On Walks in Molecular Graphs

Ivan Gutman,[†] Christoph Rücker,*[,‡] and Gerta Rücker[‡]

Faculty of Science, University of Kragujevac, P.O. Box 60, YU-34000 Kragujevac, Yugoslavia, and Institut für Organische Chemie und Biochemie, Universität Freiburg, Albertstrasse 21, D-79104 Freiburg, Germany

Received October 31, 2000

Walks in molecular graphs and their counts for a long time have found applications in theoretical chemistry. These are based on the fact that the (*i*, *j*)-entry of the *k*th power of the adjacency matrix is equal to the number of walks starting at vertex *i*, ending at vertex *j*, and having length *k*. In recent papers (refs 13, 18, 19) the numbers of all walks of length *k*, called molecular walk counts, mwc_k, and their sum from $k = 1$ to $k = n - 1$, called total walk count, *twc*, were proposed as quantities suitable for QSPR studies and capable of measuring the complexity of organic molecules. We now establish a few general properties of *mwc*'s and *twc* among which are the linear dependence between the *mwc*'s and linear correlations between the *mwc*'s and *twc*, the spectral decomposition of *mwc*'s, and various connections between the walk counts and the eigenvalues and eigenvectors of the molecular graph. We also characterize the graphs possessing minimal and maximal walk counts.

INTRODUCTION

The study of walks in molecular graphs, or more exactly, of their counts, has a long tradition in theoretical chemistry. The fact that the number of self-returning walks of length *k* is equal to the *k*th spectral moment of the respective molecular graph is of importance in the theory of total π-electron energy.[1−6] (For an exhaustive bibliography see refs 5 and 6.)

The sequence of the numbers of self-returning walks of length *k*, $k = 1, 2, ...$, starting at a given vertex *i*, was proposed by Randić[7] for characterization of the environment of vertex *i*. For later elaboration of this idea see refs 8−10.

Counts of walks starting at a particular atom are encountered in chemical documentation; they are identical to the extended connectivities obtained by the Morgan algorithm and used for discriminating and numbering the atoms in an organic molecule.[11−15]

Walk counts play a significant role in algorithms for the perception of topological (graph-theoretical) symmetry.[9,16,17]

Two of the present authors defined the quantities mwc_k, called *molecular walk counts*, as the total number of walks of length *k*, and their sum from $k = 1$ to $k = n - 1$ (where *n* is the number of vertices of the molecular graph, see below), called *total walk count*.[13] These descriptors of molecular structure were found to be suitable in QSPR studies.[13,18] Recently[19] the concept of labyrinthicity (of a molecular graph) was put forward and *twc* was proposed as its numerical measure. Labyrinthicity, on the other hand, is closely related to the notion of molecular complexity.[20,21]

In view of all this we have undertaken studies of some basic (mathematical) properties of *mwc* and *twc*. Before we outline our findings we give a few necessary explanations and definitions.

* Corresponding author phone: +49-761-203-6034; fax: +49-761-203-5987; e-mail: ChristRekr@aol.com.
[†] University of Kragujevac.
[‡] University of Freiburg.

WALKS IN GRAPHS AND THEIR COUNTS

In this paper we are concerned with ordinary molecular graphs,[22] namely graphs in which the vertices represent non-hydrogen (usually carbon) atoms of an organic molecule and the edges covalent bonds between them. Such graphs are necessarily connected and the degree of a vertex (= number of its first neighbors) does not exceed 4. Let *G* be such a molecular graph, possessing *n* vertices, and let its vertices be labeled by 1, 2, ..., *n*.

A *walk* in the graph *G* is a sequence of vertices, say (v_0, v_1, ..., v_{k-1}, v_k), such that vertex v_0 is adjacent to vertex v_1, vertex v_1 is adjacent to vertex v_2, ..., and vertex v_{k-1} is adjacent to vertex v_k. Such a walk starts at vertex v_0, ends at vertex v_k, and has length *k*. It is not required that the vertices v_0, v_1, ... , v_{k-1}, and v_k be mutually distinct. In particular, if $v_0 = v_k$, then we speak of self-returning walks.

For instance, (1, 10, 9, 8, 7, 6, 5, 4, 3, 2), (1, 5, 6, 7, 8, 9, 10, 1, 2), (1, 5, 1, 5, 1, 10, 1, 2), (1, 2, 3, 4, 5, 1, 2), (1, 5, 1, 5, 1, 2), (1, 5, 4, 3, 2), and (1, 2, 1, 2) are walks in the graph depicted in Figure 1 (which may represent azulene or bicyclo[5.3.0]decane), all starting at vertex 1 and ending at vertex 2, being of length 9, 8, 7, 6, 5, 4 and 3, respectively. Only in the first and sixth walks all vertices are distinct.

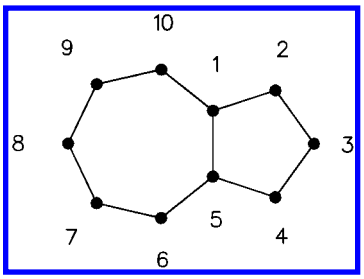

Figure 1. A molecular graph and a numbering of its vertices.

Definition 1. Let *G* be a molecular graph possessing *n* vertices. The number of all walks in *G*, having length *k*, $k \geq 1$, is called the *k*th *molecular walk count* and is denoted

10.1021/ci000149u CCC: $20.00
Published on Web 04/05/2001

by mwc_k or $mwc_k(G)$. In addition, in view of the fact that A^0 = the unit matrix, it is consistent with eq 3 to assume that $mwc_0 = n$.

