Graph-Geometric Invariants for Molecular Structures

H.-Y. Zhu and D. J. Klein*

Texas A&M University at Galveston, Galveston, Texas 77553-1675

Received February 22, 1996[⊗]

The shortest-path distance on graphs might naturally be imagined to be implicated in some sort of innate geometric structure on graphs, such as are oft used to represent molecular structures. In fact it is noted here that in addition to this shortest-path metric, several other metrics may be naturally defined on graphs, sometimes there being more formal similarities with the Euclidean metric, such as can also be associated with a graph once it is embedded in a Euclidean space as is ultimately the case with molecular structures. Five natural purely graph-theoretic metrics are briefly discussed and are utilized in conjunction with several novel graph invariants defined in correspondence with classical geometric invariants for structures in Euclidean space. These graph invariants thereby attain potentially useful geometrical interpretations in terms of curvatures, torsions, and volumina. Such so-interpretable "graph-geometric" invariants then are evaluated for graphs in several different types of sequences, and the results are compared with one another and the corresponding Euclidean-geometric quantities for simple embeddings of these graphs in three-dimensional Euclidean space. It is thence suggested that certain metrics (especially our "quasi-Euclidean" and "resistance-distance" metrics) form a more suitable foundation for a "graph geometry".

1. GRAPHS, MOLECULES, AND GEOMETRY

During the last couple of decades there has been an enthusiastic development of the idea that mathematical graphs are the natural correspondent to the molecular valence structures of chemistry—much of the work developing this idea being summarized in Trinajstić's book[1] "*Chemical Graph Theory*". But the graph-theoretic labeling for a molecule leaves out much of the detail of the molecular geometry. Thence questions naturally arise: How much molecular geometric detail might be reflected in a molecular graph? Or more generally, is there any such thing as an intrinsic "geometry of graphs"? And if so does it have chemical relevance? Such questions are somewhat ambiguous but not necessarily devoid of significant meaning.

Here this question concerning graphs and geometry is to be addressed. In a fundamental sense some part of geometry arises[2] once one has a metric (or distance) function on a set, which in classical cases is infinite and embued with additional "ordering" and continuity features, though the consequent geometry may be non-Euclidean. Within the usual view[1] for graphs the sets are discrete, so that the geometry is more primitive than classically. The crucial initial step then is to set a metric, in a natural way—indeed for a particular choice of metric this has already been done and rather well-studied as reviewed in the book[3] "*Distance in Graphs*" by Buckley and Harary. This particular graph metric we term the "shortest-path" metric with the anticipation that there may be other natural graph metrics, some examples of which we have previously noted.[4,5] The general idea of a metric and five of these various natural metrics on graphs are discussed in section 2.

That a "metric" implicates some sort of innate geometric structure is also indicated in the alternative name of "distance function", and the realization of such a "graph geometric" structure then brings to the fore the relevance of the geometry of graphs and its comparison to the Euclidean geometry for actual molecules, corresponding to graphs embedded in three-dimensional Euclidean space. To initiate such a study and the associated comparisons there are introduced in sections 3, 4, 5, and 6 several novel graph invariants designed (at least potentially) to have geometric interpretation. Several of these *graph-geometric* invariants are built up from local quantities which may be interpreted as some sort of local (linear) curvature (in section 3), curvilinear torsion (in section 4), or local Gaussian curvature (in section 5). These local quantities each are summed over all the local regions (sites or bonds) to obtain overall invariants for each graph. In section 6 we consider a sequence of "volumina-power-sum" invariants—the summands for this sequence are a power of the distances between graph vertices (for the first member), the areas between triples of vertices (for the second member), the volumes between quartets of vertices (for the third member), etc. Here the simple distance-sum invariant (involving the first power) is essentially the now widely recognized[1,6] invariant of Wiener[7] if the shortest-distance metric is used, though the possibility of using other metrics has been noted[4,5,8,9] as has[5] also the possibility of extending this idea to a sequence of sums of areas, of volumes, etc. In general, any previous invariants developed in terms of the shortest-distance metric might also be developed in terms of any of our other four natural graph-theoretic metrics.

For each type of invariant the ideas are illustrated in application to five different types of sequences (or sets) of chemically interesting graphs. These representative sets of graphs include: n-bond linear chains L_n, n-bond tree-stars S_n, n-bond cycles C_n, m-circumbenzenes, $C_6^{[m]}$, and icosahedral-symmetry fullerenes $F_{h,k}$. Examples of such graphs are presented in Figure 1. What the sequences of chains, stars, and cycles are is clear enough, but the specification of the last two classes of graphs should be expanded upon. The mth circumbenzene is obtained by circumscribing a ring of hexagons around the $(m-1)$th circumbenzene, with the $m = 0$ member being the single hexagon of benzene and the $m = 1$ member being coronene. The mth circumbenzene

[⊗] Abstract published in *Advance ACS Abstracts,* October 1, 1996.

S0095-2338(96)00025-X CCC: $12.00

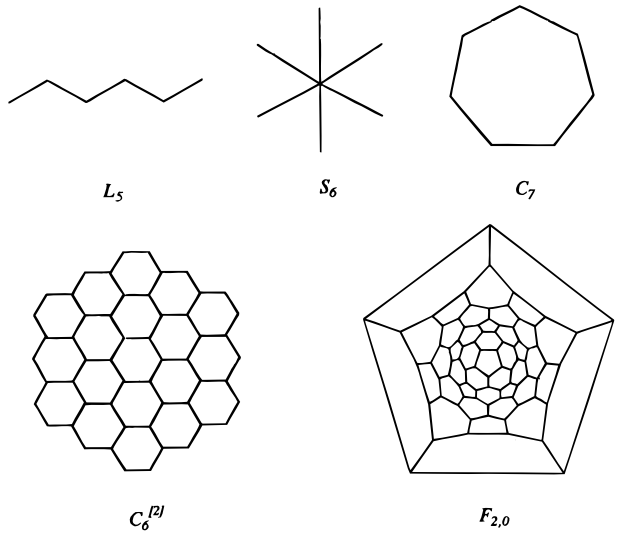

Figure 1. Examples of graphs from each of the five classes considered.

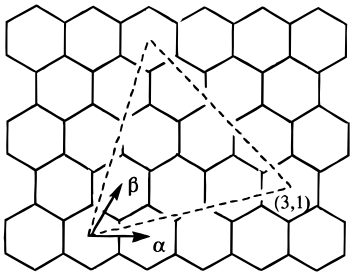

Figure 2. An equilateral triangular *h,k*-patch for the case that $h = 3$, $k = 1$.

has $6(m+1)^2$ sites and $3(m+1)(3m+2)$ edges. The *h,k*-icosafullerene graphs may be conveniently obtained[10] by gluing suitable equilateral-triangular patches cut from the honeycomb lattice onto the faces of (a suitably sized) icosahedron. An *h,k*-patch has its three apices in the centers of hexagons on the underlying honeycomb lattice, and if one of these apices is chosen as the origin with a skew α,β-coordinate system with axes along the lines of centers of adjacent hexagons, then the coordinates of the second apex of the patch is (h,k). As an example see Figure 2. An *h,k*-icosafullerene has $20\cdot(h^2 + hk + k^2)$ vertices, $30\cdot(h^2 + hk + k^2)$ edges, and $2 + 10\cdot(h^2 + hk + k^2)$ "faces".

Beyond the comparison of the different results with different metrics for different graphs we also make comparisons using the standard Euclidean metric as applied to certain presumed "standard" embeddings of the graphs in Euclidean space. Section 7 presents some overall synopsis and discussion of the interpretation and possible utility of the graph-geometric invariants introduced. It is found that the graph invariants based on the four new metrics considered here tend to mimic more closely the associated Euclidean geometric quantities (for standard graph embeddings). The final section 8 makes some further brief comment on the potential for chemical applications.

