Extensions of the Wiener Number

H.-Y. Zhu,[†] D. J. Klein,*[,†] and I. Lukovits[‡]

Texas A&M University at Galveston, Galveston, Texas 77553-1675, and Central Research Institute for Chemistry, Hungarian Academy of Science, H-1525 Budapest, P.O. Box 17, Hungary

Received September 8, 1995[⊗]

Particularly for structure−property correlations there are many chemical graph-theoretic indices, one of which is Wiener's "path number". Because Wiener's original work focused on acyclic structures, one can imagine different ways of extending it to cycle-containing structures, several of which are noted here. Many of these different formulas in fact yield like numerical values for general connected graphs—that is, different formulas sometimes correspond to the same graph invariant. Indeed it is found that there are two "dominant" classes of formulas each corresponding to one of two distinct invariants. Extensions to sequences of invariants (with the Wiener index the first member) are more often found to give distinct sequences. Further a powerful vector-space theoretic view for characterizing and for comparing different sequences is described. This is illustrated for a collection of eight sequences in application to a set of molecular graph structures (corresponding to the octanes). Another type of extension is to generate a sequence or partially ordered set of graph invariants for which the Wiener index is the natural first member of this set. Certain such sets of invariants (corresponding to contributions from different types of subgraphs) are noted to be "complete" (or form a basis) in the sense that any invariant can be faithfully linearly expanded in terms of the members of the set.

1. INTRODUCTION

The use of graph invariants as molecular descriptors has come to be widely recognized, studied, and utilized. The general idea[1] of course dates back more than a century to when[2] it was viewed as a possibility that molecular structure might be completely subsumed by "graph theory" (though this mathematical field was not yet recognized).

In the late 1940s Wiener[3] proposed a graph invariant and utilized it for (saturated) acyclic hydrocarbons. This invariant now denoted $W(G)$ for graph G was identified as the "path number" by Wiener.[3] Moreover, $W(G)$ was somewhat different from most previous invariants which tended to be atom or bond counts or atom- or bond-type counts. More recently $W(G)$ has come to be much studied, with a few reviews[4] in the chemical literature, and with some work[5−7] in the mathematical literature.

Because Wiener focused on (saturated) acyclic structures and described $W(G)$ as the "path number" and as the sum of distances between all pairs of sites, while giving a rather "different" computational formula, there arises different natural possibilities for extending $W(G)$ to cycle-containing (connected) graphs. Indeed some such possibilities have been suggested.[8,9]

Here in sections 2 and 3 a number of extensions are identified. Whenever a "nice" formula for $W(G)$ for the case of trees (i.e., acyclic structures) is found, an extension beyond trees to cycle-containing structures might be imagined. Indeed though this approach has been taken before,[8,9] several of the present formulas seem not to have been so previously contemplated, and though the formulas (along with some attendant notation) are introduced, proofs are delayed to section 4. There it is found that many equalities occur amongst these extensions of formulas beyond trees—in fact many of these extensions fall into one of two classes each corresponding to a distinct graph invariant (for non-trees).

In section 5 another type of extension is made to a natural sequence of invariants such that $W(G)$ is the first member of the sequence. For instance, in a summation formula for $W(G)$ for the case of trees, one might generate the nth member of the sequence via the same summation with the summand now raised to the nth power. Again a few such possibilities have been previously noted,[10−12] though again we find a greater variety of possibilities. There seem to be a paucity of equalities between the different sequences, though in sections 6 and 7 we take the first steps in a study of dependencies within and between sequences.

Section 8 continues with extensions to partially ordered sets of invariants such that $W(G)$ is the first member of the set. For instance, in a formula for $W(G)$ in terms of a summation over edge contributions, one might generate other members of the set as summations over analogous contributions from other types of subgraphs. There seems not to have been previous consideration of such Wiener-number-based sets, though the approach is quite similar to rather classical substructure counting schemes based upon site-contributions first, bond-contributions second, adjacent bond-pair contributions next, etc. In common with these substructure counting schemes,[13−16] the present sets can be "complete" in the sense that any other invariant can be faithfully expanded in terms of the members of such a set. This type of completeness result in fact has generalization[17] beyond the Wiener-number-related context here.

Throughout this paper graphs G appearing as arguments of the Wiener number are assumed to be *connected*. More-or-less standard notation[18] is used. The vertex and edge sets for a graph G' (whether connected or not) are denoted $V(G')$ and $E(G')$, and $|G'|$ denotes the number of vertices in $V(G')$.

[†] Texas A&M University at Galveston.
[‡] Hungarian Academy of Science.
[⊗] Abstract published in *Advance ACS Abstracts,* December 15, 1995.

0095-2338/96/1636-0420$12.00/0

2. DISTANCE-MATRIX FORMULAS

There are a number of different ways of expressing Wiener-number-related quantities in terms of distance matrices. Let $\mathbf{D} = \mathbf{D}_1$ be the ordinary *shortest-path* distance matrix for a connected graph G, or more generally $\mathbf{D}_q$ is the matrix with (i,j)th element $(d_{ij})^q$ where d_{ij} is the length of a shortest path between vertices i and j of G. The standard formula for (twice) the Wiener number[4] is

$$\sum_{i,j}^{G} d_{ij} = (\phi|\mathbf{D}_1|\phi) \tag{2.1}$$

where the sum is over all elements of $\mathbf{D}_1$, and where $(\phi|$ and $|\phi)$ are row and column vectors of all ones. Also we let $(i|$ and $|i)$ denote row and column vectors of all 0s except for a 1 in the ith position, $i \in V(G)$. Then another formula

$$\sum_{i}^{G} (i|\mathbf{D}_{1/2}{}^2|i) = \mathrm{tr}(\mathbf{D}_{1/2}{}^2) \tag{2.2}$$

is the trace of the square matrix $\mathbf{D}_{1/2}$, which may itself be viewed[12,19] as a distance matrix. That is, the function $d_{ij}^{1/2}$ satisfies the postulates of a distance function: it is symmetric in its arguments (i and j), it is positive if $i \neq j$, it is 0 if $i = j$, and satisfies the triangle inequality. But in common with real Euclidean distances, and unlike the distances $\mathbf{D}_1$, the distances of $\mathbf{D}_{1/2}$ generally satisfy the "tetrahedron inequality" (wherein the "area" of a face of a tetrahedron defined by four points never exceeds the sum of the areas of the other three faces). That this "square-rooted" path distance so more faithfully mimics Euclidean distances suggests it (as well as related invariants) as a natural candidate for investigation.

But there are yet other fundamental distance functions beyond that of $\mathbf{D}_1$ or $\mathbf{D}_{1/2}$. Of particular note is the so-called *resistance distance* Ω_{ij} being the (effective) "electrical" resistance between vertices i and j if unit resistors are imagined on each edge of the (connected) graph G. It has been argued[8] that this distance function has some features more desirable than the shortest-path distance function. Now in analogy to (2.1) and (2.2) one has

$$\sum_{i,j}^{G} \Omega_{ij} = (\phi|\Omega_1|\phi) \tag{2.3}$$

and

$$\sum_{i}^{G} (i|\Omega_{1/2}{}^2|i) = \mathrm{tr}(\Omega_{1/2}{}^2) \tag{2.4}$$

where the matrix $\Omega_{1/2}$ of elements $(\Omega_{ij})^{1/2}$ again is the distance matrix for a distance function satisfying the "tetrahedron inequality" (and more).