Immediately from Definition 1 it follows that

$$mwc_1 = \sum_{i=1}^{n} \delta_i = 2m \tag{1}$$

where δ_i is the degree of vertex i and m the number of edges of G.

To obtain a simple formula for mwc_2 consider the walks of lengths 2 whose central vertex is vertex i. There are δ_i ways to arrive at vertex i on a walk starting at one of its neighbors, and there are δ_i ways to elongate such a walk one step further. The sum of two-step walks over all vertices is mwc_2:

$$mwc_2 = \sum_{i=1}^{n} \delta_i^2$$

Definition 2. Let G be a molecular graph possessing n vertices. The *total walk count* of G, denoted by *twc* or *twc*(G) is the sum of the first $n - 1$ *mwc* values:[23]

$$twc = \sum_{k=1}^{n-1} mwc_k \tag{2}$$

The reason why the summation is chosen to end at $k = n - 1$ is given below.

The adjacency matrix A of the graph G is by definition[22] a square matrix of order n whose (i, j)-entry is equal to unity if the vertices i and j are adjacent and is equal to zero otherwise. Bearing in mind the rules for matrix multiplication it immediately follows that

$$(A^k)_{ij} = \begin{cases} \text{number of walks of length } k \text{ in the graph } G \\ \text{starting at vertex } i \text{ and ending at vertex } j \end{cases}$$

In view of Definitions 1 and 2 we have

$$mwc_k = \sum_{i=1}^{n}\sum_{j=1}^{n} (A^k)_{ij} \tag{3}$$

and

$$twc = \sum_{k=1}^{n-1}\sum_{i=1}^{n}\sum_{j=1}^{n} (A^k)_{ij}$$

WALK COUNTS IN REGULAR GRAPHS

A graph is said to be regular if all its vertices have equal degrees. If $\delta_1 = \delta_2 = \cdots = \delta_n = r$, then the graph is said to be regular of degree r. In chemical graph theory regular graphs of degree ≤ 4 only are encountered,[22] while the case $r = 1$ is trivial. Regular graphs of degree 3 became especially important after the discovery of fullerenes.

If we start a walk at any vertex of a regular graph, then at any step we may proceed in r different directions. Consequently, for n-vertex regular graphs of degree $r > 1$

$$mwc_k = nr^k \quad \text{and} \quad twc = n\,\frac{r^n - r}{r - 1} \tag{4}$$

These formulas hold for both connected and disconnected regular graphs.

LINEAR RELATIONS AND CORRELATIONS BETWEEN *MWC*'S AND *TWC*

Denote by $\phi(G; \lambda)$ the characteristic polynomial[22] of the graph G and write it in the form

$$\phi(G; \lambda) = \lambda^n + a_1\lambda^{n-1} + a_2\lambda^{n-2} + \cdots + a_{n-1}\lambda + a_n$$

Then according to a famous result of linear algebra, the so-called Cayley−Hamilton theorem, the adjacency matrix of G obeys the identity

$$A^n + a_1A^{n-1} + a_2A^{n-2} + \cdots + a_{n-1}A + a_nI = 0$$

where I stands for the unit matrix. This, in turn, means that the nth and all higher powers of A can be recursively calculated from the first $n - 1$ powers of A. Thus,

$$A^n = -(a_1A^{n-1} + a_2A^{n-2} + \cdots + a_{n-1}A + a_nI)$$

$$A^{n+1} = -(a_1A^n + a_2A^{n-1} + \cdots + a_{n-1}A^2 + a_nA)$$

$$A^{n+2} = -(a_1A^{n+1} + a_2A^n + \cdots + a_{n-1}A^3 + a_nA^2)$$

$$\cdots \cdots \cdots$$

Applied to the molecular walk counts this implies

$$mwc_n = -(a_1mwc_{n-1} + a_2mwc_{n-2} + \cdots + a_{n-1}mwc_1 + a_nmwc_0) \tag{5}$$

$$mwc_{n+1} = -(a_1mwc_n + a_2mwc_{n-1} + \cdots + a_{n-1}mwc_2 + a_nmwc_1) \tag{6}$$

$$mwc_{n+2} = -(a_1mwc_{n+1} + a_2mwc_n + \cdots + a_{n-1}mwc_3 + a_nmwc_2)$$

$$\cdots \cdots \cdots$$

In other words, for all $k \geq n$ the mwc_k's are linearly dependent on the first $n - 1$ *mwc* values. Therefore the higher *mwc*'s do not contain any new structural information on the underlying molecular graph. Because of this it seems to be sufficient to consider only the counts of the walks of lengths less than n. This leads to the concept of total walk count as defined via eq 2.

It is worth mentioning that for all graphs (either molecular or nonmolecular) the coefficient a_1 is zero. Therefore, mwc_n is, in fact, a linear function of mwc_k, $k = 1, 2, ..., n - 2$ but is independent of mwc_{n-1}. Analogous "anomalies" hold also for the higher molecular walk counts.

In addition to the (mathematically exact) linear relations between the *mwc*'s within a graph, eqs 5, 6, etc., we empirically found remarkably good linear correlations between *twc* and the higher *mwc*'s within classes of isomeric compounds.

