pp. 11-12.
19. D.M. Cvetković and I. Gutman, Note on branching, Croat. Chem. Acta 49, (1977), 115-121.
20. J.T. Edward, A note on the applications of graph theory to the thermodynamic properties of isomeric alkanes, Can. J. Chem. 58, (1980), 1897-1901.
21. J.T. Edward, The relation of physical properties of alkanes to connectivity indices: a molecular explanation, Can. J. Chem. 60, (1982), 480-485.
22. S. El-Basil and A. Hussean, Matrix inverse of chemical graphs, Math. Chem. 16, (1984), 135-151.
23. S. El-Basil and A.S. Shalabi, On the ordering of Kekulé structures, Math. Chem. 14, (1983), 191-211.
24. J.W. Essam and M.E. Fisher, Some basic definitions in graph theory, Revs.

Mod. Phys. 42, (1970), 272-288.

25. J.W. Essam, J.W. Kennedy, and M. Gordon, The graph-like state of matter. Part 8. LCGI schemes and the statistical analysis of experimental data, J. Chem. Soc. Faraday II 73, (1977), 1289-1307.

26. J.L. Gay-Lussac and L.J. Thenard, Recherches Physico-Chimiques, (Déterville, Paris, 1811), Vol. II, pp. 340-345.

27. M. Gordon and J.W. Kennedy, The graph-like state of matter. Part 2. LCGI schemes for the thermodynamics of alkanes and the theory of inductive inference, J. Chem. Soc. Faraday II 69, (1973), 484-504.

28. S.C. Grossman, Chemical ordering of molecules: A graph-theoretical approach to structure-property studies, Int. J. Quant. Chem. 28, (1985), 1-16.

29. I. Gutman, A regularity for the boiling points of alkanes and its mathematical modeling, Z. Phys. Chem. in press.

30. I. Gutman and M. Randić, Algebraic characterization of skeletal branching, Chem. Phys. Lett. 47, (1977), 15-19.

31. I. Gutman, B. Ruscić, N. Trinajstić, and C.F. Wilcox, Graph theory and molecular orbitals. XII. Acyclic polyenes, J. Chem. Phys. 62, (1975), 3399-3405.

32. F. Harary, The determinant of the adjacency matrix of a graph, SIAM Rev. 4, (1962), 202-210.

33. G.H. Hardy, J.E. Littlewood, and G. Pólya, Inequalities, (Cambridge University Press, London, 1934), pp. 45-50.

34. W. Hässelbarth, Die Verzweigtheit von Graphen, Math. Chem. 16, (1984), 3-17.

35. H.R. Henze and C.M. Blair, The number of isomeric hydrocarbons of the methane series, J. Amer. Chem. Soc. 53, (1931), 3077-3085.

36. W.C. Herndon and J.E. Leonard, Canonical numbering, stereochemical

descriptors, and unique linear notations for polyhedral clusters, Inorg. Chem. 22, (1983), 554-557.

37. H. Hosoya, and M. Randić, Analysis of the topological dependancy of the characteristic polynomial in its Chebyshev expansion, Theor. Chim. Acta 63, (1983), 473-495.
38. H. Hosoya, Topological index. A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons, Bull. Chem. Soc. Japan 44, (1971), 2332-2339.
39. H. Hiz, A linearization of chemical graphs, J. Chem. Docum. 4, (1964), 173-180.
40. J. Karamata, Sur une inegalité relative aux fonctions convexes, Publ. Math. Univ. Belgrade 1, (1932), 145-148.
41. L.B. Kier and L.H. Hall, Molecular Connectivity in Chemistry and Drug Research, (Academic Press, New York, 1976).
42. J.V. Knop, W.R. Müller, Z. Jericević, and N. Trinajstić, Computer enumeration and generation of trees and rooted trees, J. Chem. Inf. Comp. Sci. 21, (1981), 91-99.
43. J.V. Knop, W.R. Müller, K. Szymanski, and N. Trinajstić, Computer Generation of Certain Classes of Molecules, Association of Chemists and Technologists of Croatia, (Zagreb, Yugoslavia, 1985).
44. L. Lovász and J. Pelikán, On the eigenvalues of trees, Period. Math. Hung. 3, (1973), 175-182.
45. A.L. MacKay, On rearranging the connectivity matrix of a graph, J. Chem. Phys. 62, (1975), 308-309.
46. R.B. Mallion, Some chemical applications of the eigenvalues and eigenvectors of certain finite, planar graphs in Applications of Combinatorics, ed. R.J. Wilson, (Shiva Press, Cheshire, United Kingdom, 1982), chap. 7, pp. 87-114.
47. O. Mekenyan, D. Bonchev, and A.T. Balaban, Unique description of chemical