As it turns out that both $\mathbf{D}$ and Ω are related (at least for trees) to yet another fundamental matrix. The *Laplacian* matrix of G is $\Delta - \mathbf{A}$ where $\mathbf{A}$ is the standard adjacency matrix of G and Δ is the diagonal matrix of vertex degrees. Being a discretized version of the classical Laplacian of calculus, this matrix is of natural interest, and much graph-theoretic work on it has been reviewed.[20] Then another matrix Γ is defined as the generalized inverse of $\Delta - \mathbf{A}$—that is, Γ is 0 on the one-dimensional *null* space (spanned by $|\phi)$ and otherwise Γ is the inverse to $\Delta - \mathbf{A}$ (projected onto the orthogonal compliment to this null space). This Green's-function-like matrix Γ is related to Ω, with $\Omega_{ij} = \Gamma_{ii} - 2\Gamma_{ij} + \Gamma_{jj}$.[8] One might then consider the invariant

$$\sum_{i}^{G} \Gamma_{ij} = \mathrm{tr}(\Gamma) = \sum_{a}^{\neq 0} \lambda_a^{-1} \tag{2.5}$$

where the last sum is over the nonzero eigenvalues λ_a of the Laplacian $\Delta - \mathbf{A}$.

3. SUBGRAPH-SUM FORMULAS

There are a number of different ways of expressing Wiener-number related quantities in terms of summations over contributions of all supgraphs of a certain type. The subgraph types considered are all subgraphs with one vertex (i.e., isomorphic to K_1), all edge subgraphs (i.e., isomorphic to K_2), all geodiesic paths, and all spanning trees.

First we consider different types of vertex-contributions, for a vertex $i \in V(G)$. Most simply there are five possibilities with contributions for vertex i obtained as the ith summand in any one of the formulas of section 2. These five possibilities might be abbreviated to

$$(2.m)_i \tag{3.1}$$

corresponding to the ith summand of equation (2.m) for $m = 1, 2, 3, 4$, and 5. Indeed the distance sums $(2.1)_i = \sum_j d_{ij} \equiv s_i$ to a given point i have been considered sometime back[21,22] and have found useage.[23] Another approach is based on the graph $G-i$ obtained from G upon deleting vertex i (and all edges incident thereon). Thence a vertex-i contribution might be

$$\frac{N(N-1)}{2} + \#_2(G-i) \equiv \frac{N(N-1)}{2} + \sum_{\gamma \neq \gamma'}^{G-i} |\gamma|\cdot|\gamma'| \tag{3.2}$$

where N is the number of vertices in G and the sum is over products of numbers of vertices in pairs of components $\{\gamma,\gamma'\}$ of $G-i$, with the sum being understood to be 0 if $G-i$ has just one component. Another vertex-invariant utilizes *geodesics* g, such being a path g for which there is no shorter length path between the end points of g. Then a possible vertex contribution is

$$\sum_{g}^{i} \delta_i(g)/\#_g(G) \tag{3.3}$$

where the sum is over geodesics going through i, and $\#_g(G)$ is the number of different geodesics in G between the endpoints of g. The ambiguity here in what we mean by paths "going through i" is resolved by taking the factor $\delta_i(g) = 1$ if the path g fully passes through with i as an interior vertex to g, while $\delta_i(g) = {}^1/_2$ if the path ends at i (and $\delta_i(g) = 0$ for the 1-vertex path of length 0). Here (3.3) has an interpretation as the degree of "centrality" of a vertex i in terms of geodesics passing through that vertex, while (3.2) is a measure of "strong centrality" where attention is paid to whether the vertex is actually a "cut vertex" disconnecting the graph.

Second, we consider different types of edge-contributions, for an edge $e \in E(G)$. Let $G-(e)$ denote the subgraph obtained from G by deleting e but not the vertices incident

at the ends of e. Then

$$\#_2(G-(e)) \equiv \sum_{\gamma \neq \gamma'}^{G-(e)} |\gamma| \cdot |\gamma'| \tag{3.4}$$

is a sum over products of pairs of distinct components of $G-(e)$. A related[9] so-called[24] "Szeged" invariant is

$$\#_{\text{near}}(e,G) = \prod_{i}^{\in e} \#_i(e,G) \tag{3.5}$$

where for a general subgraph $G' \subseteq G$ we define the quantity $\#_i(G',G)$ for $i \in V(G')$ as the number of vertices of G nearer to i than any other vertex in $V(G')$. Of course "nearness" can generally be differently judged by different distance functions. That is, for $e = \{i,j\}$, $\#_{\text{near}}(e,G)$ is the product of the number of vertices nearer i times the number nearer j. Also another possibility is

$$\sum_{g}^{e} 1/\#_g(G) \equiv W_e(G) \tag{3.6}$$

where the sum is over geodesics containing e. Much as (3.2) and (3.3) were measures of vertex centrality, (3.5) and (3.6) are measures of edge centrality. Another edge-invariant utilizes *spanning* trees T, each such being a connected acyclic subgraph containing all the vertices of G. Then a possible edge-e contribution is

$$\sum_{T}^{e} \#_2(T-(e))/\{\#_{\text{tree}}(T-(e),G) \cdot \#_{\text{tree}}(G)\} \equiv W_e(G) \tag{3.7}$$

where the sum is over spanning trees of which the total number is $\#_{\text{tree}}(G)$ and $\#_{\text{tree}}(G,G')$ is the number of spanning trees of G containing G'. Again this is some sort of measure of edge "centrality" now in terms of the edge's centrality within spanning trees containing that edge.

Third, contributions for geodesics or for spanning trees may be identified. As a contribution for geodesic g we consider

$$(|g| - 1)/\#_g(G) \tag{3.8}$$

As a contribution for a spanning tree T we consider

$$\sum_{e}^{T} \#_2(T-(e))/\{\#_{\text{tree}}(T-(e),G) \cdot \#_{\text{tree}}(G)\} \equiv W_T'(G) \tag{3.9}$$

where the sum is over edges in T.

Because in trees (as originally considered by Wiener) there is no distinction between geodesics and general paths, the formulas of (3.3), (3.6), and (3.8) could in fact be viewed to involve paths instead of just geodesics. Indeed from Wiener's nomenclature[3] of "path number" one could even imagine that formulas could be taken as in (3.6) without the factor $1/\#_g(G)$—the summation being either over geodesics or over paths. Yet another possibility reproducing Wiener's result for trees could even go to a further extreme viewing (3.3), (3.6), and (3.8) to involve longest paths (or "detours"), with or without the denominators counting such paths—in fact one such choice ((3.6) with the denominator) has been studied.[25] But these possibilities are not addressed further here. Further it might be noted that other types of formulas beyond those we have listed are possible, as Doyle and Garver's[6] formula in terms of triples of "noncolinear" points.