To illustrate this kind of mutual dependence of the walk counts, in Figure 2 the correlations between *twc* and mwc_9, mwc_8, and mwc_7, respectively, are shown for the set of 75 acyclic decanes, $C_{10}H_{22}$. The pertinent statistical data, embracing the values of k between 2 and 9, are given in

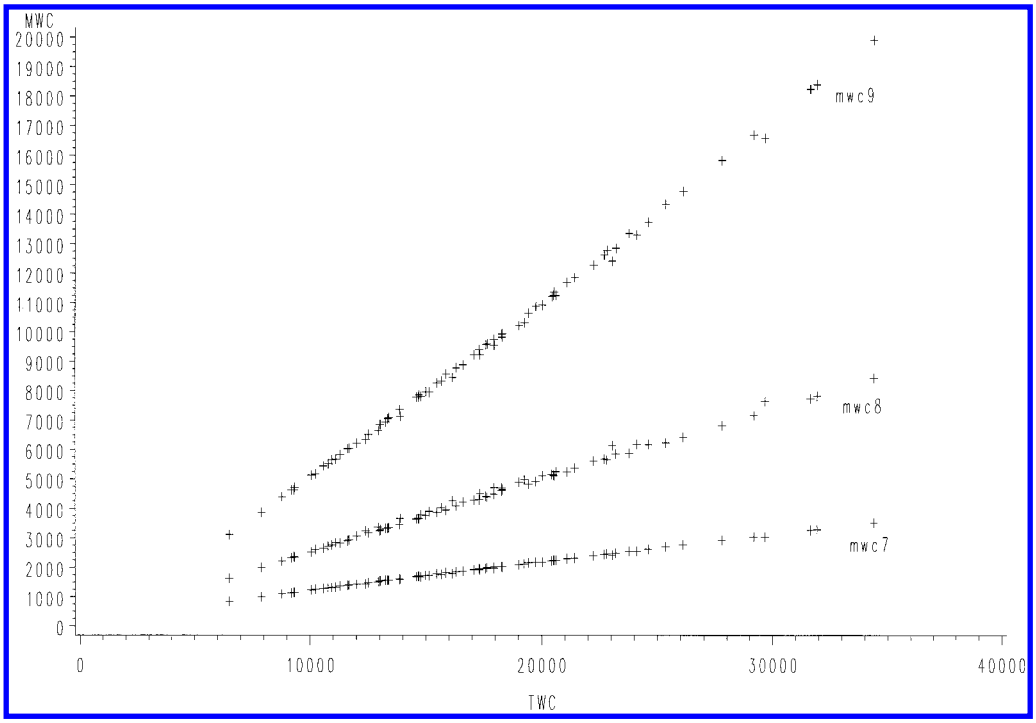

Figure 2. Correlation between the number of walks of various lengths (mwc_9, mwc_8, mwc_7) and the total walk count (twc) of the 75 decane isomers $C_{10}H_{22}$; for details see Table 1.

Table 1. (Recall that for all acyclic decanes, as a consequence of eq 1, $mwc_1 = 18$.)

The quality of the correlation between mwc_k and twc diminishes with decreasing k but only slowly and not monotonically. In the examples studied, the value of mwc_{n-1} was roughly 50% of the value of twc, whereas mwc_{n-2} was, in average, 50% of mwc_{n-1}. Anyway, the three highest walk counts (of length $n - 1$, $n - 2$, and $n - 3$) together represent some 80−90% of the total walk count.

Table 1. Statistical Data Pertaining to the Correlation of the Form $mwc_k = \alpha twc + \beta$ for the 75 Decane Isomers[a]

k	α	β	correlation coefficient	av sq. error	F
9	0.60106	−1017.4	0.9994	121.4	67187.4
8	0.24617	98.8	0.9981	91.9	19670.7
7	0.09476	289.1	0.9993	20.8	56728.5
6	0.03749	253.1	0.9935	26.3	5570.3
5	0.01356	165.5	0.9970	6.4	12261.6
4	0.00500	102.5	0.9797	6.3	1743.7
3	0.00151	57.8	0.9918	1.2	4378.9
2	0.00046	37.7	0.9055	1.3	332.5

[a] Cf. Figure 2.

GRAPHS AND MOLECULAR GRAPHS EXTREMAL WITH RESPECT TO WALK COUNTS

It is obvious that among all n-vertex graphs the complete graph K_n has maximal, and the edgeless graph O_n has minimal walk counts. Indeed, for $k \geq 1$ and for any n-vertex graph G, possessing at least one pair of adjacent vertices and at least one pair of nonadjacent vertices

$$mwc_k(O_n) < mwc_k(G) < mwc_k(K_n)$$

where $mwc_k(O_n) = 0$ and, according to eq 4, $mwc_k(K_n) = n(n-1)^k$ as K_n is an n-vertex regular graph of degree $r = n - 1$.

It should also be obvious that if e is any edge of a graph G, then for any value of k, $k \geq 1$

$$mwc_k(G - e) < mwc_k(G) \tag{7}$$

From inequality 7 it follows that the connected graph possessing minimal mwc_k must be a tree. We now show that this tree is just the path graph P_n. (Recall that P_n is the tree in which two vertices have degree one, $n - 2$ vertices have degree two, and there are no branching vertices with degrees greater than two; cf. Figure 4.)