The graph-theoretic nomenclature is standard[1] and often too so is the metric nomenclature.[2] A graph G is specified first by its set $V = V(G)$ of vertices (corresponding to atomic centers) and second its set $E = E(G)$ of edges, each edge consisting of an unordered pair of (neighbor) vertices. Throughout the present manuscript all molecular graphs are presumed to be *connected* (in the sense that there is presumed to be a path of edges of G between every pair of vertices of G).

2. METRICS AND GRAPHS

Because of the central importance of the "metric" in setting a geometry it is appropriate to begin with a careful definition of this concept (such as also discussed generally in Blumenthal's book[2] on *Distance Geometry*). A metric (or distance function) on G is a function from the Cartesian product $V \times V$ to the non-negative real numbers such that for any $i,j,k \in V$

$$\rho(i,j) = 0 \Leftrightarrow i = j$$

$$\rho(i,j) = \rho(j,i) \geq 0$$

$$\rho(i,j) + \rho(j,k) \geq \rho(i,k)$$

This definition generally is viewed to extend to rather general sets V of points, say for the points of a plane P, whence the classical plane Euclidean geometry has a (Euclidean) metric $\rho_E(i,j) \equiv \{(x_i - x_j)^2 + (y_i - y_j)^2\}^{1/2}$ with (x_i,y_i) and (x_j,y_j) being Cartesian coordinates for points $i,j \in P$. This is in fact just one of several possible metrics even for points in a plane: e.g., the functions $\rho_{P-\alpha}(i,j) \equiv \{|x_i - x_j|^\alpha + |y_i - y_j|^\alpha\}^{1/\alpha}$ are (Minkowski) metrics for each α so long as $\alpha \geq 1$.

The three defining conditions for a metric on G can be interpreted to respectively say that each singlet, doublet, and triplet of points must be *isometrically* (or *faithfully*) embeddable into zero-, one-, and two-dimensional Euclidean spaces, in the sense that the Euclidean distances in the embedding match those of ρ. Thence we might term[5] these conditions the zero-, one-, and two-*Euclidean conditions* with the prefix on Euclidean indicating the dimension of the space into which the embeddings are made—though this is not standard nomenclature. But now it is clear that there is a whole hierarchy of conceivable *m-Euclidean* conditions, which (for $m \geq 3$) may be used to distinguish between qualitatively different metrics. A little more formally we say that ρ is m-Euclidean if for every subset of $m + 1$ vertices there is a faithful embedding into m-dimensional Euclidean space E_m.

(a) The Shortest-Path Metric *d*. First, the standard[1,2] graph metric is the *shortest-path* metric. This gives $\rho(i,j)$ as the minimum number $d(i,j)$ of steps along a path from i to j in G. In the bulk of previous graph-theoretic work this has almost always been the only metric recognized—e.g., virtually no other metrics appear in the review of Buckley and Harary[3] aside from the simple decoration of weighting each of the edges in G to give bond steps of different lengths (and consequent path lengths given as sums of the weights for the constituent bonds). This metric has been commonly denoted by $d(i,j)$, and this convention is followed here. Though conventional this graph metric is not generally three-Euclidean. For instance, for a isometric E_3-embedding of the four-cycle C_4 with vertices a,b,c,d successively located around the cycle $d(a,c) = 2$, while $d(a,b) = d(b,c) = 1$ so that for a faithful Euclidean embedding b would be midway on the line segment between a and c—but likewise d would be midway between a and c on this same line segment, so that in E_3 b and d would be located at the same position, contrary to the condition that $d(b,d) = 2$. Indeed one sees that for the shortest-path metric there is no dimension n for

which C_4 can be faithfully embedded in E_n—moreover, there are no faithful embeddings of C_4 in many of the 19th century non-Euclidean geometries. Perhaps such "poor" behavior has been responsible for the dimming of any search for an intrinsic "geometry of graphs".

(b) The Resistance-Distance Metric Ω. A second fundamental graph metric is[4] the *resistance distance*. This gives $\rho(i,j)$ here denoted $\Omega(i,j)$ the value which is the effective electrical resistance between i and j if unit resistors are imagined on each edge of G. Notably this distance function has the feature of diminishing the distance between i and j the greater the number of "separate" paths there are between i and j—such a feature being comparable to the common practice of distinguishing between single and double bonds with the double-bond length being taken of lesser length than the single-bond length. Further the fundamental nature of this metric Ω is attested to by a number of fundamental graphico-combinatorial formulas for it. E.g., the value for $\Omega(i,j)$ is[11] the same as $1/\Delta_i p(j \leftarrow i)$ where Δ_i is the degree of vertex i and $p(j \leftarrow i)$ is the probability that a random walker from i reaches j before returning to i. Also $\Omega(i,j)$ is[11] the same as $t_{ij}(G)/t(G)$ where $t(G)$ is the number of spanning trees of G while $t_{ij}(G)$ is the number of spanning "two-trees" the two different components of which each contain one of the two vertices i and j. Yet further this distance function may be expressed[13] in terms of the so-called *Laplacian* matrix $\mathbf{L}(G)$ which has rows and columns labeled by vertices of G and has (i,j)th element given by

$$(i|\mathbf{L}(G)|j) =$$

$$(i|\{\Delta(G) - \mathbf{A}(G)\}|j) = \begin{cases} -1, & \{i,j\} \in E(G) \\ \Delta_i, & i = j \\ 0, & \text{otherwise} \end{cases}$$

where $\mathbf{A}(G)$ is the adjacency matrix and $\Delta(G)$ is the diagonal matrix of vertex degrees. Then for connected G the matrix $\mathbf{L}(G)$ has all eigenvalues positive except for one which is 0, whence $\mathbf{L}(G)$ has a generalized inverse $\Gamma(G)$ which is 0 on this null eigenspace and the "true" inverse on the subspace orthogonal to this null space. Then the resistance distance between i and j is

$$\Omega(i,j) = (\phi_{i-j}|\Gamma(G)|\phi_{i-j})$$

where the vectors $|\phi_{i-j})$ are with all elements 0 except for $+1$ and -1 in the ith and jth positions. Finally there are widely known "series" and "parallel" (and Y,Δ-transformation) rules for simplifying the (hand) computation of the $\Omega(i,j)$. With several such fundamental expressions for the resistance distance, one might guess that there should be a rich array of associated theorems, some of which are indicated elsewhere,[4,5,11,14] though only the more recent works interpret the results "metrically".

(c and d) Square-Rooted Metrics $d_{1/2}$ and $\Omega_{1/2}$. Next though neither of these preceding two graph metrics generally permit *faithful* $m+1$-point embeddings into m-dimensional Euclidean space for any $m \geq 3$, it turns out that there is a simple way[2,15] to modify any such ρ so as to obtain a new corresponding metric $\rho_{1/2}$ which does. Simply, for each $i,j \in V$ one takes

$$\rho_{1/2}(i,j) = \{\rho(i,j)\}^{1/2}$$

Thence corresponding to the shortest-path and resistance-distance metrics (of the two preceding paragraphs) we have *square-root-shortest-path* and *square-root-resistance-distance* metrics, respectively denoted $d_{1/2}$ and $\Omega_{1/2}$. The theorematic results on these two (m-Euclidean) metrics seems to be somewhat less than for d and Ω, though some initial results are given in ref 5.

(e) The Quasi-Euclidean Metric ρ_{q-E}. Fifth there is[5] another fundamental graph metric based on the *Laplacian* matrix $\mathbf{L}(G)$ and its generalized inverse $\Gamma(G)$, already discussed in connection with the resistance distance. This final metric here termed *quasi-Euclidean* is defined as

$$\rho_{q-E}(i,j) = (\phi_{i-j}|\{\Gamma(G)\}^2|\phi_{i-j})^{1/2}$$

This metric satisfies the higher m-Euclidean conditions, as do the previously mentioned "square-rooted" metrics. But further like the shortest-path and resistance-distance metrics, the metric ρ_{q-E} scales linearly with inverse distance weights introduced into the adjacency matrix (and similarly too into the degree matrix). Even more so than these two other square-rooted metrics relatively little work seems to have been done.