structures based on hierarchically ordered extended connectivities (HOC procedures). V. New topological indices, ordering of graphs, and recognition of graph similarity, J. Comp. Chem. 5, (1984), 629-639.

48. H.L. Morgan, The generation of a unique machine description for chemical structure, J. Chem. Docum. 5, (1965), 107-113.
49. I. Motoc, A.T. Balaban, O. Mekenyan, and D. Bonchev, Topological indices: Inter-relations and composition, Math. Chem. 13, (1982), 369-404.
50. R.F. Muirhead, Inequalities relating to some algebraic means, Proc. Edin. Math. Soc. 19, (1901), 36-45.
51. J.J. Mulckhuyze, Molecules and Models: Investigations on the Axiomatization of Structure Theory in Chemistry, Thesis, (University of Amsterdam, Amsterdam, 1960), chap. 2, pp. 6-7.
52. J.F. Nagle, On ordering and identifying undirected linear graphs, J. Math. Phys. 7, (1966), 1588-1592.
53. M. Randić, G.M. Brissy, and C.L. Wilkins, Computer perception of topological symmetry via canonical numbering of atoms, J. Chem. Inf. Comp. Sci. 21, (1981), 52-59.
54. M. Randić, Characterization of atoms, molecules and classes of molecules based on paths enumerations, Math. Chem. 7, (1979), 5-64.
55. M. Randić, Compact codes 3. On nomenclature of acyclic chemical compounds, J. Chem. Inf. Comp. Sci., in press.
56. M. Randić, On alternative form of the characteristic polynomial and the problem of graph recognition, Theor. Chim. Acta 62, (1983), 485-498.
57. M. Randić, On canonical numbering of atoms in a molecule and graph isomorphism, J. Chem. Inf. Comp. Sci. 17, (1977), 171-180.
58. M. Randić, On characterization of molecular branching, J. Amer. Chem. Soc. 97, (1975), 6609-6615.

59. M. Randić, On comparability of structures, Chem. Phys. Lett. 55, (1978), 547-551.

60. M. Randić, On molecular identification numbers, J. Chem. Inf. Comp. Sci. 24, (1984), 164-175.

61. M. Randić, On the characteristic equations of the characteristic polynomial, SIAM J. Alg. Disc. Meth. 6, (1985), 145-162.

62. M. Randić, On the recognition of identical graphs representing molecular topology, J. Chem. Phys. 60, (1974), 3920-3928.

63. M. Randić, Private communcation.

64. M. Randić and N. Trinajstić, On isomeric variations in decanes, Math. Chem. 13, (1982), 271-290.

65. M. Randić, N. Trinajstić and T. Zivković, Molecular graphs having identical spectra, J. Chem. Soc. Faraday Trans II 72, (1976), 244-256.

66. M. Randić and C.L. Wilkins, Graph-theoretical analysis of molecular properties. Isomeric variations in nonanes, Int. J. Quant.Chem. 18, (1980) 1005-1027.

67. M. Randić and C.L. Wilkins, Graph-theoretical approach to recognition of structural similarity in molecules, J. Chem. Inf. Comp. Sci. 19, (1979), 31-37.