Theorem 1. If G is a connected n-vertex graph, $n \geq 4$, different from K_n and P_n, then for all values of k, $k \geq 2$

$$mwc_k(P_n) < mwc_k(G) < mwc_k(K_n)$$

Proof. In view of relation 7 it is sufficient to demonstrate that

$$mwc_k(P_n) < mwc_k(T) \tag{8}$$

holds for any n-vertex tree, different from P_n.

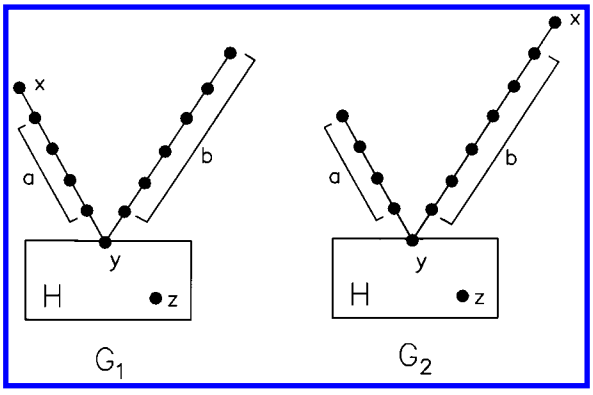

Figure 3. Graphs used in the proof of Theorem 1.

Consider the two graphs G_1 and G_2 depicted in Figure 3.

The graph G_1 is obtained from an arbitrary graph H (not required to be a tree), by attaching to its vertex y two branches, one with $a + 1$, the other with b vertices. (There may be other branches attached to vertex y, but these are considered as parts of H.) It is required that, in addition to the vertex y, the graph H possesses at least one more vertex, say z, connected to y by at least one walk. Therefore the degree of vertex y in G_1 must be at least three. Otherwise H may have an arbitrary structure. It is further required that $a < b$, that is $a + 1 \leq b$.

The graph G_2 can be viewed as obtained from G_1 by moving vertex x from the a-branch to the b-branch. Thus the a-branch in G_2 is strictly shorter than the b-branch.

We intend to demonstrate that for any $k > 1$, the graph G_1 has the same or a higher number of walks of length k than G_2. Consider therefore the difference $\Delta_k = mwc_k(G_1) - mwc_k(G_2)$.

Almost all walks in the graphs G_1 and G_2 coincide, and their counts in Δ_k cancel out. Therefore these need not to be considered any further. The only walks that differ are those involving vertex x and some vertex of H different from y, say vertex z. Now, to get from x to z a walk must first get from x to y. In G_1 this requires $a + 1$ steps, whereas in G_2 this requires $b + 1$ steps. Because $a < b$, in graph G_1 more steps remain free for walking from y to z than in G_2, resulting in a larger number of walks between x and z.

Consequently, $\Delta_k \geq 0$.

In the transformation $G_1 \rightarrow G_2$ a terminal vertex from the shorter branch is added to the longer branch, causing a decrease of the *mwc* values. Repeating this transformation several times (more precisely: $a + 1$ times), the a-branch will completely disappear, and its vertices will all be moved to the b-branch (which ultimately will have $a + b + 1$ vertices). By this the degree of vertex y will be diminished by one.

In the case of trees transformations of this kind can be continued as long as we do not arrive at the tree without vertices of degree greater than two—namely at P_n.

By this we proved inequality 8. Theorem 1 follows.

In view of eq 2 we now immediately have the following:

Corollary 1.1. The graphs O_n, K_n, and P_n, being extremal with respect to the *mwc* values (as specified above), are extremal also with respect to *twc*.

Applying the reverse construction $G_2 \rightarrow G_1$ to trees and performing it as many times as possible, we arrive at the tree with maximum *mwc*'s:

Corollary 1.2. Among n-vertex trees the star S_n (see Figure 4) has maximal *mwc* values.

Corollary 1.3. Among n-vertex trees, the trees P'_n and S'_n (see Figure 4) have second-minimal and second-maximal *mwc* values, respectively.

Corollary 1.4. The trees specified in Corollaries 1.2 and 1.3 are extremal also with respect to their *twc* values.

SPECTRAL DECOMPOSITION AND ASYMPTOTIC BEHAVIOR OF WALK COUNTS

Let, as before, G be an n-vertex molecular graph and let A be its adjacency matrix. Then the numbers $\lambda_1, \lambda_2, \ldots, \lambda_n$, and the orthonormal vectors $C_1 = (C_{11}, C_{12}, \ldots, C_{1n})^t$, $C_2 = (C_{21}, C_{22}, \ldots, C_{2n})^t$, $\ldots$, $C_n = (C_{n1}, C_{n2}, \ldots, C_{nn})^t$, satisfying

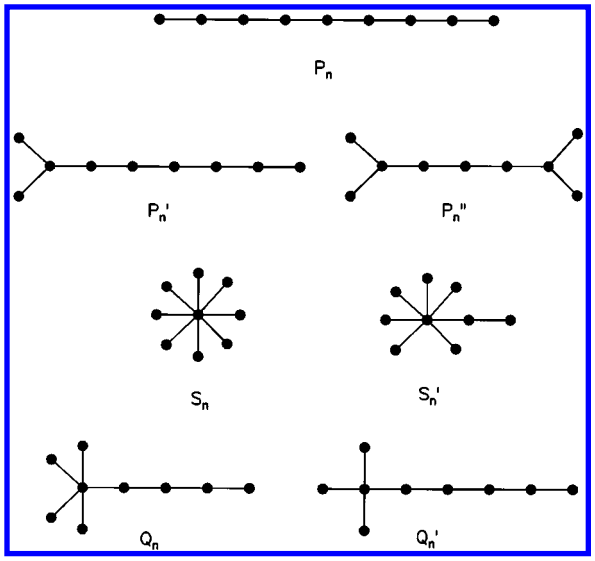

Figure 4. Trees (drawn here with $n = 9$ vertices) that are extremal with respect to their walk counts; for details see text.