(f) The Euclidean Metric ρ_E. Finally once a graph is embedded in m-dimensional Euclidean space E_m (by mapping the sites to particular locations) a Euclidean metric ρ_E is implicitly defined on the graph. Of course the consequent distances $\rho_E(i,j)$ depend on the particular embedding. For L_n we take the standard embedding to be equally spaced along a straight line. For S_n we presently forego an embedding. For C_n we take the standard embedding to be the regular n-gon with unit edges. For $C_6^{[m]}$ we take the standard embedding to be the corresponding section of the regular honeycomb lattice. For $F_{h,k}$ we take the standard embedding to have all the sites on the surface of a sphere and bond lengths varying but slightly—the actual site positions have been taken to be given by the "topological coordinates" of Fowler and Manolopoulos[16] and exhibit icosohedral point-group symmetry. Since ρ_E depends on the embedding, it is not ordinarily viewed as a "graph metric".

More precisely we say that a metric on G is a *graph metric* if always the distances between corresponding pairs of sites in all graphs isomorphic to G are the same. Thence metrics (a)−(e) are so graphical, whereas ρ_E is not. But still there often is some correlation between these graph metrics and ρ_E for embeddings associated to typical (equilibrium) molecular structures. Indeed part of the point of the present investigation concerns what this correlation might be like.

3. LINEAR CURVATURE

A first type of graph-geometric invariant assigns a *net linear curvature* k_i associated to each vertex $i \in V(G)$. A vertex i with just a single neighbor is viewed to exhibit no curvature, i.e., $k_i = 0$. Thence consideration defaults to consideration of a typical vertex with at least two vertices, say a and b. The two-Euclidean condition on ρ guarantees that i,a,b may be faithfully embedded in Euclidean two-space E_2, say as in Figure 3. Evidently in first proceeding from a to i along the first leg of this triangle and then second from i to b along the second leg one turns through an angle θ indicated in Figure 3. This is conveniently obtained via (the law of cosines)

$$\cos(\pi - \theta) =$$

$$[\rho(a,b)^2 - \rho(i,a)^2 - \rho(i,b)^2]/[2\cdot\rho(i,a)\cdot\rho(i,b)]$$

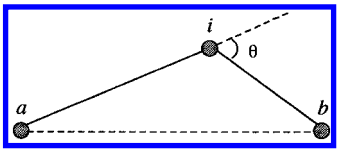

Figure 3. A prototypical imagined triangle formed from a vertex *i* and two distinct neighbors *a* and *b*. The embedding into E_2 is to be faithful (in the sense that the Euclidean distances between the vertices match the graph-theoretic ones given by ρ).

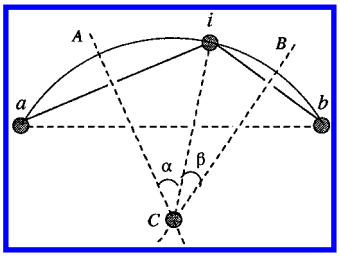

Figure 4. The prototypical triangle of Figure 3 now developed to show the "center of curvature" and the associated circle through *a*, *i*, and *b*.

This angle θ generally is defined for every choice of neighbor pair to *i*, so that the summation of θ overall such choices then is reasonably interpreted as the net curvature k_i.

But there is another possible definition parralleling closely usual differential-geometric ideas, whence one imagines the sites faithfully embedded in E_2 and considers the unique circle passing through the three points *i*,*a*,*b*, as indicated in Figure 4. In particular this circle of curvature has a center *C* located by the perpendicular bisectors *A* and *B* of the two legs *ai* and *ib*, and if the radius of this circle is *r* then the differential curvature at *i* is 1/*r*. But now to obtain the *net* curvature associated to the "curve" from *a* to *b* through vertex *i* one plausably integrates over the arc from the position of the bisector *A* (at an angle α to the left of *Ci*) to the bisector *B* (at an angle β to the right of *Ci*), so that one finds this net result as

$$(1/r)\cdot\{arc\ length\} = (1/r)\cdot(\alpha+\beta)r = \alpha+\beta = \theta$$

with the last equality seen because the angle $(\alpha+\beta)$ turned through by the normal is the same as the angle θ turned through by the tangent. Thus this alternative leads to the same result (and the same sum k_i), so that either view can be taken as the definition.

A third proposal for linear curvature is that of Menger (as discussed in ref 15 or sections 30−33 of ref 2). This takes the curvature at *i* to be twice the area of the triangle aib divided by the product $\rho(a,i)\cdot\rho(i,b)\cdot\rho(b,a)$. But in fact this is just the same as 1/*r* from the preceding paragraph, so that it is a *differential* linear curvature. With the view that graphs are discrete, it seems to us that for such structures a more natural invariant is the *net* linear curvature at a vertex *i* (say as integrated over the region of *i*). For the present we forego this third alternative.

Granted net linear curvatures k_i at the sites, we define the (absolute linear) *curvature sum*

$$k(G,\rho) = \sum_{i\in V} k_i$$

which of course should generally depend on the choice of metric ρ. For a Euclidean metric for a graph embedded in Euclidean two-space $k(G,\rho_E)$ has a classical interpretation, as the net angle turned through at the different vertices. Thus for a graph around the boundary of a convex polygon $k(G,\rho_E)$ would take a value of 2π, and for a graph around the boundary of a nonplanar or even of a planar nonconvex polygon $k(G,\rho_E)$ would take a value greater than 2π. For a graph embedding with $V(G)$ corresponding to the vertices of a trivalent convex polyhedron (and $E(G)$ corresponding to the polyhedron's edges), $k(G,\rho_E)$ sums over the curvatures around each face and so takes the value $2\pi f$ where *f* is the number of faces of the polyhedron.

For the linear chain, tree-star, and simple cycle graphs analytic results are obtainable for each of our five graph metrics as well as for the Euclidean metric for "standard" graph embeddings. Moreover for tree graphs the shortest-path and resistance-distance metrics are equivalent. We find

$$k(G,d) = 0, \qquad G = L_n, S_n, C_n (n \geq 4), C_6^{[m]}, F_{h,k}$$

$$k(L_n,\Omega) = k(S_n,\Omega) = 0$$

$$k(L_n,d_{1/2}) = k(L_n,\Omega_{1/2}) = (n-1)\cdot\pi/2$$

$$k(S_n,d_{1/2}) = k(S_n,\Omega_{1/2}) = n\cdot(n-1)\cdot\pi/2$$

$$k(C_n,d_{1/2}) = n\cdot\pi/4$$

$$k(C_n,\Omega) = n\{\pi - \cos^{-1}[(n^2-6n+7)/(n^2-2n+1)]\}$$

$$k(C_n,\Omega_{1/2}) = n\{\pi - \cos^{-1}[(n^2-6n+7)/(n^2-2n+1)]^{1/2}\}$$

For more explicit numerical comparison for each of the five sequences of graphs considered we report the data of Table 1. It is to be noted that for the cycle-containing graphs *G* each of the "new" graph metrics Ω, $d_{1/2}$, $\Omega_{1/2}$, and ρ_{q-E} give values for $k(G,\rho)$ more closely in concert with that for our standard Euclidean embeddings. (Of incidental note is the fact that $k(F_{h,k},\rho_E)/2\pi$ for the higher fullerenes slightly exceeds the number *f* of faces of $F_{h,k}$ since the face rings are not quite planar.)