4. EQUIVALENCES

Amongst the different formulas introduced in the previous two sections there are in fact but only a few really distinct graph invariants, the two most frequently occurring ones being denoted $W(G)$ and $W'(G)$. To describe these equivalences we identify the different formulas by their equation numbers in the following:

Theorem A—For a connected graph G the invariants of the preceding two sections are related:

$$\begin{aligned}
W(G) &= \frac{1}{2}(2.1) = \frac{1}{2}\sum_{i,j}^{G} d_{ij} \\
&= \frac{1}{2}(2.2) = \text{tr}(\mathbf{D}_{1/2}{}^2) \\
&= \frac{1}{2}\sum_{i}^{G}(3.3)_i = \frac{1}{2}\sum_{i}^{G}\{\sum_{g}^{i} \delta_i(g)/\#_g(G)\} \\
&= \sum_{e}^{G}(3.6)_e = \sum_{e}^{G} W_e(G) \\
&= \sum_{g}^{G}(3.8)_g = \sum_{g}^{G}(|g|-1)/\#_g(G) \\
W'(G) &= \frac{1}{2}(2.3) = \frac{1}{2}\sum_{i,j}^{G} \Omega_{ij} \\
&= \frac{1}{2}(2.4) = \text{tr}(\Omega_{1/2}{}^2) \\
&= N \cdot (2.5) = N \cdot \text{tr}(\Gamma) \\
&= \frac{1}{2}\sum_{e}^{G}(3.7)_e = \sum_{e}^{G} W_e'(G) \\
&= \sum_{T}^{G}(3.9)_T = \sum_{T}^{G} W_T'(G)
\end{aligned}$$

Proof: First the equalities to $W(G)$ are addressed. Of these the first equality here may be taken as the standard definition of $W(G)$, while the second and third lines are trivial. For the fourth line, the vertex-sum over all the terms as of (3.3) leads to a contribution of $1/\#_g(G)$ for each internal vertex of each geodesic as well as a contribution of $1/2\#_g(G)$ for each end vertex of each (nonzero) geodesic—thence this gives a contribution of $length(g)/\#_g(G)$ for each geodesic, or upon summing just the geodesics with a common pair of endpoints $\{i,j\}$ one obtains a contribution d_{ij} for each $\{i,j\}$—so that, finally the first equality is reproduced. The equality involving the contributions of (3.6) is obtained in a similar fashion, having already been noted.[26] The result for (3.8) involves just a reshuffling of summations.

Next the equalities to $W'(G)$ are addressed. The first line of these equalities here may be taken as the standard definition of $W'(G)$, and the second line is trivial (though in

the next section they lead to different sequences). The equivalence of the third equality to the first has previously been established.[8] As a first step toward establishing the equality involving (3.7) we note that the resistance distance Ω_{ij} can be given[27] a graphico-combinatorial interpretation

$$\Omega_{ij} = \frac{\#_{bi-t}(i,j;G)}{\#_{\text{tree}}(G)} \tag{4.1}$$

where the numerator here is the number of spanning bi-trees with i and j in different components (a *bi-tree* being a graph with two components each of which is a tree). Deletion of an edge from a spanning tree clearly gives a bi-tree (and every spanning bi-tree B can be so obtained, for on adding any edge of G between the two sets of vertices associated to the two components of B there results a spanning tree of G)—moreover, the number of times a spanning bi-tree B is obtained upon deletion of an edge from some spanning tree is just $\#_{\text{tree}}(B,G)$, in the notation of (3.7). Now the summation of the edge contributions of (3.7) over all edges of G can be rewritten as

$$\sum_{e}^{G} (3.7)_e = \sum_{B}^{G} \#_2(B)/\#_{\text{tree}}(G) \tag{4.2}$$

Now $\#_2(B)$ is just the product of the numbers of vertices in its two components, say τ_1 and τ_2—but then this can be replaced by a double sum over vertices of τ_1 and τ_2, thusly

$$\sum_{e}^{G} (3.7)_e = \sum_{B}^{G} \sum_{i}^{\tau_1} \sum_{j}^{\tau_2} \frac{\#_2(B)}{\#_{\text{tree}}(G)} = \sum_{i<j} \sum_{B}^{ij} \frac{1}{\#_{\text{tree}}(G)} \tag{4.3}$$

where the last sum on B is just over those spanning bi-trees with i and j in its different components (τ_1 and τ_2). The summand of this B-summation is now independent of B, so that this sum just gives $\#_{\text{bi-t}}(i,j;G)$, and the result for $\Omega_{i,j}$, so that this equality for $W'(G)$ is established. Finally the equality involving the tree-contributions of (3.9) comes from just reversing the summations involved in the summation over edges of (3.7).

There are some other (rather trivial) equalities which could be included in theorem A—namely those associated with (3.1). That is, i-summation of $(2.m)_i$ would give $(2.m)$ for $m = 1, 2, 3, 4, 5$, and thence lead to $W(G)$ or $W'(G)$. The expressions of (3.1) have been separately noted since they will often give new results in the next section.

That all the noted quantities extend Wiener's ideas for the special case of trees is rigorously enunciated as the following:

Theorem B—For a tree T there are equalities beyond those of theorem A:

$$W(T) = W'(T) = \sum_{i}^{T} (3.2)_i = \sum_{e}^{T} (3.4)_e = \sum_{e}^{T} (3.5)_e$$

Proof: The equality between $W(T)$ and $W'(T)$ is immediately seen from the equality[8] between d_{ij} (appearing in (2.1) for $W(G)$) and Ω_{ij} (appearing in (2.3) for $W'(G)$) when G is a tree—with only one path between i and j the effective resistance Ω_{ij} is just the sum of the values of the intervening (series-connected) resistors. The equality for trees of $W(T)$ to the formula of (2.5) (here associated with W') has also been previously noted.[7] The formula of (3.2) for trees essentially reproduces that of (3.3) already identified with W in theorem A. Likewise for trees (3.4) and (3.5) reproduce that of (3.6) identified with W.

We see that there are several conceivable extensions of Wiener's path number, but that of those several given prominence here the different formulas really identify but a few distinct invariants for general connected G. The formulas (3.2) and (3.4) may in some sense be viewed to imitate (3.5). However, (3.2) and (3.4) are not very successful for the many graphs without cut vertices or cut edges, since then the summations in these formulas become vacuous. Perhaps the most fundamental new formula is that of (3.7) which is in some sense the resistance-distance analogue of the shortest-path-distance formula of (3.6).

5. EXTENSIONS TO SEQUENCES

The various possible expressions identified for the Wiener number or related invariants may be extended in different ways to a linear sequence of invariants. That is, one might simply modify (by inclusion of a suitable power n the expressions of sections 2 and 3 to give sequences as $W_n(G)$, $n = 1, 2, \ldots$ such that $W_1(G)$ is just $W(G)$ or $W'(G)$. But even so there are different manners of power modification:

(a) Rather simply raise the initial (Weiner-number) invariant to powers, such being a classical technique in least-squares fittings, though here we have in mind more "fundamental" extensions.