the relations

$$AC_r = \lambda_r C_r\,;\ r = 1, 2, \ldots, n \tag{9}$$

are called[22,25] the eigenvalues and the eigenvectors, respectively, of the graph G. The eigenvalues are labeled so that

$$\lambda_1 > \lambda_2 \geq \cdots \geq \lambda_n$$

The eigenvector C_1, corresponding to the greatest eigenvalue λ_1, is sometimes referred to as the *principal eigenvector*.

Using routine matrix-theoretical methods, from eq 9 it follows that

$$(A^k)_{ij} = \sum_{r=1}^{n} \lambda_r^k C_{ri} C_{rj}$$

which combined with eq 3 yields

$$mwc_k = \sum_{r=1}^{n} \sigma_r \lambda_r^k \tag{10}$$

where[24]

$$\sigma_r = (C_{r1} + C_{r2} + \cdots + C_{rn})^2 \tag{11}$$

Formula 10 represents the spectral decomposition of the molecular walk counts; this result is known for some time.[15,25]

Clearly, $\sigma_r \geq 0$; some σ_r's may be equal to zero, e.g. because of symmetry. In the case of regular graphs, $\sigma_r = 0$ for all $r \geq 2$. Note that because the graph eigenvectors are normalized, for all $r = 1, 2, \ldots, n$

$$C_{r1}^2 + C_{r2}^2 + \cdots + C_{rn}^2 = 1 \tag{12}$$

An intriguing and hitherto not reported consequence of eq 10 is the following. We may rewrite eq 10 as

$$mwc_k = \sum_{+} \sigma_r \lambda_r^k + (-1)^k \sum_{-} \sigma_r |\lambda_r|^k \tag{13}$$

Table 2. Examples of Molecular Graphs Obeying (+) and Violating (−) the Inequalities 16

molecule	mwc_3/mwc_2	mwc_4/mwc_3	mwc_5/mwc_4	mwc_6/mwc_5	mwc_7/mwc_6	mwc_8/mwc_7	eq 16
2,2-dimethylpropane	1.6000	2.5000	1.6000	2.5000	1.6000	2.5000	(+)
n-octane	1.8461	1.8750	1.8667	1.8809	1.8734	1.8817	(+)
2-methylnonane	1.8889	2.0000	1.9118	2.0000	1.9231	2.0000	(+)
1,4-diethylcyclohexane	2.1818	2.2083	2.2075	2.2137	2.2135	2.2147	(+)
azulene	2.2800	2.2982	2.2977	2.3056	2.3040	2.3083	(+)
benzocyclobutadiene	2.3333	2.3469	2.3478	2.3519	2.3528	2.3541	(−)
biphenylene	2.4706	2.500	2.5143	2.5227	2.5270	2.5294	(−)

where $\sum_{+}$ and $\sum_{-}$ indicate summations over positive and negative eigenvalues, respectively. Note that both sums on the right-hand side of eq 13 are positive-valued. (Exceptionally, for regular graphs the second sum on the right-hand side of eq 13 is equal to zero.)

Now, for the counts of walks of even length we obtain

$$mwc_k = \sum_{+} \sigma_r \lambda_r^k + \sum_{-} \sigma_r |\lambda_r|^k \tag{14}$$

whereas for walks of odd length we have

$$mwc_k = \sum_{+} \sigma_r \lambda_r^k - \sum_{-} \sigma_r |\lambda_r|^k \tag{15}$$

This shows that walks of even length are somehow preferred over walks of odd length.

Eqs 14 and 15 can be understood as follows. When the parameter k is increased, then the molecular walk count increases anyway (roughly as an exponential function of k, cf. eq 19 below). However, when k is increased by one, then if k is odd, the increase of *mwc* is greater than the analogous increase when k is even. This reasoning suggests that the inequalities

$$\frac{mwc_{2h-1}}{mwc_{2h-2}} < \frac{mwc_{2h}}{mwc_{2h-1}} > \frac{mwc_{2h+1}}{mwc_{2h}} \tag{16}$$

might be satisfied.

Numerical testing revealed that the inequalities 16 are indeed satisfied by many, but not by all, molecular graphs.

In Table 2 are given the quotients mwc_k/mwc_{k-1} of a few molecular graphs that satisfy relations 16 and of some that violate them. Note that the inequalities 16 are not obeyed by regular graphs, for which the quotients mwc_k/mwc_{k-1} are independent of k and are all equal to r.