4. CURVILINEAR TORSION

In differential geometry after describing the curvature of a curve in Euclidean three-space one often continues to the "torsion" describing how the curve twists out of a plane. This too then is to be dealt with for a graph metric which is defined on a discrete set (of vertices), so that a net torsion (imagined to be a differential torsion integrated over the region of a suitable small region) is more appropriate than a differential torsion (such as are usually considered, even in the non-Euclidean but continuous case, as discussed in section 34 of ref 2). Thence to this end it is natural to consider the *net torsion* $\tau_{\{i,j\}}$at each edge $\{i,j\} \in E(G)$. A "curve" in *G* is imagined to pass through edge $\{i,j\}$ coming from a preceding site *a* and going on to a succeeding site *b*, with $a \neq b$, as in Figure 5. There are two triangles *aij* and *ijb* which give net curvatures at *i* and *j*, and the net torsion associated to the noted curve then is reasonably imagined to be the (coplanarity) deviation taken as the dihedral angle ϕ between the planes of these two triangles. At least in a Euclidean space the value of this angle ϕ is determined from the tetrahedron *aijb* with the additional edge-length $\rho(a,b)$,

Table 1. Linear Curvature Sums Divided by 2π for Various Distance Metrics

graph		d	Ω	$d_{1/2}$	$\Omega_{1/2}$	ρ_{q-E}	ρ_E
L_n	$n=2$	0.0000	0.0000	0.2500	0.2500	0.1667	0.0000
	$n=3$	0.0000	0.0000	0.5000	0.5000	0.3041	0.0000
	$n=4$	0.0000	0.0000	0.7500	0.7500	0.4241	0.0000
	$n=5$	0.0000	0.0000	1.0000	1.0000	0.5320	0.0000
S_n	$n=3$	0.0000	0.0000	0.7500	0.7500	0.5877	
	$n=4$	0.0000	0.0000	1.5000	1.5000	1.2587	
	$n=5$	0.0000	0.0000	2.5000	2.5000	2.1795	
	$n=6$	0.0000	0.0000	3.7500	3.7500	3.3502	
C_n	$n=4$	0.0000	1.0709	1.0000	1.2163	1.1282	1.0000
	$n=5$	0.0000	1.1503	1.2500	1.4511	1.2500	1.0000
	$n=7$	0.0000	1.3050	1.7500	1.9366	1.4685	1.0000
	$n=8$	0.0000	1.3779	2.0000	2.1825	1.5673	1.0000
$C_6^{[m]}$	$m=0$	0.0000	1.2290	1.5000	1.6923	1.3631	1.0000
	$m=1$	0.0000	9.5410	12.0000	13.4937	10.0685	8.0000
	$m=2$	0.0000	29.1440	31.5000	35.7265	25.9416	21.0000
	$m=3$	0.0000	51.0284	60.0000	68.4036	48.9014	40.0000
$F_{h,k}$	(1,0)	0.0000	14.9074	15.0000	17.8044	15.1917	12.0000
	(1,1)	0.0000	42.4950	45.0000	52.6208	41.7739	32.0200
	(2,0)	0.0000	56.2991	60.0000	70.0340	54.8141	42.0136
	(2,1)	0.0000	97.7071	105.0000	122.2722	93.3573	72.0224
	(3,0)	0.0000	125.3136	135.0000	157.0981	118.7729	92.0279
	(2,2)	0.0000	166.7230	180.0000	209.3368	156.6105	122.0316

whence through standard trigonometric manipulation (as in the Appendix) one finds

$$\cos(\phi) = \left|\frac{\Delta(a,j;i)\cdot\Delta(b,i;j) + 2\cdot\rho(i,j)^2\cdot[\Delta(a,j;i) - \Delta(a,b;i)]}{16\cdot A(a,i,j)\cdot A(i,j,b)}\right|$$

where $\Delta(i,j;k)$ and $A(i,j,k)$ respectively denote a "Pythagorean defect" and the area of a triangle, thusly

$$\Delta(i,j;k) = \rho(i,j)^2 - \rho(i,k)^2 - \rho(j,k)^2$$

$$A(i,j,k) = \{s\cdot[s - \rho(i,j)]\cdot[s - \rho(j,k)]\cdot[s - \rho(k,i)]\}^{1/2}$$

with $s \equiv \{\rho(i,j) + \rho(j,k) + \rho(k,i)\}/2$ half the triangle's perimeter. The sum of such deviations ϕ for all such curves through $\{i,j\}$ then is here taken to give the net torsion $\tau_{\{i,j\}}$ at this edge.

It is to be emphasized that this definition can be seen as somewhat "contentious" unless the metric ρ is three-Euclidean. Basically if it is not three-Euclidean there can in general be a problem constructing the tetrahedron suggested in Figure 5 since concievably either the distance $\rho(a,b)$ can turn out to be greater than the Euclidean distance between a and b when the two triangles aij and ijb are laid flat (sharing their ij-bases) in a single Euclidean plane more-or-less as in the figure, or alternatively this distance $\rho(a,b)$ can be too small to allow a faithful embedding in E_3. Of course, for the metrics $d_{1/2}$, $\Omega_{1/2}$, and ρ_{q-E} the three-Euclidean condition is satisfied, so that no such problem arises in these cases.

Granted the net torsions at each edge of the graph G we define the *net torsion sum*

$$\tau(G,\rho) = \sum_{\{i,j\}\in E} \tau_{\{i,j\}}$$

For the Euclidean metric for a graph embedded in Euclidean three-space, this has a classical interpretation, as the total amount of twisting (e.g., as in a helically twisted chain) within the whole structure.

Again for the linear-chain, tree-star, and simple cycle graphs analytic results are obtainable. For any graphs

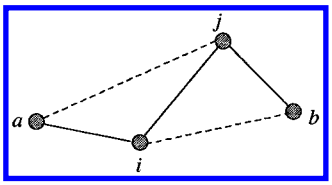

Figure 5. A prototypical curve a,i,j,b through a pair of neighbor sites i,j. In general we do not imagine the four points to be coplanar, though here we imagine the two triples a,i,j and i,j,b to each be separately embedded into an E_2 space.

containing no cycles smaller than 6 we have $\tau_{\{i,j\}}(G,d) = 0$, and the case of the tree-star is particularly trivial, since it contains no four-site chains, thereby giving $\tau(S_n,\rho) = 0$. For L_n and C_n we find

$$\tau(L_n,d_{1/2}) = \tau(L_n,\Omega_{1/2}) = n\cdot\pi/2,\ n \geq 3$$

$$\tau(C_n,d_{1/2}) = n\cdot\pi/2,\ n \geq 6$$

$$\tau(C_n,\Omega) = n\cdot\cos^{-1}\{(n^3 - 11n + 12)/[2(n-2)^2(2n-3)]\}$$

$$\tau(C_n,\Omega_{1/2}) = n\cdot\cos^{-1}[1/(n-2)]$$

Again for more explicit comparison for all our sequences of cycle-containing graphs we report numerical data in Table 2. The entries left blank in this table are for cases where the i,a,b,j-tetrahedron discussed in our definition of k_i is not faithfully embeddable in Euclidean space. This problem is manifested with the formula for cos ϕ in that it gives $+\infty$. For L_n (and indeed for any trees with paths of length 4 or greater) both d and Ω actually give 0/0 for the value of cos ϕ, via the given formula, but the tetrahedron is degenerately realizable as a straight line with torsion 0. Of course since $d_{1/2}$, $\Omega_{1/2}$, and ρ_{q-E} are three-Euclidean, they always allow an isometric embedding in E_3 of the four-site torsion curves and therefore never lead to catastrophes for the torsions. Moreover, our data seems to indicate that Ω seldom has "catastrophic" problems.