(b) Raise matrix elements (within a summation) to a power, as $(d_{ij})^n$ in (2.1), $(i|\mathbf{D}_{1/2}{}^2|i)^n$ as in (2.2), $(\Omega_{ij})^n$ as in (2.3), $(i|\Omega_{1/2}{}^2|i)^n$ as in (2.4), or $(\Gamma_{ii})^n$ as in (2.5). Indeed the choices based on $(d_{ij})^n$ or $(\Omega_{ij})^n$ have previously been noted[10] as moments of the distribution of distances, and the first of these sequences (up to a slight rescaling) has been emphasized by Randić.[11]

(c) Raise sums of matrix elements to a power, as $(2.1)_i{}^n$ or $(2.3)_i{}^n$, the $n = 1$ sums being mentioned in (3.1). Here the final invariant is obtained after summation on i. The $n = 2$ invariant based on (2.1) has already being considered,[21] as have also closely related indices.[28] The sequences of (d) next also may be interpreted as powers of (special) sums.

(d) Raise matrices to a power, as $\mathbf{D}_1{}^n$ in (2.1), as $\mathbf{D}_{1/2}{}^n$ or $\mathbf{D}_{1/2}{}^{2n}$ in (2.2), as Ω^n in (2.3), as $\Omega_{1/2}{}^n$ in (2.4), or Γ^n in (2.5).

(e) Raise subgraph contributions to a power, any one of the expressions in (3.1 → 9) being raised to a power n (then followed by a summation on subgraphs of the given type).

(f) Raise just a subpart of a subgraph contribution to a power, as just the inner summand in (3.4, 3.6, 3.7, or 3.9), or as just $|g|^{-n}$ in (3.8), or just $\#_2(T-(e))^{-n}$ or $\#_{\text{tree}}(T-(e),G)^{-n}$ in (3.7) or (3.9).

Here there is seen to be a great number of possible choices for sequences, any one of which is a sequence of moments of some sort and might be argued to be of comparable "naturalness". Moreover, though the first members of these sequences give the same invariant for trees (theorem B) and but a few different invariants for connected graphs (theorem A), the consequent sequences now noted seldom turn out generally to represent the same sequence of invariants. With limitation to trees there are some coincidences of different sequences.

Another approach is to consider sequences of coefficients in the secular (or characteristic) equation for suitable matrices. That is, for a matrix $\mathbf{M}$ one considers the sequence

of coefficients of x^{N-n} in the polynomial $\det\{x\mathbf{I}-\mathbf{M}\}$, where **I** is the identity matrix. Suitable choices for the matrix **M** are $(\mathbf{D}_{1/2})^2$ or $(\Omega_{1/2})^2$ or Γ. That the $n = 1$ coefficients are traces of these matrices which in turn give $W(G)$ (for trees) identifies these as sequence extensions—moreover, they are somewhat reminescent of the sequence of coefficients for the choice $\mathrm{M} = \mathbf{D}_1$, as has already been [29] suggested to be meaningful.

A final sequence is that of "*n*-volumina" sums.[12] The $n = 1$ term is a sum over distances between all pairs of vertices of G; the $n = 2$ term is a sum over "areas" among all triples of vertices; the $n = 3$ term is a sum over "volumes" among all quartets of vertices; etc. Of course, one needs to define these *n*-volumina using one of several different choices for the "geometry" associated to the different distance functions noted.

A more limited idea is to extend the Wiener number to but just a single additional invariant rather than a whole sequence. One such proposal by Randić[30] (for trees) involves in some sense extending a summation over contributions of the form (3.4) to a summation over related path contributions, but the result turns out[10] to be closely related to the second member of the $(d_{ij})^n$ sequence noted in (b) above (and is thereby naturally extended beyond trees and beyond just a single additional invariant).

6. INTRASEQUENCE DEPENDENCES

Granted the fair number of possible sequences already noted, questions arise as to their character and interrelations. Two natural mathematical questions are as follows:

•How many independent terms are there in the various sequences?

•Are there linear relations between different sequences?

To address such questions an inner product on the space of invariants is of utility. Thence if X and Y are two invariants, we define

$$\langle X|Y\rangle = \sum_{G}^{\mathcal{S}} X(G)\cdot Y(G) \tag{6.1}$$

where the summation is over some suitable set $\mathcal{S}$ of graphs. Indeed one should perhaps label the inner product by a subscript $\mathcal{S}$, and perhaps $X(G)$ in (6.1) should be complex-conjugated, though we do not do so here. This is essentially just the standard[31] vector-space inner product where $|X\rangle$ is viewed as a vector with components $X(G)$ for each $G \in \mathcal{S}$. An invariant is *normalized* if

$$\langle X|X\rangle = 1 \tag{6.2}$$

which if not satisfied may be accomplished through a rescaling of the invariant.

Possible linear dependences (over $\mathcal{S}$) within a sequence X_n are encoded (in some manner) within the *overlap* (or metric) matrix **S** with elements

$$S_{mn} = \langle X_m|X_n\rangle \tag{6.3}$$

This is non-negative definite, taking a 0 eigenvalue if there is a linear dependence. That is, if for a normalized eigenvector with components φ_n one defines

$$X_\varphi \equiv \sum_n \varphi_n X_n \tag{6.4}$$

then

$$\langle X_\varphi|X_\varphi\rangle = \sum_{m,n} \varphi_m S_{mn} \varphi_n = \lambda_\varphi \tag{6.5}$$

where λ_φ is the eigenvalue. Consequently if $\lambda_\varphi = 0$, then evidently $X_\varphi = 0$, and there is linear dependence amongst the X_n, at least on restriction to the set $\mathcal{S}$. Moreover (for a sequence of normalized invariants) the nearness of the (non-negative) eigenvalues to 0 is a measure of the proper space nearness to linear dependence manifested by the corresponding X_φ. Thus the minimum eigenvalue $\lambda_{\min}$ or $\det(\mathbf{S})$ can serve as an overall measure of nearness to linear dependence (best when the X_n are normalized). Löwdin[32] suggests $\lambda_{\min}$ as a measure, but a more sort of overall comparison is made by the geometric-mean eigenvalue which is given as $\{\det(\mathbf{S})\}^{1/N}$, where N is the number of invariants in $\mathcal{S}$. The type of orthogonalization involved here is much like that of "principal-components" [33] analysis or of "canonical orthogonalization" [32] in quantum chemistry.

A nice "geometric" interpretation can be given to the invariants arising from such an eigen-analysis. For an N-dimensional parallelotope (or hyperparallelapiped) P with edges issuing from one vertex specified by vectors $\mathbf{X}_i$, $i = 1$ to N, it is a standard result[34] that its N-dimensional hypervolume is specified by the determinant of the matrix **X** with columns being the $\mathbf{X}_i$. Thence the square of the volume of parallelotope P is

$$\{\mathrm{vol}(P)\}^2 = \det\mathbf{X}\cdot\det\mathbf{X}^+ = \det(\mathbf{X}\mathbf{X}^+) = \det\mathbf{S} \tag{6.6}$$

where $\mathbf{X}^+$ is the transpose of **X**. Thence $\{\det(\mathbf{S})\}^{1/2N}$ is seen to be the required length for the edge of a hypercube with the same hypervolume as the parallelotope specified by the $\mathbf{X}_i$, $i = 1$ to N. Evidently for (normalized) unit vectors $|\mathrm{vol}(P)| \leq 1$, and the closer to zero $\mathrm{vol}(P)$ is the more nearly (at least some) vectors $\mathbf{X}_i$ point in the same direction. Moreover $\lambda_{\min}^{1/2}$ and $\lambda_{\max}^{1/2}$ seem to be roughly associated to minimal and maximal linear dimensions of P. The smaller $\lambda_{\min}^{1/2}$ is, the more "squashed" the parallelotope is. It seems $\lambda_{\max}^{1/2}$ is related to a maximum length internal to P, and indeed approaches $\sqrt{M}$ when there are M very nearly parallel (normalized) vectors in the set of $\mathbf{X}_i$ (while all the remaining are nearly orthogonal). That is, we view $\lambda_{\max}$ as a measure of the effective (maximum) number of mutually aligned vectors $\mathbf{X}_i$.