The effect is most obvious in star graphs S_n (cf. Figure 4), and at least for this narrow class of graphs it can be understood. For a star S_n there are $n-1$ one-step walks starting at the central vertex and $n-1$ one-step walks starting at one of the peripheral vertices. Each walk of the first kind can be elongated by one step in exactly one direction. Each walk of the second kind may be elongated in $n-1$ different directions. Therefore

$$mwc_1 = (n-1) + (n-1) = 2(n-1)$$

$$mwc_2 = (n-1) + (n-1)^2 = n(n-1)$$

By similar reasoning a general formula is obtained for stars:

$$mwc_k(S_n) = \begin{cases} 2(n-1)^{(k+1)/2} & \text{if } k \text{ is odd} \\ n(n-1)^{k/2} & \text{if } k \text{ is even} \end{cases}$$

Thus,

$$\frac{mwc_{k+1}}{mwc_k} = \begin{cases} n/2 & \text{if } k \text{ is odd} \\ 2(n-1)/n < 2 & \text{if } k \text{ is even} \end{cases}$$

Molecular graphs are necessarily connected and therefore their greatest eigenvalue λ_1 is nondegenerate. Furthermore, $\lambda_1 \geq |\lambda_n|$, and $\lambda_1 = |\lambda_n|$ if and only if G is bipartite.[25]

Rewriting eq 10 as

$$\frac{mwc_k}{\lambda_1^k} = \sum_{r=1}^{n} \sigma_r \left(\frac{\lambda_r}{\lambda_1}\right)^k \tag{17}$$

we now see that in the limit $k \rightarrow \infty$ all summands on the right-hand side of eq 17 will vanish, except the first and, perhaps, the last. In view of this we arrive at the following asymptotic formulas[15,25,26]

$$mwc_k(G) \sim \begin{cases} \sigma_1 \lambda_1^k & \text{if } G \text{ is nonbipartite} \\ [\sigma_1 + (-1)^k \sigma_n] \lambda_1^k & \text{if } G \text{ is bipartite} \end{cases} \tag{18}$$

Observe that relation 18 implies that the inequalities 16 must be obeyed by all bipartite graphs, provided k is sufficiently large and σ_n is nonzero.

For all connected graphs (and therefore for all molecular graphs) all components of the principal eigenvector are positive-valued.[25] Therefore, $\sigma_1 > 0$. On the other hand, σ_n is necessarily smaller than σ_1, usually much smaller and in many cases equal to zero. Anyway, the leading term in the spectral decomposition of mwc_k is the one coming from the principal eigenvector,

$$\sigma_1 \lambda_1^k \tag{19}$$

and, consequently, this term is responsible for both the main features of the structure-dependence of the molecular walk counts and for the gross part of the numerical values of the *mwc*'s. In particular, from expression 19 we see that the mwc_k's form a nearly geometric series.

The structure-dependence of λ_1 was much investigated.[26,27] For the present considerations the view[25,27] that λ_1 is the *dynamic average* of the vertex degrees seems to be particularly relevant. Under "dynamic average" is meant the average value of the vertex degrees, as seen by someone making a very long random walk in the labyrinth formed by the graph G.

Recall that the greatest eigenvalue of a regular graph of degree r is equal to r.

The dependence of σ_1 on molecular or graph structure was so far not examined. Bearing in mind eq 12 it is easy to show that σ_1 will attain its maximal value if $C_{11} = C_{12} = \cdots$

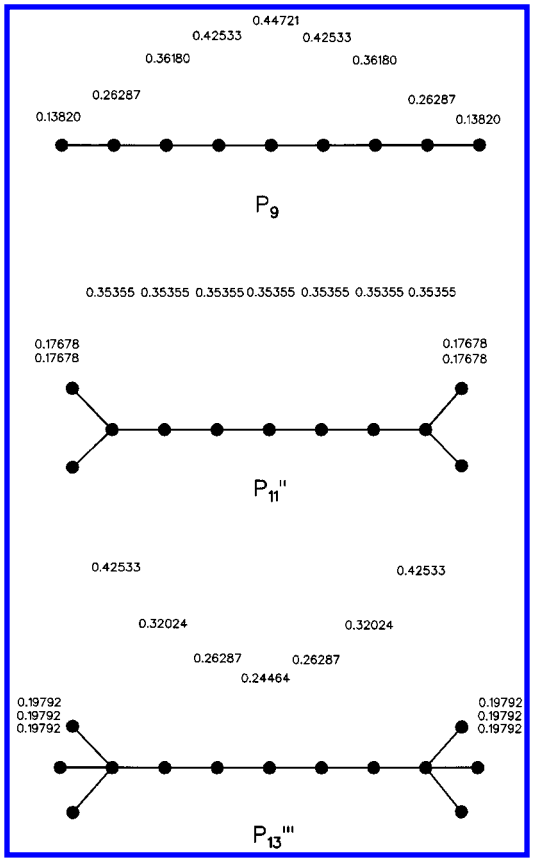

Figure 5. Coefficients of the principal eigenvector for some trees; their vertical positions indicate their magnitude; for details see text.

$= C_{1n}$. This happens in the case of regular graphs (irrespective of the value of r), for which $\sigma_1 = \max\{\sigma_1\} = n$.

This fact suggests to interpret σ_1 or σ_1/n as a measure of regularity, i.e., of equality or similarity of the coefficients C_{1j}. Such an interpretation is supported by the following reasoning.

A measure of the variability of $C_{11}, C_{12}, \ldots, C_{1n}$ is their variance, denoted here by $Var(C_1)$, that is

$$Var(C_1) = \frac{1}{n}\sum_{j=1}^{n} C_{1j}^2 - \left[\frac{1}{n}\sum_{j=1}^{n} C_{1j}\right]^2$$

$$= \frac{1}{n} - \frac{\sigma_1}{n^2}$$

Therefore, $\sigma_1 = n - n^2 Var(C_1)$ indeed is a measure of "nonvariability" or "equality" of the coefficients of the principal eigenvector C_1.