5. GAUSSIAN CURVATURE

A further possibility is to introduce net *Gaussian curvatures* κ_i at each vertex i of a graph G. Just as a net linear

Table 2. Torsion Curvature Sums Divided by 2π for Various Distance Metrics

graph		d	Ω	$d_{1/2}$	$\Omega_{1/2}$	ρ_{q-E}	ρ_E
L_n	$n=3$	0.0000	0.0000	0.2500	0.2500	0.2500	0.0000
	$n=4$	0.0000	0.0000	0.5000	0.5000	0.5000	0.0000
	$n=5$	0.0000	0.0000	0.7500	0.7500	0.7500	0.0000
C_n	$n=4$		0.4097	0.0000	0.6667	0.5354	0.0000
	$n=5$		0.6861	0.8333	0.9796	0.8333	0.0000
	$n=7$	0.0000	1.1596	1.7500	1.5257	1.3714	0.0000
	$n=8$	0.0000	1.3831	2.0000	1.7868	1.6311	0.0000
$C_6^{[m]}$	$m=0$	0.0000	0.9295	1.5000	1.2587	1.1070	0.0000
	$m=1$	0.0000	14.6133	19.5000	17.3399	14.6114	0.0000
	$m=2$	0.0000	44.2854	55.5000	49.1617	39.9153	0.0000
	$m=3$	0.0000	89.8385	109.5000	96.6864	76.3327	0.0000
$F_{h,k}$	(1,0)		23.9230	25.0000	25.1718	25.0080	10.5725
	(1,1)		74.8542	85.0000	75.5113	71.8732	15.9993
	(2,0)		100.3949	115.0000	103.7146	94.9148	26.1888
	(2,1)		176.6514	205.0000	182.1475	161.3660	36.7614
	(3,0)		227.5631	265.0000	234.4713	204.9408	43.0288
	(2,2)		303.5776	355.0000	312.7991	267.9301	44.4077

Table 3. Gaussian Curvature Sums Divided by 4π for $C_6^{[m]}$ and $F_{h,k}$ for Various Distance Metrics

graph		d	Ω	$d_{1/2}$	$\Omega_{1/2}$	ρ_{q-E}	ρ_E
$C_6^{[m]}$	$m=1$	−3.0000	0.4323	1.5000	2.0205	0.7605	0.0000
	$m=2$	−9.0000	2.0100	4.5000	6.2541	2.1586	0.0000
	$m=3$	−18.0000	4.7262	9.0000	12.7096	4.1296	0.0000
$F_{h,k}$	(1,0)	−5.0000	2.4537	2.5000	3.9022	2.5958	1.0000
	(1,1)	−15.0000	6.2475	7.5000	11.3104	5.8870	1.0000
	(2,0)	−20.0000	8.1496	10.0000	15.0170	7.4070	1.0068
	(2,1)	−35.0000	13.8536	17.5000	26.1361	11.6787	1.0112
	(3,0)	−45.0000	17.6568	22.5000	33.5490	14.3863	1.0139
	(2,2)	−60.0000	23.3615	30.0000	44.6684	18.3053	1.0158

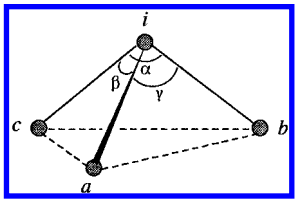

Figure 6. A prototypical tetrahedral region around a vertex *i*.

curvature measures the angle turned through (say as determined from the set of normals) in traveling along a curve, so does a net Gaussian curvature measure the solid angle covered by the normals from a patch of surface. Here the result is understood to be 0 for vertices of degree 0, 1, or 2, and again there are two equivalent ways to define the net Gaussian curvature. With the first view one considers a vertex *i* with three neighbors, say *a*,*b*,*c* much as imagined in Figure 6, whence the net Gaussian curvature for this triplet of neighbor sites is taken as (the so-called Descartes defect) $2\pi-\alpha-\beta-\gamma$, and if there are just three neighbors to *i* this is κ_i. It may be noted that regardless of whether the metric allows a faithful Euclidean embedding of the quartet *i*,*a*,*b*,*c* one may still define the angles α, β, and γ (occurring in triangles *iab*, *ibc*, and *ica* using the law of cosines) and consequently define κ_i also.

Again as for the case of linear curvatures (in section 3 here) there is an alternative view involving inverse radii of curvature. One can imagine a spherical surface fit through *i*,*a*,*b*,*c* to determine a differential Gaussian curvature and take an integration over the portion of this surface nearer to *i* than *a*, *b*, or *c*—but again the result is the same. Too there have been previous general developments with the continuous case in mind (in ref 17 and section 35 of ref 2), whence differential Gaussian curvatures are developed. But we again prefer the present type of discretized view. Indeed the discretized view here (and then too that in the preceding two sections) is related to that of Regge[18] who treats geometric structures (curves, surfaces, volumes, etc.) in terms of piecewise linear simplicial sections each taken as an unsubdivided whole in a continuous geometric space. Regge's formula for (net) Gaussian curvature then coincides with that given here, and further extensions to discretized formulas for the full Riemann curvature tensor are given.

The question of how to treat vertices with degree four or greater we leave for further resolution. Basically we see two plausable approaches, the first of which follows in analogy to the preceding two sections and sums $2\pi-\alpha-\beta-\gamma$ over all triples of nearest neighbors. The second would somehow imagine all the neighbors to *i*, say *a*,*b*,*c*,*d* (for the degree-four case) to be arranged cyclicly around *i* all in a single (polyhedral) surface and take κ_i to be $2\pi-\alpha-\beta-\gamma-\delta$. In any event the graphs (other than S_n) currently considered are no more than degree 3.

Granted net Gaussian curvatures for each vertex of *G*, we define *net* (absolute) *Gaussian-curvature sums*

$$\kappa(G,\rho) = \sum_{i \in V} \kappa_i$$

Again at least for the Euclidean metric for a graph embedding in Euclidean three-space this sum has a classical interpretation as the total amount of absolute Gaussian curvature integrated over the whole surface and is simply bounded via the Gauss−Bonnet theorem in terms of the Euler characteristic for the surface. Only a bound rather than an equality is obtained because in many chemical cases there is a lack of planarity of what are otherwise perceived to be faces of polyhedra. Thus for a convex polyhedral structure such as buckminsterfullerene the net Gaussian curvature is just 4π, while it is greater than 4π if a polyhedral surface has "bent" faces.

For the case of Gaussian curvatures the linear-chain and simple cycle graphs the result is trivial, being $\kappa_i(L_n,\rho) = 0$ and $\kappa_i(C_n,\rho) = 0$ since there are no vertices of degree 3. For the tree-stars S_n there is a vertex of degree exceeding 3 (for $n \geq 4$), so that we have not yet defined κ_i (or $\kappa(S_n,\rho)$) and do not consider this sequence of graphs. Numerical results are given in Table 3 for the two remaining types of graphs ($C_6^{[m]}$ and $F_{h,k}$). It is seen that there is a better consonance between the $\kappa(G,\rho_E)$ and $\kappa(G,\rho)$ for the four new graph metrics ($\rho = \Omega$, $d_{1/2}$, $\Omega_{1/2}$, ρ_{q-E}) than there is for the conventional shortest-path graph metric.

6. ROOT-MEAN-SQUARE VOLUMINA

A metric gives distances so that naturally also one might consider areas and higher volumina. But this entails a presumptive definition for these quantities, and we follow an earlier graph-theoretically applied suggestion[5] of using the so-called "Cayley−Menger" determinant to make this definition—this suggestion being in the spirit of the general "coordinate-free" geometry reviewed by Blumenthal.[2] That is, for a given metric ρ on G we define the square of the *l-voluminum* of the l-simplex determined by the $l + 1$ points $i_0, i_1, i_2, ..., i_l$ to be

$$\{V_l(\mathbf{D}_l)\}^2 = -\frac{(-2)^{-l}}{(l!)^2}\cdot\det\begin{pmatrix}\mathbf{D}_2 & \mathbf{1}_{l+1}\\ \mathbf{1}^+_{l+1} & \mathbf{0}\end{pmatrix}$$

where $\mathbf{D}_q$ is here the $(l + 1) \times (l + 1)$ matrix with (i,j)th element the qth power of the distance between the ith and jth vertices of this l-simplex, $\mathbf{1}_{l+1}$ is the $(l + 1) \times 1$ column vector with all elements 1, and $\mathbf{1}^+_{l+1}$ is the transpose of $\mathbf{1}_{l+1}$. This voluminum is normalized so that for a unit l-simplex (with unit distance between every pair of vertices) it takes a magnitude of $1/(l!)$.

