Chemical applications of the Laplacian spectrum of molecular graphs: Studies of the Wiener number

Ivan Gutman, Shyi-Long Lee, Chih-Hung Chu & Yeung-Long Luo

*Institute of Chemistry, Academia Sinica, Taipei, Taiwan, ROC; †Department of Chemistry, National Chung-Cheng University, Chia-Yi, Taiwan, ROC; and ‡Department of Chemistry, National Taiwan Normal University, Taipei, Taiwan, ROC

Received 17 January 1994; accepted 25 February 1994

The Wiener number (W) of an acyclic molecule can be expressed as a function of the Laplacian spectrum of the corresponding molecular graph. This algebraic relation enables inferences about the dependence of W on molecular structure. The graph invariants which determine the gross part—some 98% or more of W—are identified. As a consequence of this, an approximate formula for W is obtained which is capable of reproducing W of the molecular graphs of alkanes with an error of about 2%.

Introduction: The Wiener number and its chemical applications

In this paper we are concerned with the sum of the distances between all pairs of vertices in a molecular graph. This quantity is nowadays commonly referred to as the Wiener number or the Wiener index and is denoted by W, after Harry Wiener who first conceived it as a means to relate the physico-chemical properties of organic molecules to molecular structure.

The numerous physico-chemical and pharmacologic applications of the Wiener number as well as its basic mathematical properties are outlined in a recent review and elsewhere. Anyway, the Wiener number turns out to be one of the most frequently and most successfully employed structural descriptors that can be deduced from the molecular graph. It is nowadays well understood that the Wiener number measures the compactness of the corresponding molecule and reflects its surface-volume ratio. Consequently, W is related to the intermolecular forces, especially in the case of hydrocarbons where polar groups are absent. Those physical, chemical and pharmacologic properties of substances which may be expected to depend on the volume to surface ratio of their molecules and/or on the branching of the molecular carbon-atom skeleton, are usually well correlated with W. Among them are heats of formation, vaporization and atomization, density, boiling point, critical pressure, refractive index, surface tension, viscosity, sound velocity, liquid-liquid partition coefficients, chromatographic retention times and the consequences as well as antigens, cytostatic and estrogen-binding activities.

In view of all these applications it is somewhat surprising that the dependence of the Wiener number on molecular structure has not been studied extensively and is not completely understood. In order to contribute towards filling this gap we have undertaken a study of the Wiener number, employing the mathematical apparatus of Laplacian-graph-spectral theory, and exploiting a newly discovered connection between W and the Laplacian eigenvalues of the molecular graph.

The Laplacian spectrum of a graph

We denote the molecular graph being considered by G and its adjacency matrix by A. It is assumed that G has n vertices, i.e., that A is a square matrix of order n. The eigenvalues of A form the (ordinary) spectrum of the graph G. This spectrum has several well-known and well-elaborated chemical applications.

In this paper we are concerned with a similar, but in the chemical literature almost unknown graph spectrum, namely, the Laplacian graph spectrum.

We denote by d, the degree (= number of first neighbours) of the r-th vertex of the graph G. Let
D be the matrix whose diagonal elements are d₁, d₂, ..., dₙ and whose off-diagonal elements are all zero. Then L = D - A is the Laplacian matrix of the graph G. Its eigenvalues λ₁, λ₂, ..., λₙ₋₁, λₙ form the Laplacian spectrum of G. It will be assumed that they are labelled in a non-increasing order, i.e., λ₁ ≥ λ₂ ≥ ... ≥ λₙ₋₁ ≥ λₙ.

The Laplacian spectrum of a graph has been studied in detail by mathematicians [11-13]. For the present work the following properties of the Laplacian eigenvalues are of importance:

1° The Laplacian eigenvalues are non-negative real numbers.

2° For all graphs, λₙ = 0.

3° The (n-1)-th Laplacian eigenvalue is non-zero if and only if the respective graph is connected.

Recall that a molecular graph is necessarily connected. Hence, its Laplacian spectrum consists of a zero and n-1 positive numbers.

4° \[ \sum_{i=1}^{n-1} \lambda_i = \sum_{r=1}^{n} d_r \] ... (1)

where dᵢ is the degree of the r-th vertex.