A tree with $n \geq 3$ vertices obviously cannot be regular in the strict sense. Nevertheless, we may ask the question which n-vertex tree is the "most regular" one, in the sense that all the coefficients $C_{11}, C_{12}, \ldots, C_{1n}$ have most similar values and thus maximal σ_1. An obvious candidate would be the path P_n. However, the principal eigenvector coefficients of a path graph are described by the sinus function[28]

$$C_{1j} = \sqrt{\frac{2}{n+1}} \sin\frac{j\pi}{n+1}$$

The eigenvector coefficients C_{1j} , $j = 1, 2, \ldots, n$, are shown in Figure 5 for P_9.

An approximation for σ_1 of P_n (replace summation over j by integration) is given by

$$\sigma_1(P_n) \approx \frac{2(n+1)}{\pi^2}\left(1 - \cos\frac{n\pi}{n+1}\right)^2$$

Whence,

$$\lim_{n\to\infty}\frac{\sigma_1(P_n)}{n} = \frac{8}{\pi^2} < 1 \qquad (20)$$

To enlarge the peripheral coefficients in the chain, branching may be introduced at the path's end. We thus arrive at P_n'' (cf. Figure 4), which surprisingly has all its principal eigenvector coefficients exactly the same, except for the four pendent vertices whose coefficients are exactly one-half of the former. Generally, for P_n'' it can be shown that $\lambda_1 = 2$ and that the principal eigenvector has the form

$$\frac{1}{\sqrt{n-3}}\left(\frac{1}{2}, \frac{1}{2}, 1, 1, \ldots, 1, 1, \frac{1}{2}, \frac{1}{2}\right)^t$$

So we have

$$\sigma_1(P_n'') = \frac{(n-2)^2}{n-3}$$

i.e.,

$$Var(C_1(P_n'')) = \frac{1}{n^2}\frac{n-4}{n-3}$$

We see that for increasing n, $Var(C_1(P_n''))$ rapidly tends to zero, i.e.,

$$\lim_{n\to\infty}\frac{\sigma_1(P_n'')}{n} = 1 \qquad (21)$$

i.e., $\sigma_1(P_n'')/n$ assumes the maximal allowed σ_1/n value, namely unity. At this point it is instructive to compare eqs 20 and 21.

The eigenvector coefficients of P_{11}'' are shown in Figure 5.

If peripheral branching of a chain is further enhanced, then the coefficients of the principal eigenvector become less equal again, as shown in Figure 5 for a chain P_{13}''', doubly branched at both its ends.

Anyway, a computer search for the trees with maximal σ_1 (done for n up to 10) resulted in the following:

Observation 1. For $n \geq 8$ the n-vertex tree with maximal σ_1 is P_n''. Of course, P_n'' is also the chemical tree with maximal σ_1.

The structure of the graph(s) with minimal σ_1 value is not known. We have undertaken a computer search for the trees with minimal σ_1-values and arrived at the following peculiar results:

Observation 2. Let Q_n and Q_n' be trees shown in Figure 4. For $n \geq 8$ the n-vertex tree with minimal σ_1 is Q_n (having a single branching vertex of degree five). For $n \geq 5$ the n-vertex chemical tree (i.e., tree with vertices whose degrees do not exceed 4) with minimal σ_1 is Q_n'.

It is plausible that a tree in order to possess very diverse C_{1j}'s should consist of a long unbranched chain (to provide

small C_{1j}-values) and a highly branched part (to provide large C_{1j}-values). That it is a *k*-comet graph, as we call it (defined as a star with the central vertex of degree *k* and one unbranched tail), was perhaps not to be anticipated. That $k = 5$ is the optimal choice would be even more difficult to rationalize.

It would be interesting (and remains as a challenge to mathematical chemists) to prove the above observations, or, perhaps, to show that they are violated at higher values of *n*.

Among cyclic graphs on *n* vertices, the one with minimal σ_1 is plausibly a long chain of $n - k$ vertices (to provide small C_{1j}'s), bearing at one end a highly cyclic part, a complete subgraph K_k (to provide large C_{1j}'s). Such a graph may, for obvious reasons, be called a *k*-kite. Among the kites, the 4-kite seems to possess minimal σ_1.

CONCLUDING REMARKS

In this work we outlined some previously known and proved some hitherto unknown properties of the walk counts of molecular graphs. By this we hope to have shed some more light on the recently introduced walk-based structure descriptors and help applying them in QSPR and QSAR studies as well as in the analysis of molecular complexity.

We pointed out a few open problems and made a few conjectures (stated here as "Observations"). Especially intriguing seems to be the finding that in many (but not in all) molecular graphs the counts of even walks increase faster than the counts of the odd walks. It would be interesting to get an explanation of this phenomenon.

ACKNOWLEDGMENT

I.G. thanks the Alexander von Humboldt Foundation for financial support and for enabling him to visit Germany in Fall 2000.