Granted the l-voluminum for each l-simplex of G, we define the *root-mean-square (rms) l-voluminum* for G

$$\nu_l(G,\rho) = \{\sum_{simplex}[V_l(\mathbf{D}_1)]^2/\binom{N}{l}\}^{1/2}$$

where the summation is over all l-simplices of G and $\binom{N}{l}$ is the standard binomial coefficient (counting these l-simplices). Efficient means for computing $\nu_l(G,\rho)$ without going through all the $\binom{N}{l}$ individual l-simplexes is available.[5] (But if a power other than 2 for the individual volumina of the l-simplices were to appear in the summand, then such a result has not yet been established.)

Further in Tables 4−8 we report numerical data for the (lower) *rms l*-volumina, each table being devoted to one of the sequences of graphs L_n, S_n, C_n, $C_6^{[m]}$, and $F_{h,k}$. Again any blanks for the *rms l*-volumina as given by the new graph metrics do not indicate zero-values but rather ill-defined values. We see that the standard shortest-path metric frequently has problems for graphs other than the linear chain L_n: often the formula gives negative values for the squares of l-volumina, when $l \geq 3$. The other graph metrics do however typically give nonzero l-volumina for dimensions $l \geq 4$, though for the graphs considered here these higher *rms l*-volumina seem to decrease as l increases.

7. DISCUSSION

Overall it seems that the new graph metrics (Ω, $d_{1/2}$, $\Omega_{1/2}$, ρ_{q-E}) each exhibit closer correspondences with standard embeddings of some chemically interesting graphs than is the case with the "classical" shortest-path metric, which in the past has often been considered to the exclusion of other natural graph metrics. Moreover we surmise that the correspondence for these metrics is closer for graphs representing molecular species with lesser flexibility—in these cases the rigidity is mandated by cycles to which especially Ω, $\Omega_{1/2}$, and ρ_{q-E} rather naturally pay keen attention. From our numerical results we might try to draw a slightly more detailed conclusion, ordering the various graph metrics according to their "similarity" to the results for the Euclidean metric for standard embeddings in E_3: for "cycle-rich" graphs it seems

$$d < \Omega_{1/2} < d_{1/2} < \Omega \sim \rho_{q-E}$$

(with d bearing the least similarity to ρ_E and both Ω and ρ_{q-E} bearing comparable maximal similarities to ρ_E). For circumstances where there are cut edges it seems to us that there is less reason to expect a graph metric to reasonably represent distances between vertices separated by such a cut edge—and our data may be construed to support this contention also, though ρ_{q-E}, $\Omega_{1/2}$, and $d_{1/2}$ here even give noncatastrophic results.

Perhaps most notably the four new graph metrics seldom suffer the "catastrophes" which the conventional shortest-path metric rather often exhibits. These catastrophes include "cosines" of torsion angles with magnitude greater than 1 and "squares" of volumina being negative. Rather

Table 4. rms l-Volumina of L_n for Various Distance Metrics

n	l	d	Ω	$d_{1/2}$	$\Omega_{1/2}$	ρ_{q-E}	ρ_E
2	1	1.4142	1.4142	1.1547	1.1547	1.0541	1.4142
	2	0.0000	0.0000	0.5000	0.5000	0.2887	0.0000
3	1	1.8257	1.8257	1.2910	1.2910	1.4720	1.8257
	2	0.0000	0.0000	0.6124	0.6124	0.5000	0.0000
	3	0.0000	0.0000	0.1667	0.1667	0.0833	0.0000
4	1	2.2361	2.2361	1.4142	1.4142	1.9494	2.2361
	2	0.0000	0.0000	0.7246	0.7246	0.8093	0.0000
	3	0.0000	0.0000	0.2108	0.2108	0.1563	0.0000
	4	0.0000	0.0000	0.0417	0.0417	0.0186	0.0000
5	1	2.6458	2.6458	1.5275	1.5275	2.4788	2.6458
	2	0.0000	0.0000	0.8367	0.8367	1.2383	0.0000
	3	0.0000	0.0000	0.2582	0.2582	0.2755	0.0000
	4	0.0000	0.0000	0.0538	0.0538	0.0367	0.0000

Table 5. rms l-Volumina of S_n for Various Distance Metrics

n	l	d	Ω	$d_{1/2}$	$\Omega_{1/2}$	ρ_{q-E}
3	1	1.5811	1.5811	1.2247	1.2247	1.1726
	2	0.8660	0.8660	0.6124	0.6124	0.5303
	3			0.1667	0.1667	0.0833
4	1	1.6733	1.6733	1.2649	1.2649	1.2329
	2	1.0955	1.0955	0.6708	0.6708	0.6245
	3	0.2981	0.2981	0.2108	0.2108	0.1764
	4			0.0417	0.0417	0.0186
5	1	1.7321	1.7321	1.2910	1.2910	1.2693
	2	1.2247	1.2247	0.7071	0.7071	0.6770
	3	0.4714	0.4714	0.2357	0.2357	0.2152
	4			0.0538	0.0538	0.0439
6	1	1.7728	1.7728	1.3093	1.3093	1.2936
	2	1.3093	1.3093	0.7319	0.7319	0.7107
	3	0.5634	0.5634	0.2520	0.2520	0.2381
	4	0.1409	0.1409	0.0610	0.0610	0.0549

Table 6. rms l-Volumina of C_n for Various Distance Metrics

n	l	d	Ω	$d_{1/2}$	$\Omega_{1/2}$	ρ_{q-E}	ρ_E
4	1	1.4142	0.8416	1.1547	0.9128	0.6124	1.1547
	2	0.0000	0.2795	0.5000	0.3536	0.1531	0.5000
	3		0.0417	0.0000	0.0833	0.0208	0.0000
5	1	1.5811	1.0198	1.2247	1.0000	0.7746	1.3450
	2	0.6847	0.3909	0.5863	0.4183	0.2345	0.6396
	3		0.0638	0.1443	0.1054	0.0365	0.0000
	4		0.0066	0.0233	0.0186	0.0037	0.0000
7	1	2.1603	1.3801	1.4142	1.1547	1.1547	1.7603
	2	1.1068	0.6821	0.7583	0.5477	0.4928	1.0393
	3		0.1355	0.2075	0.1543	0.0987	0.0000
	4		0.0159	0.0366	0.0302	0.0124	0.0000
8	1	2.5071	1.5613	1.5119	1.2247	1.3693	1.9753
	2	1.0690	0.8612	0.8452	0.6124	0.6806	1.2905
	3		0.1878	0.2254	0.1809	0.1534	0.0000
	4		0.0238	0.0315	0.0369	0.0212	0.0000

Table 7. rms *l*-Volumina of $C_6^{[m]}$ for Various Distance Metrics

n	*l*	*d*	Ω	$d_{1/2}$	$\Omega_{1/2}$	ρ_{q-E}	ρ_E
0	1	1.9494	1.1995	1.3416	1.0801	0.9560	1.5492
	2	0.5477	0.5253	0.6708	0.4830	0.3459	0.8216
	3		0.0946	0.1491	0.1291	0.0613	0.0000
	4		0.0104	0.0000	0.0241	0.0069	0.0000
1	1	3.9891	1.7540	1.9054	1.2938	1.9028	3.1485
	2	4.6013	1.1705	1.4185	0.6983	1.3342	3.1035
	3		0.4247	0.5948	0.2289	0.4840	0.0000
	4		0.0985	0.1630	0.0528	0.1106	0.0000
2	1	6.0166	2.0218	2.3360	1.3899	2.8047	4.7345
	2	10.8013	1.5883	2.1393	0.8102	2.8663	6.9289
	3		0.7033	1.1165	0.2885	1.5016	0.0000
	4		0.2059	0.3883	0.0727	0.4931	0.0000
3	1	8.0381	2.2008	2.6984	1.4518	3.7081	6.3179
	2	19.4570	1.9055	2.8573	0.8872	4.9836	12.2865
	3		0.9456	1.7300	0.3322	3.4120	0.0000
	4		0.3147	0.7016	0.0883	1.4576	0.0000