68. M. Randić and C.L. Wilkins, Graph-theoretical ordering of structures as a basis for systematic searches for regularities in molecular data, J. Phys. Chem. 83, (1979), 1525-1540.

69. M. Randić and C.L. Wilkins, On a graph-theoretical basis for ordering of structures, Chem. Phys. Lett. 63, (1979), 332-336.

70. M. Randić, W.L. Woodworth, and A. Graovac, Unusual random walks, Int. J. Quant. Chem. 24, (1983), 435-452.

71. C. Raychaudhury, S.K. Ray, J.J. Ghosh, A.B. Roy, and S.C. Basak, Discrimination of isomeric structures using information theoretic topological indices, J. Comp. Chem. (1984), 5, 581-588.

72. M. Razinger, J.R. Chrétien, and J.E. Dubois, Structural selectivity of topological indexes in alkane series, J. Chem. Inf. Comp. Sci. 25, (1985), 23-27.

73. D.H. Rouvray, Isomer enumeration methods, Chem. Soc. Revs. (London) 3, (1974), 355-372.

74. D.H. Rouvray, Should we have designs on topological indices?, in Chemical Applications of Topology and Graph Theory, (ed. R.B. King, Elsevier, Amsterdam, 1983), pp. 159-177.

75. D.H. Rouvray, Some reflections on the topological structure of covalent molecules, J. Chem. Educ. 52, (1975), 768-773.

76. D.H. Rouvray, The role of the topological distance matrix in chemistry, in Applications of Mathematical Concepts to Chemistry, (ed. N. Trinajstić, Ellis Horwood Publishers, Chichester, United Kingdom, 1986).

77. E. Ruch, Algebraic aspects of the chirality phenomenon in chemistry, Accts. Chem. Res. 5, (1972), 49-56.

78. E. Ruch and I. Gutman, The branching extent of graphs, J. Comb. Inform. and Syst. Scis 4, (1979), 285-295.

79. E. Ruch and A. Schönhofer, Theorie der Chiralitätsfunktionen, Theor. Chim. Acta 19, (1970), 225-287.

80. E. Ruch, The diagram lattice as structural principle, Theor. Chim. Acta 38, (1975), 167-183.

81. H. Sachs, Beziehungen zwischen den in einem Graphen enthaltenen Kreisen und seinem charakteristischen Polynom, Publ. Math. Debrecen 11, (1964), 119-134.

82. W. Schubert and I. Ugi, Constitutional symmetry and unique descriptors of molecules, J. Amer. Chem. Soc. 100, (1978), 37-41.

83. A.J. Schwenk, W.C. Herndon, and M.L. Ellzey, The construction of cospectral composite graphs, Annals New York Acad. Scis 319, (1979), 490-496.

84. Z. Slanina, Chemical isomerism and its contemporary theoretical description, Advs. Quant. Chem. 13, (1981), 89-153.

85. P.J. Slater, Counterexamples to Randić's Conjecture on Distance Degree Sequences for Trees, J. Graph Theory 6, (1982), 89-92.

86. G.R. Somayajulu and B.J. Zwolinski, Generalized treatment of Alkanes. Part 3. Triatomic additivity, J. Chem. Soc. Faraday II 70, (1974), 967-972.

87. K. Szymanski, W.R. Müller, J.V. Knop and N. Trinajstić, On Randić's Molecular Identification Numbers, J. Chem. Inf. Comput. Sci. 25, (1985), 413-415.

88. H. Wiener, Structural determination of paraffin boiling points, J. Amer. Chem. Soc. 69, (1947), 17-20.

89. A. Young, On quantitative substitutional analysis, Proc. Lond. Math. Soc. [1]33, (1901), 97-146.

90. J. Yuangsheng, Problem on isospectral molecules, Sci. Sinica 27B, (1984), 236-248.

91. T. Zivković, N. Trinajstić and M. Randić, On conjugated molecules with identical topological spectra, Mol. Phys. 30, (1975), 517-533.