As an explicit numerical illustration of how to address this problem concerning dependences within a sequence, we restrict attention to the set of 18 octane trees and evaluate our linear-dependence measures for several of the sequences of the preceding section. The different sequences used are indicated in Table 1, where also the unnormalized sequence-member values for a representative octane is given. The notation $Q(p)$ for the pth of these sequences extended to different lengths N is then used in reporting the linear-dependence measures reported in Table 2. Notably from consideration of $\lambda_{\min}$ or $\det(\mathbf{S})$ one sees that a couple of the sequences of invariants seem to exhibit "early" linear

Table 1. Sequences of Invariants for 2,2,3,3-Tetramethylbutane

n	1	2	3	4	5	6
Q(1): $\{W(G)\}^n$	58	3364	195112	11316496	656356768	38068692544
Q(2): $\sum_{i,j}(d_{ij})^n$	116	272	692	1856	5156	14672
Q(3): $\sum_i(\sum_j d_{ij})^n$	116	1736	26576	413216	6491456	102663296
Q(4): $\sum_i(\Gamma_{ii})^n$	7	7	8	9	9	10
Q(5): $\mathrm{tr}(\mathbf{D}_{1/2}{}^{2n})$	116	9961	984422	97878117	9735156774	968301464697
Q(6): $\mathrm{tr}(\mathbf{\Gamma}^n)$	7	12	27	68	183	510
Q(7): $\sum_i \#_2(G-i)^n$	58	478	6778	101278	1518778	22781278
Q(8): $\sum_e \#_2(G-(e))^n$	58	550	6154	79942	1149418	17483100

dependence (over the current set S of octanes), one after $N = 4$ terms and one after six terms—but all the sequences exhibit linear dependence after eight terms. Evidently the two which show early linear dependence are less "complete", and might *a priori* be less favored, though one of these sequences is as earlier utilized.[11] Of course up to the point of linear dependence each sequence is generally expected to span a different space. Indeed though a sequence $Q(r)$ maintains linear independence beyond that of another $Q(p)$, the smaller space spanned by $Q(p)$ generally is not a subspace of that of the larger space spanned by $Q(r)$, so that for certain properties the shorter sequence can be superior. Of the several sequences tested, that of $Q(6)$ based on traces of powers of the inverse Laplacian matrix Γ shows the least rapid approach towards linear dependence. Interestingly the effective number $\lambda_{\max}$ of mutually aligned elements in a sequence is quite high for these sequences—it is almost always close to the number of members in a sequence—this thereby also indicating significant approach toward linear dependence in these sequences.

7. DEPENDENCES BETWEEN SEQUENCES

To compare different sequences of invariants, say X_m and Y_n, $m,n = 1$ to N, there are at least a couple possible approaches. One approach[35] makes use of the idea of "posets" and yields a distance between the two sequences. But here a second approach using ideas like those of the preceding section is used—in particular the vector-space view with the inner product of (6.1) is retained.

Comparisons between two sequences X_m and Y_n are naturally encoded (in some manner) within the *intersequence overlap* matrix $\mathbf{R}$ with elements

$$R_{mn} = \langle X_m | Y_n \rangle \tag{7.1}$$

The mth row of $\mathbf{R}$ can be viewed as representation coefficients for X_m in terms of the Y-sequence of invariants, and dually the columns can be viewed as representation coefficients for Y-invariants in terms of X-sequence invariants. Making normalization-preserving linear transformations on rows and columns leads to a new matrix with the same interpretations for the corresponding new X- and Y-sequence invariants. Thence it is natural to make the double unitary transformations on rows and columns bringing $\mathbf{R}$ to diagonal form. The consequent diagonal elements are arranged to be non-negative, are termed[36] *singular values* of $\mathbf{R}$, and are measures of linear dependence between the two sequences. Such singular values are conveniently obtained as (positive) square roots of the eigenvalues of $\mathbf{R}^+\mathbf{R}$ (or equivalently of $\mathbf{R}\mathbf{R}^+$). This singular-value decomposition is of the form

$$\mathbf{R} = \mathbf{U}^+\Delta\mathbf{V} \tag{7.2}$$

with $\Delta_{ij} = \rho_i\delta_{ij}$ and may be accomplished through the (conventional) diagonalization of $\mathbf{R}^+\mathbf{R}$ and $\mathbf{R}\mathbf{R}^+$. Indeed $\mathbf{R}^+\mathbf{R}$ and $\mathbf{R}\mathbf{R}^+$ have the same eigenvalues which are ρ_i^2, and the unitary matrices accomplishing these two diagonalizations (with the eigenvalues appearing in the diagonal form Δ^2 being like-ordered) are $\mathbf{U}$ and $\mathbf{V}$, as evidenced in the relations

$$\mathbf{U}(\mathbf{R}\mathbf{R}^+)\mathbf{U}^+ = \Delta^2 = \mathbf{V}(\mathbf{R}^+\mathbf{R})\mathbf{V}^+ \tag{7.3}$$

The transformed X and Y variables

$$X_i^U \equiv \sum_j U_{ji}X_j,\ Y_i^V \equiv \sum_j V_{ji}Y_j \tag{7.4}$$

then are *biorthogonal*

$$\langle X_i^U | X_j^V \rangle = (\mathbf{U}\mathbf{R}\mathbf{V}^+)_{ij} = \delta_{ij}\rho_i \tag{7.5}$$

That is, there is correlation only between corresponding invariants in each sequence. Indeed

$$r_{ij} = \delta_{ij}\rho_i/\langle X_i^U | X_i^U \rangle^{1/2}\langle Y_i^V | Y_i^V \rangle^{1/2} \tag{7.6}$$

may be viewed as statistical correlation coefficients between the invariants of the two sequences. The arithmetic mean value r of these diagonal elements r_{ij} is a mean measure of linear dependence.

As an explicit numerical illustration addressing the question concerning dependences between sequences we again restrict attention to octane trees and evaluate the linear-dependence measure r between several pairs of sequences. The results in matrix form are shown in Table 3 for the same eight sequences of invariants considered in Tables 1 and 2. In Table 3 we report the lower off-diagonal parts of the **r**-matrices for these different sequences of invariants for different choices of the number N of terms retained in the sequences. Evidently there is a high degree of correlation between these different sequences. The correlation for $Q(8)$ becomes indefinite as N increases past 4 since this sequence itself contains no more that 4 independent invariants (so that the corresponding individual r-values between $Q(8)$ and any other sequences involve ratios of $\rho_i = 0$ to $\langle X_i^U | X_i^U \rangle^{1/2} = 0$).