Table 8. rms *l*-Volumina of $F_{h,k}$ for Various Distance Metrics

(*h*,*k*)	*l*	*d*	Ω	$d_{1/2}$	$\Omega_{1/2}$	ρ_{q-E}	ρ_E
(1,0)	1	2.8470	0.9761	1.6222	0.9805	0.8648	2.0332
	2	2.0775	0.3888	1.0260	0.4091	0.3009	1.5016
	3		0.0914	0.3576	0.1055	0.0608	0.5071
	4		0.0147	0.0774	0.0194	0.0082	0.0000
(1,1)	1	5.0990	1.3156	2.1707	1.1385	1.5377	3.5473
	2	7.6636	0.7131	1.8510	0.5524	0.9347	4.4870
	3		0.2314	0.8914	0.1660	0.3240	2.5445
	4		0.0521	0.2797	0.0357	0.0743	0.0000
(2,0)	1	6.0640	1.3997	2.3600	1.1747	1.7752	4.0977
	2	10.2191	0.8096	2.1703	0.5886	1.2422	5.9745
	3		0.2814	1.1148	0.1828	0.4943	3.8966
	4		0.0681	0.3683	0.0407	0.1297	0.0000
(2,1)	1	7.9620	1.5594	2.7087	1.2408	2.3407	5.4124
	2	18.6743	1.0107	2.8767	0.6578	2.1512	10.3946
	3		0.3960	1.7261	0.2165	1.1192	8.9047
	4		0.1087	0.6778	0.0511	0.3820	0.0000
(3,0)	1	9.1181	1.6302	2.8961	1.2690	2.6499	6.1347
	2	24.1767	1.1071	3.2819	0.6886	2.7532	13.3432
	3		0.4556	2.0959	0.2321	1.6166	12.9350
	4		0.1316	1.3736	0.0562	0.6218	0.0000
(2,2)	1	10.5118	1.7106	3.1103	1.3004	3.0544	7.0799
	2	32.2359	1.2221	3.7878	0.7237	3.6528	17.7592
	3		0.5304	2.6036	0.2504	2.4650	19.8404
	4		0.1620	1.1709	0.0623	1.0877	0.0000

Table 9. Correlation Coefficients for Octane Boiling Point Fits

invariants	correlation coefficients	
	one invariant	two invariants
$k(G,d_{1/2})$, $\tau(G,d_{1/2})$	0.721	0.919
$k(G,\rho_{q-E})$, $\tau(G,\rho_{q-E})$	0.715	0.936
$V_1(G,d_{1/2})$, $V_2(G,d_{1/2})$	0.533	0.789
$V_1(G,\rho_{q-E})$, $V_2(G,\rho_{q-E})$	0.448	0.785
$W_1(G,d_{1/2})$, $W_2(G,d_{1/2})$	0.544	0.555
$W_1(G,d)$, $W_2(G,d)$	0.538	0.539
$W_1(G,\rho_{q-E})$, $W_2(G,\rho_{q-E})$	0.478	0.479

less often the resistance distance exhibits such problems (for certain circumstances where the resistance-distance and shortest-path metrics coincide). The general absence of such catastrophes for the new graph metrics presumably makes the associated graph-geometric invariants more plausably geometrically interpretable as here indicated in their names.

Sometimes the new graph metrics yield invariants in rather surprisingly close correspondence with those for the Euclidean metric applied to standard embeddings in E_3. This semiquantitative correspondence is most notable for the linear curvature sum (of section 3), this involving the most local of the here considered invariants. Beyond the surface curvatures considered here one could imagine also looking at the higher-dimensional curvatures for which Regge[18] has developed "discretized" formulas.

8. GRAPH GEOMETRY AND MOLECULAR STRUCTURE

While there is general recognition that molecular (geometric) shape is quite crucial in understanding molecular properties, there is less of a consensus as to how to encode it in a form convenient for structure−property or structure−activity correlations. See, e.g., Mezey.[19] Comparisons between graph-metrical and Euclidean-metrical (when *G* is embedded in E_3) distances have been deemed relevant[20] in characterizing shape. A more purely graph-theoretic viewpoint[21] (using the shortest-path metric) embeds each molecular graph in a larger graph (e.g., a lattice graph) which in some sense more closely mimics the structure of Euclidean space. But from our study here we suggest that some aspects of molecular (geometric) shape may in fact be intrinsic in each individual molecular graph. As to the practical utility a simple example utilizing the new graph-geometric invariants might be of value.

As an initial test of the potential for structure−property correlations we consider the boiling points for the 18 octanes C_8H_{18}. Following Wiener's[7] early work just this property has developed as a kind of test bed for an initial assessment of new graph invariants to be used for structure−property correlations. See, e.g., reviews.[6] Here the question of comparing to theoretical results involving "correct" molecular geometries is rather ambiguous, since in fact there is a statistical-mechanical distribution of geometric structures for these floppy molecules, and indeed in statistical mechanics beyond the geometry of individual molecules, the (suitably averaged) intermolecular geometries of arrangement (of large numbers) of molecules is crucial. Here we use boiling points from ref 22, and we report correlations with one or two invariants as compared to linear fits of the first and second moments of the shortest-path distance—this pair of invariants being essentially what is usually called the Wiener number[6] and hyper-Wiener number,[9,23] here denoted by $W_1(G,d)$ and $W_2(G,d)$. The correlation coefficients *r* for these various fits are reported in Table 9. Rather interestingly at this low-level of fitting it is seen that the new invariants frequently achieve better linear fits than the standard $W_1(G,d)$ and $W_2(G,d)$. Analogous pairs $W_1(G,\rho)$ and $W_2(G,\rho)$ for our other metrics ρ seem to give roughly the same results independently of the choice for ρ (though such a result is not true for the curvature-torsion pairs, where often the choice $\rho = d$ exhibits some sort of "catastrophe").

In conclusion it is proposed that the new graph metrics especially ρ_{q-E} and perhaps also Ω should prove useful in modeling molecular species. The new graph invariants, namely the linear-curvature sum, the torsion sum, the Gaussian-curvature sum, and the *rms* volumina, as based on the new graph metrics hopefully will have utility in characterizing molecular structures and their propertities. We suggest that certainly there should be more study of the *geometry of graphs* and of its possible chemical uses.

ACKNOWLEDGMENT

Finally acknowledgment is made for support of this research by the Welch Foundation of Houston, TX.

APPENDIX

To obtain the formula of section 4 for the dihedral angle ϕ imagine an embedding of the points *a*,*i*,*j*,*b* in E_3, and let

$\mathbf{r}_{q\leftarrow p}$ denote the vector from a point p to point q. Then normals to the planes of the triangles a,i,j and i,j,b are $\mathbf{r}_{a\leftarrow i} \times \mathbf{r}_{j\leftarrow i}$ and $\mathbf{r}_{j\leftarrow i} \times \mathbf{r}_{b\leftarrow j}$, respectively (where $\times$ denotes the standard cross-product of vector analysis). Thence the cosine of the dihedral angle ϕ can be obtained from the dot-product between these two normals, thusly

$$\cos\phi = \left|\frac{(\mathbf{r}_{a\leftarrow i} \times \mathbf{r}_{j\leftarrow i})\cdot(\mathbf{r}_{i\leftarrow j} \times \mathbf{r}_{b\leftarrow j})}{|\mathbf{r}_{a\leftarrow i} \times \mathbf{r}_{j\leftarrow i}||\mathbf{r}_{i\leftarrow j} \times \mathbf{r}_{b\leftarrow j}|}\right| \qquad (A.1)$$