Captions for Tables

Table 1. Number of alkane constitutional isomers (trees) for various values of n, the number of carbon atoms.

Table 2. Number of paths of length i with ($1 \leq i \leq 7$) for the alkane isomers having $n = 8$ (the octanes).

Captions for Figures

Figure 1. Hydrogen-suppressed graphs of the octane isomers (a) 2-methylheptane, (b) 3-methylheptane, and (c) 4-methylheptane.

Figure 2. The nonascending vertex degree sequence and Young diagram for each of the three isomers in Figure 1.

Figure 3. Grid of the 18 octane isomers showing an ordering based on the number of paths of length two (p_2) and of length three (p_3) in each.

Figure 4. Canonical labelings and canonical codes for the three pentane isomers.

Figure 5. Pairs of hydrogen-suppressed alkane graphs displaying identical indices of varying kinds.

Figure 6. A scatter plot of boiling point against Wiener index, W(G), for the 75 decane isomers.

Number of Carbon Atoms	Number of Alkane Isomers
1	1
2	1
3	1
4	2
5	3
6	5
7	9
8	18
9	35
10	75
20	366319
30	4111846763
40	62481801147341
50	1117743651746953270

Table 1: Number of alkane constitutional isomers (trees) for various values of $\underline{n}$, the number of carbon atoms.

TABLE 1

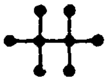
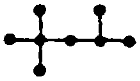
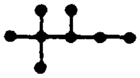
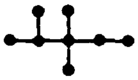
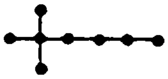
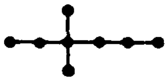
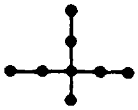
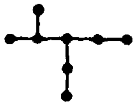
Name of Molecule	Graph	Number of Paths p_i of Length i						
		p_1	p_2	p_3	p_4	p_5	p_6	p_7
2,2,3,3-Tetramethylbutane		7	12	9				
2,2,4-Trimethylpentane		7	10	5	6			
2,2,3-Trimethylpentane		7	10	8	3			
2,3,3-Trimethylpentane		7	10	9	2			
2,3,4-Trimethylpentane		7	9	8	4			
2,2-Dimethylhexane		7	9	5	4	3		
3,3-Dimethylhexane		7	9	7	4	1		
2,5-Dimethylhexane		7	8	5	4	4		
2,4-Dimethylhexane		7	8	6	5	2		
2,3-Dimethylhexane		7	8	7	4	2		
3-Methyl-3-Ethylpentane		7	9	9	3			
2-Methyl-3-Ethylpentane		7	8	8	5			
3,4-Dimethylhexane		7	8	8	4	1		
2-Methylheptane		7	7	5	4	3	2	
3-Methylheptane		7	7	6	4	3	1	
4-Methylheptane		7	7	6	5	2	1	
3-Ethylhexane		7	7	7	5	2		
n-Octane		7	6	5	4	3	2	1