8. EXTENSIONS BY SUBGRAPHS

There is a rather different type of alternative to the linearly sequenced extensions of section 5. This involves extending the subgraph summations of section 3 to include summations over other types of subgraphs. In particular we imagine the K_1- or K_2-type subgraphs in (3.1 → 7) replaced by other types of subgraphs isomorphic to some graph other than K_1 or K_2. There then arises a choice as to the different types of subgraphs to be admitted. Classes of such subgraphs, for

Table 2. Linear-Dependence Measures for Sequences

N	2	3	4	5	6	7	8
				$(\lambda_{min})^{1/2}$			
Q(1)	6.44×10^{-2}	4.09×10^{-3}	2.37×10^{-4}	1.11×10^{-5}	4.34×10^{-7}	1.48×10^{-8}	3.74×10^{-10}
Q(2)	9.35×10^{-2}	8.95×10^{-3}	1.25×10^{-3}	1.90×10^{-4}	2.34×10^{-5}		
Q(3)	6.60×10^{-2}	4.44×10^{-3}	3.04×10^{-4}	4.67×10^{-5}	5.93×10^{-6}	1.83×10^{-6}	1.99×10^{-7}
Q(4)	7.05×10^{-2}	5.93×10^{-3}	8.59×10^{-4}	1.80×10^{-4}	6.22×10^{-5}	1.17×10^{-5}	2.07×10^{-6}
Q(5)	4.93×10^{-2}	2.44×10^{-3}	2.73×10^{-4}	2.37×10^{-4}	3.22×10^{-5}	3.42×10^{-6}	2.00×10^{-7}
Q(6)	1.79×10^{-1}	2.51×10^{-2}	7.08×10^{-3}	1.07×10^{-3}	1.54×10^{-4}	3.57×10^{-5}	4.85×10^{-6}
Q(7)	3.08×10^{-2}	2.15×10^{-2}	3.63×10^{-3}	5.83×10^{-4}	1.56×10^{-4}	1.24×10^{-5}	7.65×10^{-7}
Q(8)	5.98×10^{-2}	9.20×10^{-3}	3.86×10^{-4}				
				$\{\det(\mathbf{S})\}^{1/2N}$			
Q(1)	3.02×10^{-1}	9.69×10^{-2}	3.11×10^{-2}	9.55×10^{-3}	2.79×10^{-3}	7.79×10^{-4}	2.03×10^{-4}
Q(2)	3.63×10^{-1}	1.45×10^{-1}	6.53×10^{-2}	3.03×10^{-2}	1.39×10^{-2}		
Q(3)	3.05×10^{-1}	1.01×10^{-1}	3.47×10^{-2}	1.42×10^{-2}	6.00×10^{-3}	2.88×10^{-3}	1.31×10^{-3}
Q(4)	3.15×10^{-1}	1.16×10^{-1}	5.07×10^{-2}	2.43×10^{-2}	1.31×10^{-2}	6.97×10^{-3}	3.63×10^{-3}
Q(5)	2.64×10^{-1}	7.54×10^{-2}	2.77×10^{-2}	1.46×10^{-2}	7.09×10^{-3}	3.26×10^{-3}	1.35×10^{-3}
Q(6)	5.02×10^{-1}	2.45×10^{-1}	1.37×10^{-1}	6.92×10^{-2}	3.41×10^{-2}	1.75×10^{-2}	8.63×10^{-3}
Q(7)	2.09×10^{-1}	1.51×10^{-1}	8.18×10^{-2}	4.23×10^{-2}	2.29×10^{-2}	1.10×10^{-2}	4.77×10^{-3}
Q(8)	2.91×10^{-1}	1.20×10^{-1}	3.95×10^{-2}				
				λ_{max}			
Q(1)	1.996	2.983	3.959	4.920	5.863	6.787	7.693
Q(2)	1.991	2.960	3.895	4.793	5.656	6.490	7.302
Q(3)	1.996	2.982	3.954	4.906	5.835	6.739	7.618
Q(4)	1.995	2.977	3.937	4.870	5.774	6.653	7.512
Q(5)	1.998	2.990	3.973	4.945	5.904	6.848	7.775
Q(6)	1.968	2.881	3.755	4.607	5.448	6.283	7.118
Q(7)	1.999	2.991	3.967	4.918	5.839	6.729	7.590
Q(8)	1.996	2.988	3.976	4.960	5.942	6.922	7.564

the most part previously considered[13] (in a general "cluster-expansion" context), are as follows:

(a) the class $\mathcal{C}_{conn}(G)$ of all connected subgraphs of G;

(b) the class $\mathcal{C}_{ind}(G)$ of all connected *induced* subgraphs of G, an induced subgraph G' having all edges $\{i,j\} \in E(G)$ for which $i,j \in V(G')$;

(c) the class $\mathcal{C}_{wcvx}(G)$ of all *weakly convex* subgraphs of G, such a subgraph G' including a geodisic g of G between every pair of points of G';

(d) the class $\mathcal{C}_{cvx}(G)$ of all (strongly) *convex* subgraphs of G, such a subgraph G' including all geodesics of G between pairs of points of G'. Of course, these last two classes generally depend on the choice of distance function. In these various cases one obtains a set of invariants $W(G',G)$ with $G' \in \mathcal{C}_*(G)$ (for $* =$ conn, ind, wcvx, cvx) such that $W(K_1,G)$ or $W(K_2,G)$ is $W(G)$ (possibly up to some trivial factor, such as 2). For $G' \notin \mathcal{C}_*(G)$ it is understood that $W(G',G) = 0$. Rather than obtaining a sequence of graph invariants for G, one obtains instead a partially ordered set of invariants (with the partial ordering being that of subgraph inclusion).

Several of the choices for invariants $W(G',G)$ with $G' \in \mathcal{C}_*(G)$ may be indicated more explicitly. First generalization of the vertex invariant sums indicated in (3.1) we have

$$\sum_i^{G'} (2.1)_i = \sum_i^{G'} (\sum_j^{G} d_{ij}) \tag{8.1}$$

and

$$\sum_i^{G'} (2.2)_i = \sum_i^{G'} (i|\mathbf{D}_{1/2}{}^2|i) \tag{8.2}$$

as well as analogous expressions with $\mathbf{D}$ replaced by $\mathbf{\Omega}$. The expression of (3.2) or (3.4) can be extended to

$$\#_2(G-G') \quad \text{or} \quad \#_2(G-(G')) \tag{8.3}$$

while for (3.5) one might consider

$$\#_{near}(G',G) \equiv \prod_i^{G'} \#_i(G',G) \tag{8.4}$$

and for (3.6)

$$\sum_g^{G} |E(g) \cap E(G')|/\#_g(G) \tag{8.5}$$

There are many more possibilities.