As is well known, the right hand side of Eq. (1) is equal to twice the number of edges of the graph G. A tree (i.e., a connected acyclic graph) with n vertices has n-1 edges. Therefore, for trees Eq. (1) can be rewritten as

\[ \frac{1}{n-1} \sum_{i=1}^{n-1} \lambda_i = 2 \]

or

\[ \langle \lambda \rangle = 2 \] ... (1a)

where \( \langle \lambda^i \rangle \) denotes the arithmetic mean of the numbers \( \langle \lambda^i \rangle \), i = 1, ..., n-1. It may be recalled that the molecular graphs of alkanes are trees.

In addition to Eq. (1) the following identities hold. The way how they can be obtained is explained in the Appendix.

\[ \sum_{i=1}^{n-1} \lambda_i = \sum_{r=1}^{n} d_r + \sum_{r=1}^{n} d_r \]

... (2)

\[ \sum_{i=1}^{n-1} \lambda_i^3 = \sum_{r=1}^{n} d_r^3 + 3 \sum_{r=1}^{n} d_r^2 \]

... (3)

\[ \sum_{i=1}^{n-1} \lambda_i^4 = \sum_{r=1}^{n} d_r^4 + 4 \sum_{r=1}^{n} d_r^3 + 2 \sum_{r=1}^{n} d_r^2 - \sum_{r=1}^{n} d_r \]

... (4)

The last summation on the right-hand side of Eq. (4) goes over all pairs of adjacent vertices of G.

It may be noted that Eqs (1)-(4) hold for all trees.

The following important result seems to have been discovered in the late 1980s, independently by Russell Merris [14,15] and Brendan McKay (unpublished, referred to as private communication in [12,15,16]).

**Theorem 1.** If G is a tree with n vertices, then

\[ W = n \sum_{i=1}^{n-1} \frac{1}{\lambda_i} \]

... (5)

Using the above introduced notation, we can rewrite Eq. (5) as

\[ W = n(n-1) \langle 1/\lambda \rangle \]

... (5a)

The significance of Theorem 1 lies in the fact that the Wiener number (a graph-distance based quantity) is expressed by means of the Laplacian graph eigenvalues (that have their origin in linear algebra) [15]. By means of Eq. (5) we expect to be able to use the powerful mathematical techniques of linear algebra in studying the Wiener number.

**An approximate formula for the Wiener number of acyclic molecules**

In spite of efforts made to elucidate the relation between the Wiener number and the structure of the underlying molecule [16-23], this problem remains essentially unresolved. In this section we offer a few results that contribute towards a better understanding of the structure-dependency of W.

Bearing in mind Eq. (5a), we can approximate the Laplacian spectrum by means of a continuous spectral density operator \( \Gamma = \Gamma (\lambda) \). This is a well-elaborated technique when the (ordinary) graph spectrum is applied to chemical problems [23].

On the basis of what was mentioned about the Laplacian spectrum in the previous section, it is natural to employ a uniform-spectral-distribution approximation, namely, to choose \( \Gamma \) such that

\[ \Gamma (\lambda) = \begin{cases} \gamma & \text{if } \lambda \in [2-a, 2+a] \\ 0 & \text{otherwise} \end{cases} \]

... (6)

where a is a certain constant, whose value must lie between 0 and 2 (see Fig. 1).

The function \( \Gamma \) is chosen such that its center is at \( \lambda = 2 \). By means of this, condition (1a) is automatically obeyed. Now, from

\[ \int_{\lambda_a}^{\lambda_a+2} \Gamma d\lambda = 1 \]

... (7)

\[ 2 \]

... (8)

\[ \int_{\lambda_a}^{\lambda_a+2} \Gamma d\lambda = 1 \]

... (9)
we readily obtain \( \gamma = 1/(2a) \). Then from (5a),
\[
W = n(n-1) \int_{-a}^{a} \frac{1}{\lambda} \Gamma d\lambda
\]
which straightforwardly results in
\[
W = \frac{n(n-1)}{2a} \ln[(2+a)/(2-a)] \quad \ldots \quad (7)
\]
Formula (7), although obtained in a somewhat unorthodox way, is an exact result, provided that the parameter \( a \) depends on the underlying graph. In other words, for every graph \( G \) we can find \( a = a_G \), \( 0 < a < 2 \), such that Eq. (7) is satisfied.