TABLE 2

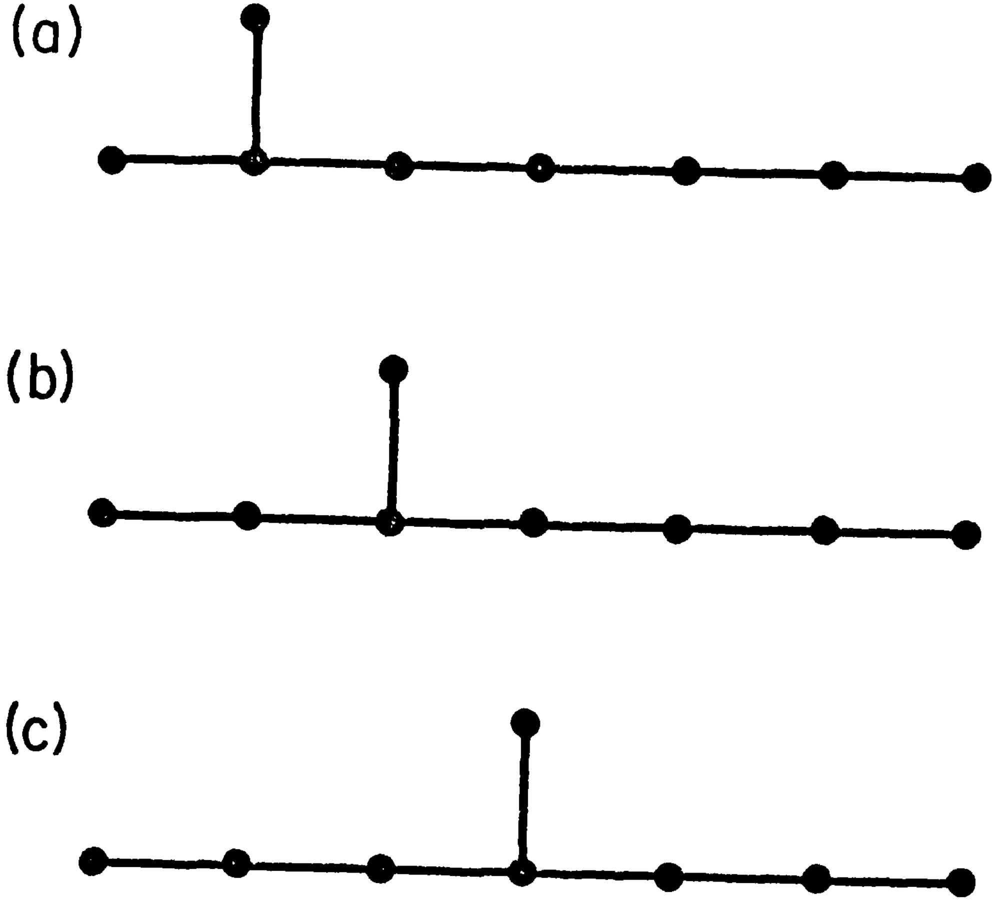


FIGURE 1