Notably the subgraph extensions of this section allow a faithful linear expansion of an arbitrary invariant in terms of the invariants of any one of the partially ordered sets identified. That is, the partially ordered sets of invariants in some sense act as linear "bases" for all other invariants. We summarize this as follows:

Theorem C—For connected graphs there is an expansion for a general invariant $X(G)$:

$$X(G) = \sum_{G'}^{G} \mathfrak{c}_X(G') \cdot W(G';G)$$

where the summation is over all subgraphs of a class $\mathcal{C}_*(G)$ identified above, and the expansion coefficients $c_X(G')$ depend only on X and G'. These coefficients are given as

$$\mathfrak{c}_X(G') = \sum_{G''}^{G'} X(G'') \cdot W^{-}(G'',G')$$

where the *W-Möbius function* W^- is obtainable as the matrix

Table 3. Sequence Interdependence Measures

	Q(1)	Q(2)	Q(3)	Q(4)	Q(5)	Q(6)	Q(7)
				$N = 2$			
Q(2)	0.99905						
Q(3)	0.99961	0.99823					
Q(4)	0.99566	0.99329	0.99787				
Q(5)	0.99913	0.99660	0.99963	0.99775			
Q(6)	0.98895	0.99388	0.98591	0.97588	0.98229		
Q(7)	0.82416	0.80677	0.83698	0.86476	0.84516	0.74958	
Q(8)	0.98478	0.98679	0.98044	0.96739	0.97905	0.98928	0.74448
				$N = 3$			
Q(2)	0.99242						
Q(3)	0.98261	0.98038					
Q(4)	0.94183	0.95404	0.98308				
Q(5)	0.99676	0.98675	0.98225	0.94747			
Q(6)	0.95913	0.97447	0.91688	0.87224	0.94245		
Q(7)	0.70580	0.70832	0.65480	0.66975	0.71914	0.74283	
Q(8)	0.70583	0.71984	0.65807	0.66115	0.70957	0.78951	0.93470
				$N = 4$			
Q(2)	0.92733						
Q(3)	0.89612	0.76780					
Q(4)	0.75650	0.78142	0.92244				
Q(5)	0.83881	0.80341	0.90274	0.89850			
Q(6)	0.85691	0.95859	0.75915	0.79357	0.81141		
Q(7)	0.53128	0.50863	0.49613	0.49172	0.54779	0.53895	
Q(8)	0.53351	0.53248	0.47855	0.61918	0.57389	0.57344	0.74355
				$N = 5$			
Q(2)	0.83077						
Q(3)	0.74178	0.70560					
Q(4)	0.65133	0.68296	0.85304				
Q(5)	0.82799	0.83479	0.81129	0.86472			
Q(6)	0.77138	0.93074	0.60440	0.65158	0.74354		
Q(7)	0.42728	0.37027	0.43813	0.39870	0.49584	0.48125	
Q(8)	–	–	–	–	–	–	–
				$N = 6$			
Q(2)	0.74024						
Q(3)	0.63401	0.66444					
Q(4)	0.56681	0.60763	0.80516				
Q(5)	0.72329	0.71014	0.72540	0.71477			
Q(6)	0.70730	0.84056	0.54276	0.53432	0.62692		
Q(7)	0.33863	0.29646	0.33446	0.31705	0.38713	0.37928	
Q(8)	–	–	–	–	–	–	–

inverse of the matrix with (G'',G')th element $W(G'',G')$, for G'', $G' \in C_*(G)$.

Proof: The sets $C_*(G)$ each contain G, so that the first expansion of the theorem contains at least one term (the $G' = G$ term) contained in no smaller subgraph of $C_*(G)$, and as a consequence can be used to eliminate whatever error there might be after limiting the expansion to all terms excepting this last $G' = G$ term. That is, one can choose the coefficients in a hierarchical fashion building them up from smaller values, thusly

$$c_X(G') = \sum_{G'}^{G}{}' c_X(G') \cdot W(G',G)/W(G,G) \qquad (8.6)$$

where now the summation goes over all $G' \in C_*(G)$ excepting $G' = G$. Iteration of this downward ultimately yields a formula for the W-Möbius function appearing in the second expansion of the theorem. To do this it is crucial to have a nonsingular W-matrix as we do since it is upper triangular and all the diagonal elements $W(G',G')$ are nonzero.

The ideas in this theorem have in fact been often stated for the case that $W(G',G)$ is replaced by the number $\#(G',G)$ of subgraphs of type G' in G and also $C_*(G)$ is identified to the case of all connected[13,15,16] or induced[14] graphs. An elegant theory[14,15,17,37] of Möbius functions applies. The inclusion of the possibility of different atom types[16,17] is straightforward. For example one might express heats of atomization as sums of atom types, bond types, adjacent bond-pair types, etc. (though in this case because of the definition of energy 0 the atom type contributions are 0). The more general idea for general (nonsingular upper-triangular) $f(G',G)$ in place of $W(G',G)$ has also been described in both chemical[17] and mathematical[37] contexts. The theorem C noted here of course answers for the present partially ordered sets the independence problem addressed in the preceding section in connection with Table 1 for sequences.

For the theorem to be most useful the question of rapidity of convergence becomes crucial and is nontrivial. For some cases as the bond-energy-related example the chemical intuition involved with the idea of "locality" of interactions aids the motivation for development in terms of $\#(G',G)$. More generally one can look[17] at different "scaling" related properties of the invariants involved in the expansion. Often it may prove chemically advantageous to rescale the $W(G',G)$ by a factor $|G|^{-m}$ with m independent of G—i.e., to make the rescaled quantity bounded by G-independent values. But to implement such ideas for the present Wiener-number-related expansions we leave to future work.

9. PROSPECT

Evidently there is some leeway in extending Wiener's definitions beyond the case of trees and perhaps even more leeway in extending Wiener numbers in chemically plausable ways to whole sets of invariants. Here first steps in a more comprehensive study of these extensions have been taken. Rigorous results (some being new) relating different extensions are found in theorems A and B, and some results concerning characteristics of sequences or sets of "hyper-Wiener" indices are described rigorously in theorem C or empirically indicated in connection with Tables 2 and 3 in sections 6 and 7.

ACKNOWLEDGMENT

The authors acknowledge support from the Welch Foundation of Houston, Texas.

REFERENCES AND NOTES

(1) The idea seems to have been fairly wide spread, but one nice reference is Brown, Crum On the Application of Mathematics to Chemistry. *Proc. R. Soc. Edin.* **1866**−**9**, *6*, 89.

(2) See, e.g.: Sylvester, J. J. On an Application of the New Atomic Theory to the Graphical Representation of the Invariants and Covariants of Binary Quantics. *Am. J. Math.* **1878**, *1*, 64−125.

(3) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1948**, *69*, 17−20.

(4) (a) Polansky, O. E.; Bonchev, D. The Wiener Number of Graphs. I. *Commun. Math. Chem.* **1986**, *21*, 133−186. (b) Rouvray, D. E. The Role of the Topological Distance Matrix in Chemistry. In *Mathematical and Computational Concepts in Chemistry*; Trinajstić, N., Ed.; Ellis Horwood: Chichester, England, 1986; pp 295−306. (c) Gutman, I.; Yeh, Y.-N.; Lee, S.-L.; Luo, Y.-L. Some Recent Results in the Theory of the Wiener Number. *Indian J. Chem.* **1993**, *32A*, 651−661. (d) Nikolic, S.; Trinajstić, N.; Mihalic, Z. The Wiener Index: Development and Applications. *Croat. Chem. Acta* **1995**, *68*, 105−129.