We denote by \( a_p \) and \( a_s \) the \( a \)-values of the \( n \)-vertex path-graph \( P_n \) and of the \( n \)-vertex star-graph \( S_n \), respectively (see Fig. 2). Then for any other \( n \)-vertex tree \( T \), we calculate \( a_T \) by means of a Walker-type\(^{25} \) interpolation:
\[
a_T = a_p + [(I_s - I_T)/(I_p - I_T)](a_p - a_s) \quad \ldots \quad (8)
\]
where \( I_T \) is an appropriately chosen graph-invariant of \( T \), representing structural features that significantly influence the value of \( W \). An inspection of Eqs (1)-(4) suggests that this graph-invariant may be either
\[
I_T = \sum_{i=1}^{n} (d_i)^{\beta} \quad \ldots \quad (9a)
\]
or
\[
I_T = \left[ \sum_{i,j} d_i d_j \right]^\beta \quad \ldots \quad (9b)
\]
or
\[
I_T = \sum_{i,j} [d_i d_j]^\beta \quad \ldots \quad (9c)
\]
for some conveniently chosen value of \( \beta \). The first of these three options was abandoned because it would result in an approximate formula for \( W \) that predicts equal \( W \)-values for all isomers with equal vertex-degree sequences.

The combination of Eqs (7), (8) & (9b) or (7), (8) & (9c) results in an approximate formula for the Wiener number, and at the same time sheds some light on the dependence of \( W \) on molecular structure. These issues are discussed in more detail in the two subsequent sections.

However, before proceeding any further we wish to emphasize that the main goal of our study is to gain insight into the dependence of the Wiener number on molecular structure (and, consequently, the structure-dependence of intermolecular forces and various related physico-chemical properties). The primary purpose of the approximation which is elaborated in the subsequent section is to help in achieving this goal. The possibility of computing an approximate value of \( W \) by means of a simple algebraic formula should be considered just as a by-product.

**Numerical work**

The approximations for the Wiener number obtained by means of Eqs (7) and (8) are adjusted so that they reproduce the exact values of \( W \) for the unbranched path-graphs \( P_n \) and for the maximally branched star-graphs \( S_n \) (see Fig. 2). The \( a_s \) and \( a_p \) values needed in the calculations are obtained by solving the equations,
\[
W = \frac{n(n-1)}{2a_s} \ln[(2+a_s)/(2-a_s)] = (n-1)^2
\]
and
\[
W = \frac{n(n-1)}{2a_p} \ln[(2+a_p)/(2-a_p)] = \frac{1}{6} (n^3 - n).
\]
They are collected in Table 1.

The parameters \( \alpha \) and \( \beta \) in Eqs (8) & (9b) or (8) & (9c) were determined by minimizing the mean relative error over the set of all alkanes with 5 to 10 carbon atoms. For the combination (7), (8) & (9b) we found,

\[ \alpha = 0.974, \quad \beta = 2.83 \]

whereas for the combination (7), (8) & (9c),

\[ \alpha = 0.417, \quad \beta = 1.55. \]

Both approximate formulae are of comparable quality. For the set of all alkanes with \( n = 5 \) to 10, the mean relative errors are 1.45% and 1.66%, the maximal observed errors are 5.5% and 6.1%, the correlation coefficients are 0.997 and 0.998, respectively. The accuracy of the approximation (7), (8) & (9b) is only slightly diminished by rounding \( \alpha \) and \( \beta \) to their nearest-integer values, namely, \( \alpha = 1 \) and \( \beta = 3 \). Then the mean relative error is 1.48%, the maximal observed error is 5.5%, whereas the correlation coefficient remains 0.997.