Requiring $\cos\phi$ to be positive here merely chooses ϕ as the smaller of the two supplementary angles ϕ_0 and $\pi - \phi_0$ appearing between two intersecting planes. First to treat the magnitudes in the denominator on the right-hand side of (A.1) recall (the standard vector-analysis result) that $|\mathbf{u} \times \mathbf{v}|$ gives the area of the parallelapiped with vectors $\mathbf{u}$ and $\mathbf{v}$ along adjacent sides, this area being twice the area of the triangle with $\mathbf{u}$ and $\mathbf{v}$ along two of its sides. Thence the two magnitudes in the denominator of eq A.1 are $2A(a,i,j)$ and $2A(i,j,b)$. Now to treat the numerator on the right-hand side of eq A.1 we use a general identity of vector analysis, namely

$$(\mathbf{s} \times \mathbf{t})\cdot(\mathbf{u} \times \mathbf{v}) = (\mathbf{s}\cdot\mathbf{u})(\mathbf{t}\cdot\mathbf{v}) - (\mathbf{s}\cdot\mathbf{v})(\mathbf{t}\cdot\mathbf{u}) \qquad (A.2)$$

thus

$$\cos\phi = \left|\frac{(\mathbf{r}_{a\leftarrow i}\cdot\mathbf{r}_{i\leftarrow j})(\mathbf{r}_{j\leftarrow i}\cdot\mathbf{r}_{b\leftarrow j}) - (\mathbf{r}_{a\leftarrow i}\cdot\mathbf{r}_{b\leftarrow j})(\mathbf{r}_{j\leftarrow i}\cdot\mathbf{r}_{i\leftarrow j})}{4A(a,i,j)A(i,j,b)}\right| \qquad (A.3)$$

For the consequent dot-product between pairs of our relative-position vectors we note

$$\mathbf{r}_{a\leftarrow i}\cdot\mathbf{r}_{i\leftarrow j} = \frac{1}{2}(r_{aj}^2 - r_{ai}^2 - r_{ij}^2) = \frac{1}{2}\Delta(a,j;i)$$

$$\mathbf{r}_{j\leftarrow i}\cdot\mathbf{r}_{b\leftarrow j} = \frac{1}{2}(r_{ib}^2 - r_{ij}^2 - r_{bj}^2) = \frac{1}{2}\Delta(i,b;j) \qquad (A.4)$$

$$\mathbf{r}_{a\leftarrow i}\cdot\mathbf{r}_{b\leftarrow j} = \frac{1}{2}(r_{aj}^2 - r_{bi}^2 - r_{ab}^2 - r_{ij}^2) = \frac{1}{2}\{\Delta(a,j;i) - \Delta(a,b;i)\}$$

$$\mathbf{r}_{j\leftarrow i}\cdot\mathbf{r}_{i\leftarrow j} = -r_{ij}^2$$

(where the magnitude of a vector $\mathbf{r}_{q\leftarrow p}$ is written as r_{pq}). Combination of these various results together along with the replacement of r_{pq} by $\rho(p,q)$, then gives the quoted result for $\cos\phi$ in section 4.

REFERENCES AND NOTES

(1) Trinajstić, N. *Chemical Graph Theory*; CRC: Boca Raton, FL, 1992.
(2) Blumenthal, L. M. *Distance Geometry*; Chelsea: New York, 1970.
(3) Buckley, F.; Harary, F. *Distance in Graphs*; Addison-Wesley: Reading, MA, 1989.
(4) Klein, D. J.; Randić, M. Resistance Distance. *J. Math. Chem.* **1993**, *12*, 81−95.
(5) Klein, D. J.; Zhu, H.-Y. Distances and Volumina for Graphs. *Disc. Math.* Submitted for publication.
(6) (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.
(7) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1948**, *69*, 17−20.
(8) (a) Bonchev, D.; Balaban, A. T.; Liu, X.; Klein, D. J. Molecular Cyclicity and Centricity of Polycyclic Graphs. I. *Intl. J. Quantum Chem.* **1994**, *50*, 1−20. (b) Balaban, A. T.; Liu, X.; Klein, D. J.; Babic, D.; Schmalz, T. G.; Seitz, W. A.; Randić, M. Graph Invariants for Fullerenes. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 396−404.
(9) 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.
(10) (a) Goldberg, M. A Class of Multi-Symmetric Polyhedra. *Tohoku Math. J.* **1937**, *43*, 104−108. (b) Caspar, D. L. D.; Klug, A. Physical Principles in the Construction of Regular Viruses. *Cold Spring Harbor Symp. Quantitative Biol.* **1962**, *27*, 1−24. (c) Coexeter, H. S. M. Virus Molecules and Geodesic Domes. In *A Spectrum of Mathematics*; Butcher, J. C., Ed.; Auckland/Oxford University Press: Wellington, New Zealand, 1971; pp 98−107. (d) Klein, D. J.; Seitz, W. A.; Schmalz, T. G. Icosahedral Symmetry Carbon Cage Molecules. *Nature* **1986**, *323*, 703−706. (e) Fowler, P. W. How Unusual is C_{60}? Magic Numbers for Carbon Clusters. *Chem. Phys. Lett.* **1986**, *131*, 444−450.
(11) Doyle, P. G.; Snell, J. L. *Random Walks and Electrical Networks*; Math. Assoc. Am.: Washington, D.C., 1984.
(12) Shapiro, L. W. An Electrical Lemma. *Math. Mag.* **1987**, *60*, 36−38.
(13) The result to be quoted may be "extracted" from different electrical engineering texts but is also proved in ref 4.
(14) (a) Kirchhoff, G. Uber die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Verteilhung Galvanischer Ströme geführt wird. *Ann. Phys. Chem.* **1847**, *72*, 497−508. For a reprint (English translation), see: *IRE Trans. Cir. Th.* **1958**, *5*, 4−7. (b) Weinberg, L. Kirchoff's "Third and Fourth Laws". *IRE Trans. Cir. Th.* **1958**, *5*, 8−30. (c) Reza, F. Some Topological Considerations in Network Theory. *IRE Trans. Cir. Th.* **1958**, *5*, 30−42. (d) Coates, C. L. Genral Topological Formulas for Linear Network Functions. *IRE Trans. Cir. Th.* **1958**, *5*, 42−54. (e) Bryant, P. R. Graph Theory and Electrical Networks. In *Applications of Graph Theory*; Wilson, R. J., Beineke, L. W., Eds.; AP, New York, 1979; pp 81−119.
(15) Menger, K. Zur Metrik der Kurven. *Math. Ann.* **1930**, *103*, 486−501.
(16) Manolopoulos, D. E.; Fowler, P. W. Molecular Graphs, Point Groups, and Fullerenes. *J. Chem. Phys.* **1992**, *96*, 7603−7614.
(17) Wald, A. Begrunding einer koordinatenlosen Differentialgeometrie der Flachen. *Ergeb. Math. Kolloq. (Wien)* **1936**, *7*, 24−36.
(18) Regge, T. General Relativity without Coordinates. *Nuovo Cimento* **1961**, *19*, 558−571.
(19) Mezey, P. G. *Shape in Chemistry*; VCH: New York, 1993.
(20) (a) Randić, M.; Kleiner, A. F.; DeAlba, L. M. Distance/Distance Matrices. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 277−286. (b) Klein, D. J. Similarity and Dissimilarity in Posets. *J. Math. Chem.* **1995**, *18*, 321−348.
(21) (a) Domb, C. Self-Avoiding Walks on Lattices. *Adv. Chem. Phys.* **1969**, *15*, 229−260. (b) Klein, D. J.; Seitz, W. A. Graphs, Polymer Models, Excluded Volume, and Chemical Reality. In *Chemical Applications of Topology and Graph Theory*; King, R. B., Ed.; Elsevier: Amsterdam, 1983; pp 430−445. (c) Kuchanov, S. I.; Korolev, S. V.; Panykov, S. V. Graphs in Chemical Physics of Polymers. *Adv. Chem. Phys.* **1986**, *72*, 115−326.
(22) Needham, D. E.; Wei, I.-C.; Seybold, P. G. Molecular Modeling of the Physical Properties of the Alkanes. *J. Am. Chem. Soc.* **1988**, *110*, 4186−4194.
(23) (a) Randić, M. Novel Molecular Description for Structure−Property Studies. *Chem. Phys. Lett.* **1993**, *211*, 478−483. (b) Randić, M. Molecular Shape Profiles. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 373−382.

CI960025A