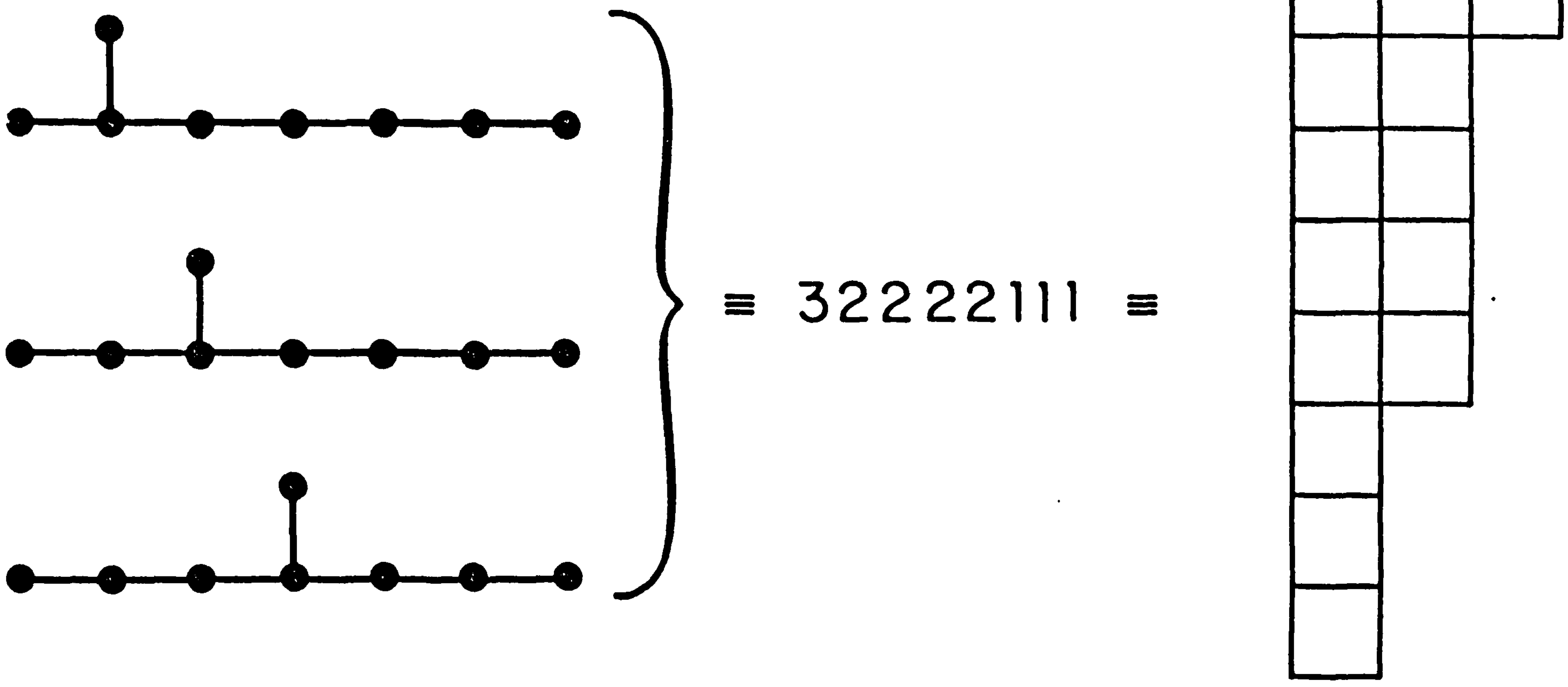
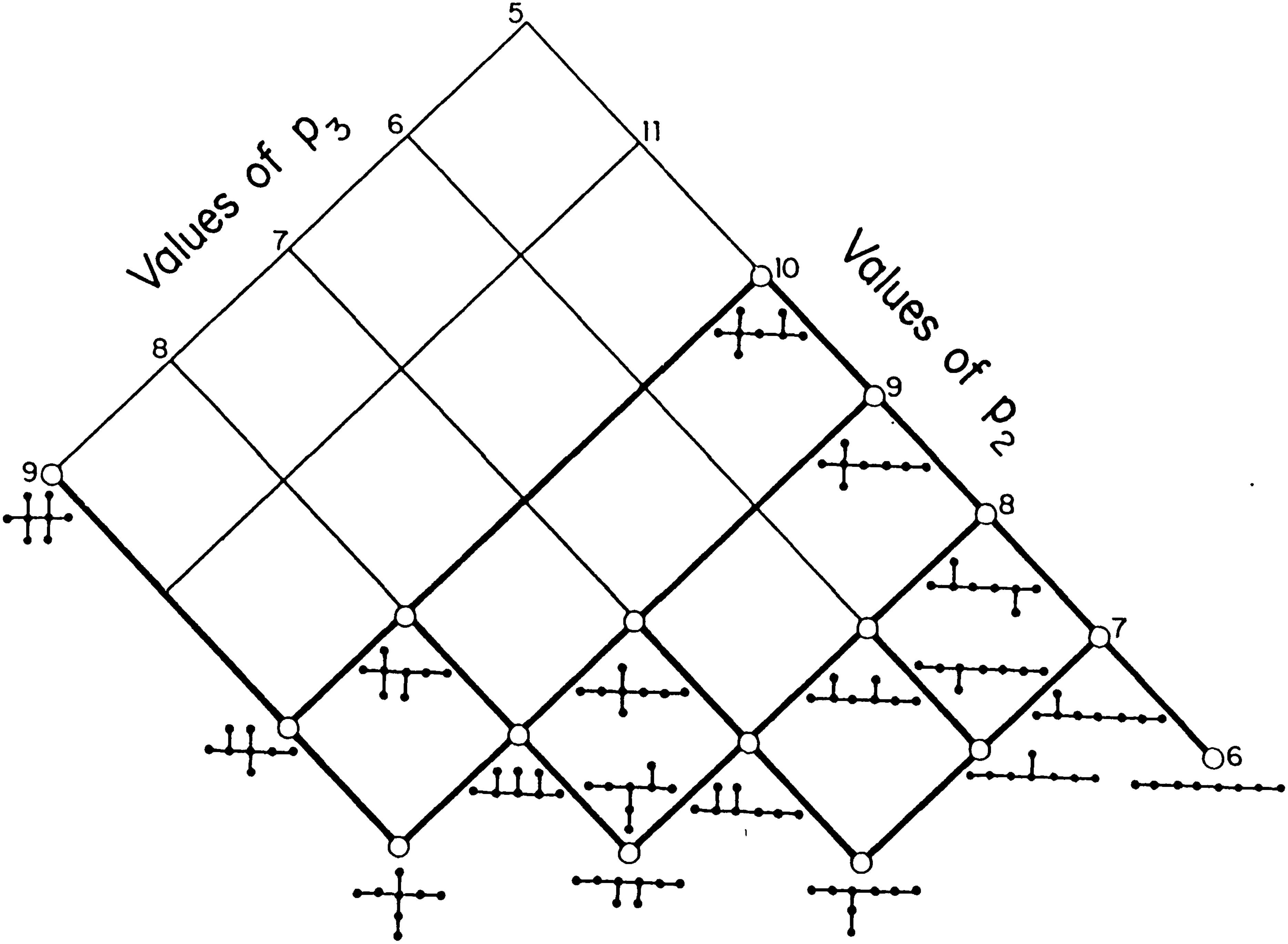
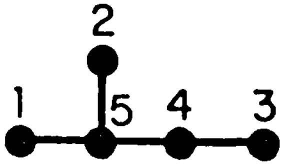