REFERENCES AND NOTES

(1) Hall, G. G. The Bond Orders of Alternant Hydrocarbon Molecules. *Proc. R. Soc. London* **1955**, *A 229*, 251−259.
(2) Gutman, I.; Trinajstić, N. Graph Theory and Molecular Orbitals. The Loop Rule. *Chem. Phys. Lett.* **1973**, *20*, 257−260.
(3) Hall, G. G. The Evaluation of Moments for Polycyclic Hydrocarbons. *Theor. Chim. Acta* **1986**, *70*, 323−332.
(4) Schmalz, T. G.; Živković, T.; Klein, D. J. Cluster Expansion of the Hückel Molecular Orbital Energy of Acyclics: Application to pi Resonance Theory. *Stud. Phys. Theor. Chem.* **1988**, *54*, 173−190.
(5) Gutman, I. Remark on the Moment Expansion of Total π-Electron Energy. *Theor. Chim. Acta* **1992**, *83*, 313−318.
(6) Gutman, I.; Marković, S.; Vesović, A.; Estrada, E. Approximating Total π-Electron Energy in Terms of Spectral Moments. A Quantitative Approach. *J. Serb. Chem. Soc.* **1998**, *63*, 639−646.
(7) Randić, M. Random Walks and Their Diagnostic Value for Characterization of Atomic Environment. *J. Comput. Chem.* **1980**, *1*, 386−399.
(8) Randić, M.; Woodworth, W. L.; Graovac, A. Unusual Random Walks. *Int. J. Quantum Chem.* **1983**, *24*, 435−452.
(9) Rücker, G.; Rücker, C. Isocodal and Isospectral Points, Edges, and Pairs in Graphs and How to Cope with Them in Computerized Symmetry Recognition. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 422−427.
(10) Rücker, C.; Rücker, G. Understanding the Properties of Isospectral Points and Pairs in Graphs: The Concept of Orthogonal Relation. *J. Math. Chem.* **1992**, *9*, 207−238.
(11) Razinger, M. Extended Connectivity in Chemical Graphs. *Theor. Chim. Acta* **1982**, *61*, 581−586.
(12) Razinger, M. Discrimination and Ordering of Chemical Structures by the Number of Walks. *Theor. Chim. Acta* **1986**, *70*, 365−378.
(13) Rücker, G.; Rücker, C. Counts of All Walks as Atomic and Molecular Descriptors. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 683−695.
(14) Figueras, J. Morgan Revisited. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 717−718.
(15) Rücker, C.; Rücker, G. Mathematical Relation between Extended Connectivity and Eigenvector Coefficients. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 534−538.
(16) Rücker, G.; Rücker, C. Computer Perception of Constitutional (Topological) Symmetry: TOPSYM, a Fast Algorithm for Partitioning Atoms and Pairwise Relationships among Atoms into Equivalence Classes. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 187−191.
(17) Rücker, G.; Rücker, C. On Using the Adjacency Matrix Power Method for Perception of Symmetry and for Isomorphism Testing of Highly Intricate Graphs. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 123−126.
(18) Rücker, G.; Rücker, C. On Topological Indices, Boiling Points, and Cycloalkanes. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 788−802.
(19) Rücker, G.; Rücker, C. Walk Counts, Labyrinthicity, and Complexity of Acyclic and Cyclic Graphs and Molecules. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 99−106.
(20) Nikolić, S.; Trinajstić, N.; Tolić, I. M. Complexity of Molecules. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 920−926.
(21) Nikolić, S.; Trinajstić, N.; Tolić, I. M.; Rücker, G.; Rücker, C. On Molecular Complexity Indices. In *Complexity in Chemistry*; Bonchev, D., Rouvray, D. H., Eds.; Gordon & Breach: New York, in press.
(22) Trinajstić, N. *Chemical Graph Theory*, 2nd ed. 1992; CRC Press: Boca Raton, FL.
(23) In our previous papers[13,18,19] the index *twc* was defined as the half-sum of the respective *mwc*'s. This definition took notice of the fact that most walks in a graph occur in pairs, starting at vertex *i* and ending in vertex *j*, and vice versa. However, as an exception the so-called palindromic self-returning walks, e.g., (1, 2, 1, 5, 1, 2, 1) or (1, 10, 1), are self-reverse and consequently do not occur in pairs. Therefore, it seems to be more appropriate to modify the earlier definition by leaving out the multiplier 1/2. Nevertheless, it can be shown that for all graphs *G* and for all values of *k*, $k \geq 1$, $mwc_k(G)$ is an even-valued integer, i.e., the count of the palindromic self-returning walks is always even. Thus, the above-mentioned palindromic self-returning walks have as counterparts the following: (5, 1, 2, 1, 2, 1, 5) and (10, 1, 10), respectively.
(24) If λ_r is a degenerate eigenvalue, then σ_r, as defined via eq 11, is not uniquely determined. In such cases only the sum of σ_r's over all degenerate graph eigenvectors is a true graph invariant.
(25) Cvetković, D. M.; Doob, M.; Sachs, H. *Spectra of Graphs−Theory and Application*; Academic Press: New York, 1980.
(26) Cvetković, D. M.; Gutman, I. Note on Branching. *Croat. Chem. Acta* **1977**, *49*, 115−121.
(27) Cvetković, D. M.; Rowlinson, P. The Largest Eigenvalue of a Graph: A Survey. *Lin. Multilin. Algebra* **1990**, *28*, 3−33.
(28) Cvetković, D. M.; Rowlinson, P.; Simić, S. *Eigenspaces of Graphs*; Cambridge University Press: Cambridge, 1997.

CI000149U