(5) (a) Entringer, R. C.; Jackson, D. E. Distance in Graphs. *Czech. Math. J.* **1976**, *69*, 283−296. (b) Plesnik, J. On the Sum of All Distances in a Graph on Digraph. *J. Graph Th.* **1984**, *8*, 1−21. (c) Chung, F. R. K. The Average Distance and the Independence Number. *J. Graph Th.*

1988, *12*, 229−235. (d) Mohar, B. Eigenvalues, Diameter, and Mean Distance in Graphs. *Graphs Comb.* **1991**, *7*, 53−64.
(6) Doyle, J. K.; Garver, J. E. Mean Distance in a Graph. *Disc. Math.* **1977**, *17*, 147−154.
(7) (a) Merris, R. An Edge Version of the Matrix-Tree Theorem and the Wiener Index. *Lin., Multilin. Alg.* **1989**, *25*, 291−296. (b) Mohar, B.; Babic, D.; Trinajstić, N. A Novel Definition of the Wiener Index for Trees. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 153−154.
(8) Klein, D. J.; Randić, M. Resistance Distance. *J. Math. Chem.* **1993**, *12*, 81−95.
(9) Gutman, I. A Formula for the Wiener Number of Trees and Its Extension to Graphs Containing Cycles. *Graph Theory Notes, NYAS* **1994**, *27*, 9−15.
(10) Klein, D. J.; Lukovits, I.; Gutman, I. On the Definition of the Hyper-Wiener Index for Cycle-Containing Structures. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 50−52.
(11) Randić, M. Molecular Shape Profiles. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 373−382.
(12) Klein, D. J.; Zhu, H.-Y. Distances and Volumina for Graphs. *Disc. Math.* Submitted for publication.
(13) (a) Smolenski, E. A. Application of the Theory of Graphs to Calculations of the Additive Structural Properties of Hydrocarbons. *Russ. J. Phys. Chem.* **1964**, *38*, 700−704. (b) Gordon, M.; Kennedy, J. M. The Graph-Like State of Matter. Part 2. *J. Chem. Soc., Faraday Trans. II* **1973**, *69*, 484−504.
(14) Essam, J. W.; Fisher, M. E. Some Basic Definitions in Graph Theory. *Rev. Mod. Phys.* **1970**, *42*, 271−283.
(15) (a) Essam, J. W.; Kennedy, J. W.; Gordon, M; Whittle, P. The Graph-Like State of Matter. Part 8. *J. Chem. Soc., Faraday Trans. II* **1977**, *73*, 1289−1301. (b) Kennedy, J. W.; Gordon, M. Graph Contractions and a Generalized Möbius Inversion. *Ann. N. Y. Acad. Sci.* **1979**, *319*, 331−348.
(16) (a) Skvortsova, M. I.; Baskin, I. I.; Slovokhotova, O. L.; Zefirov, N. S. Methodology of Constructing a General Model for the Structure-Property Relationship at the Topological Level. *Doklady Acad. Nauk.* **1994**, 336−499. (b) Baskin, I. I.; Skvortsova, M. I.; Stankevich, I. V.; Zefirov, N. S. On the Basis of Invariants of Labelled Molecular Graphs. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 527−531.
(17) Klein, D. J. Chemical Graph-Theoretic Cluster-Expansions. *Int. J. Quantum Chem.* **1986**, *S20*, 153−171.
(18) (a) Harary, F. *Graph Theory*; Addison-Wesley: Reading, MA, 1972. (b) Trinajstić, N. *Chemical Graph Theory*; CRC: Boca Raton, FL, 1992.
(19) Blumenthal, L. M. New Theorems and Methods in Determinant Theory. *Duke Math. J.* **1936**, *2*, 396−404.
(20) Mohar, B. The Laplacian Spectrum of Graphs. In *Graph Theory, Combinatorics, and Applications*; Alavi, Y., Chartrand, C., Ollermann, O. R., Schwenk, A. J., Eds.; John Wiley & Sons: New York, 1991; pp 871−898.
(21) Balaban, A. T. Topological Indices Based on Topological Distances in Molecular Graphs. *Pure Appl. Chem.* **1983**, *55*, 199−206.
(22) Miller, Z. Medians and Distance Sequences in Graphs. *Ars. Comb.* **1983**, *15*, 169−177.
(23) Balaban, A. T.; Diudea, M. V. Real Number Vertex Invariants. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 421−428.
(24) Khadikar, P. V.; Deshpande, N. V.; Kale, P. P.; Dobrynin, A.; Gutman, I.; Dömötör, G. The Szeged Index and Analogy with the Wiener Index. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 547−550.
(25) (a) Amić, D; Trinajstić, N. On the Detour Matrix. *Croat. Chim. Acta* **1995**, *68*, 53−62. (b) Lukovits, I. The Detour Index. *J. Chem. Inf. Comput. Sci.* Submitted for publication.
(26) (a) Lukovits, I. Wiener Indices and Partition Coefficients of Unsaturated Hydrocarbons. *Quant. Struct.-Act. Relat.* **1990**, *9*, 227−231. (b) Juvan, M.; Mohar, B. Bond Contributions to the Wiener Index. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 217−219.
(27) We suspect this result is "implicit" in a number of electrical engineering texts, but for an explicit proof, see: Shapiro, L. W. An Electrical Lemma. *Math. Mag.* **1987**, *60*, 36−38.
(28) (a) Gordon, M.; Scantlebury, G. R. Non-Random Polycondensation Statistical Theory of the Substitution Effect. *Trans. Faraday Soc.* **1964**, *60*, 604−621. (b) Gutman, I.; Ruscic, B.; Trinajstić, N.; Wilcox, C. F., Jr. Graph Theory and Molecular Orbitals. XII. Acyclic Polyenes. *J. Chem. Phys.* **1975**, *62*, 3399−3405.
(29) Hosoya, H. On Some Counting Polynomials in Chemistry. *Disc. Appl. Math.* **1988**, *19*, 239−257.
(30) Randić, M. Novel Molecular Description for Structure-Property Studies. *Chem. Phys. Lett.* **1993**, *211*, 478−483.
(31) See, e.g.: Halmos, P. R. *Finite-dimensional Vector Spaces*; Litton Ed. Pub., 1958; and Springer-Verlag: Berlin, 1974.
(32) Löwdin, P. O. On the Nonorthogonality Problem. *Adv. Quantum Chem.* **1970**, *5*, 185−199.
(33) Hotelling, H. Analysis of A Complex of Statistical Variables into Principle Components. *J. Ed. Psych.* **1933**, *24*, 417−441.
(34) See, e.g.: section 5.9 of Yale, P. B. *Geometry and Symmetry*; Dover Pub.: New York, 1988.
(35) Klein, D. J. Similarity and Dissimilarity in Posets. *J. Math. Chem.* In press.
(36) Marcus, M.; Minc, H. *A Survey of Matrix Theory and Matrix Inequalities*; Dover Pub.: New York, 1992. (Originally published by Prindle-Weber & Schmidt: Boston, 1964.)
(37) Rota, G.-C. On the Foundations of Combinatorial Theory. I. *Zeit. Wahr. Verw. Geb.* **1964**, *2*, 340−368.

CI950116S