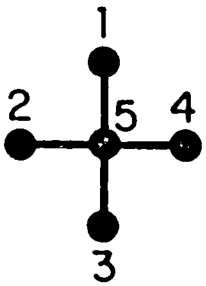
FIGURE 2


5
Values of p_3
6
7
8
9
11
10
Values of p_2
9
8
7
6

$\equiv$ 0000100010000110110010100



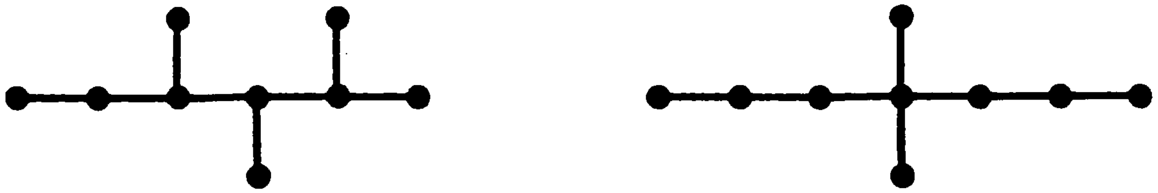
$\equiv$ 0000100001000100010111010



$\equiv$ 0000100001000010000111110

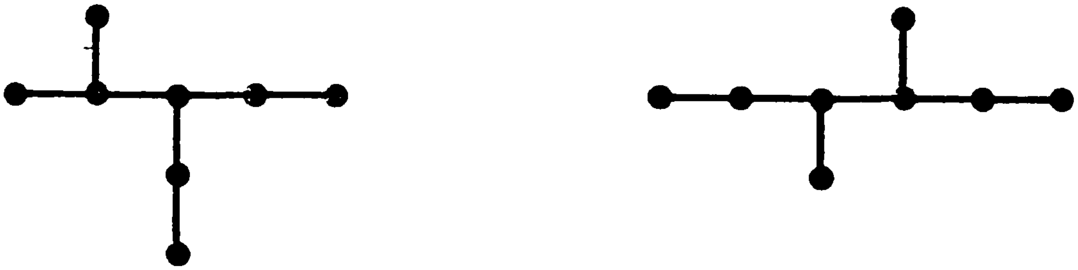
FIGURE 4

Isospectral



Same Wiener Index

Same Connectivity Index



Same Balaban Index

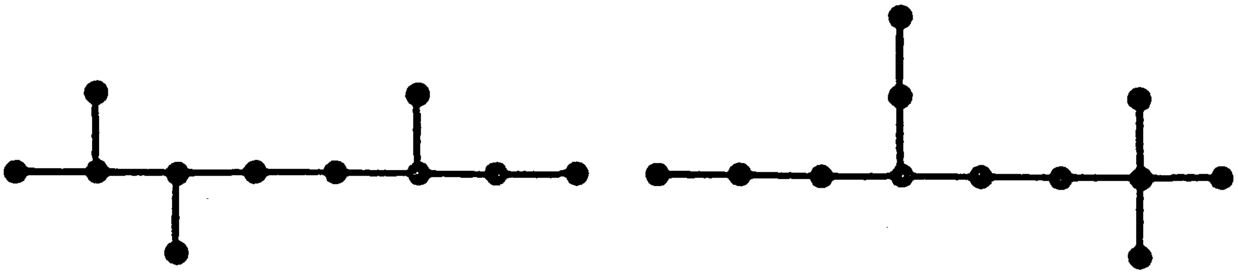


FIGURE 5

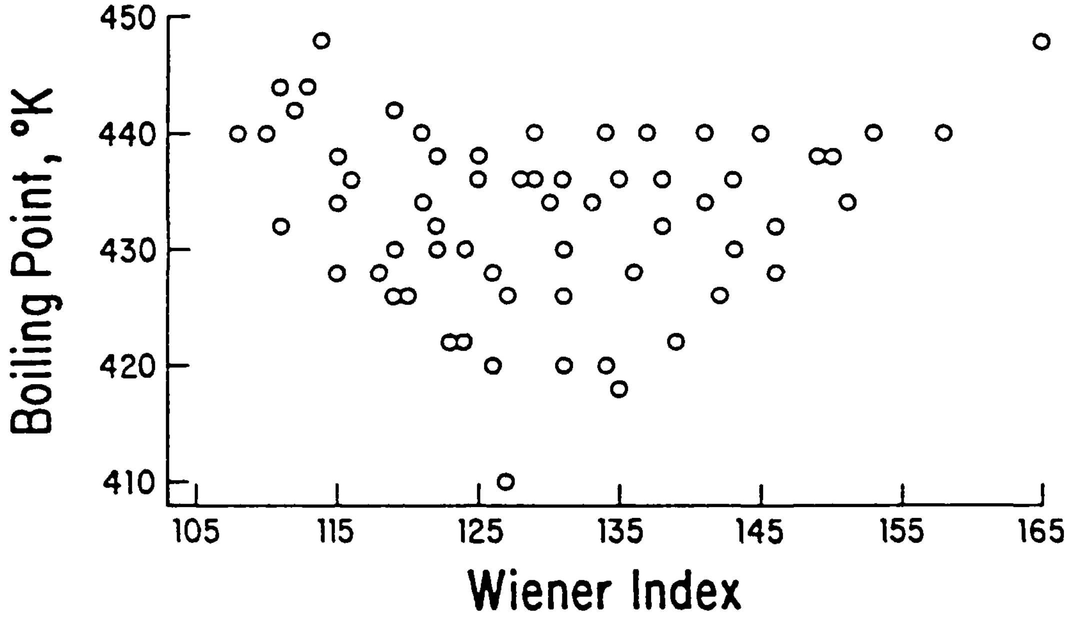

450
440
430
420
410
Boiling Point, °K
105
115
125
135
145
155
165
Wiener Index

FIGURE 6

END

DTIC

8—86