Because the combination (7), (8) & (9b) gives slightly better results than (7), (8) & (9c), and because the calculation of (9b) is much easier than that of (9c), our final conclusion is that the expression

\[ \frac{n(n-1)}{2a_T} \ln\left(\frac{2 + a_T}{2 - a_T}\right) \]

is the expression that

\( \ldots (10a) \)

with

\[ a_T = a_s + \left(1 - \frac{1}{T} \right) \frac{a_p - a_s}{\left(T - T_s\right)} \]

and

\[ T_s = \left(\sum d_s d_s\right)^{1/2} \]

and with \( a_s \) and \( a_p \) being listed in Table 1, is capable of reproducing the Wiener number of any acyclic molecular graph \( T \) with an error of about 2%. It may be observed that in (10b) and (10c), \( \alpha = 1 \) and \( \beta = 3 \).

In order to further test the quality of (10) we have applied it to the four highly branched molecular graphs, depicted in Fig. 2. The results obtained (expressed via the relative errors) are given in Table 2.

For completeness we provide the explicit expressions for the \( W \)-values of the four molecular graphs considered:

\[ W(T_{1k}) = \frac{1}{2} (3k^3 + 18k^2 + 9k + 2) \quad n = 3k + 2 \]

\[ W(T_{2k}) = \frac{1}{2} (3k^3 + 15k^2) \quad n = 3k + 1 \]

\[ W(T_{3k}) = \frac{1}{2} (3k^3 + 12k^2 - 7k) \quad n = 3k \]

\[ W(T_{4k}) = \frac{1}{2} (3k^3 + 12k^2 - 9k + 2) \quad n = 3k \]

From the data presented in Table 2 we see that for highly branched alkanes the precision of the approximation (10) first increases with the increasing size of the molecule. It yields very good
results for \( n \) up to about 15, after which its error begins rapidly to increase and for \( n > 20 \) the formula loses its applicability.

Because the vast majority of alkanes of interest in chemistry has less than 20 carbon atoms, we deem that the difficulties with large values of \( n \) are of little practical relevance.

In conclusion we may say that although formula (10) looks quite complicated and although its precision is not too high, it is the most accurate approximation for the Wiener number known at present. It reproduces \( W \) with an accuracy that is satisfactory for the usual structure-property and structure-activity studies (namely, when \( W \) is correlated with experimentally determined physico-chemical or pharmacologic quantities).

**Discussion: Dependence of the Wiener number on molecular structure**

The analysis in the previous two sections has, of course, not completely resolved the problem of how the Wiener number depends on molecular structure, but, hopefully, has contributed towards its better understanding. From the first four Laplacian-spectral moments, Eqs (1)-(4), we see that the graph invariants \( \sum d_r^k \), \( k = 1,2,3,4 \) and \( \sum d_r^k \), deserve particular attention. Theorem 1 implies that these invariants may somehow influence the Wiener number, i.e., the \((-1)\)-th Laplacian-spectral moment.

The conclusion that the branching patterns of individual vertices \( (d_r) \) and of pairs of adjacent vertices \( (d_r,d_s) \) have an effect on \( W \) is in full harmony with the previous empirical and theoretical observations. The novelty of our approach is that we arrive at this dependence from a mathematics-based reasoning, and that we can establish the concrete form of this dependence. In particular, for acyclic molecular graphs we find that the graph invariant which (together with a given value of \( n \), \( n < 20 \)) determines the gross part—some 98\% or more—of \( W \) is \( \sum d_r^3 \).

**Appendix: Towards the Identities (2), (3) and (4)**

We provide here only the details of the derivation of the identity (3). The proof of (2) is much easier whereas the proof of (4) is somewhat more difficult, but they all follow the same line of reasoning.

The \( k \)-th spectral moment of the Laplacian matrix \( L \) satisfies

\[
\sum_{i=1}^{n} \lambda_i^k = \text{Tr}(L^k)
\]

where \( \text{Tr} \) stands for the trace (= sum of the diagonal elements) of the respective matrix. Because \( L = D - A \), we obtain

\[
\sum_{i=1}^{n} \lambda_i^3 = \text{Tr}(D^3) - \text{Tr}(D^2A) - \text{Tr}(DA^2) - \text{Tr}(A^2D) + \text{Tr}(D^2A) + \text{Tr}(A^2D) - \text{Tr}(A^3). \tag{11}
\]

In order to compute the traces on the right-hand side of (11) we may recall that because \( D \) is diagonal,

\[
(D^k)_{rs} = \begin{cases} 
  d_r^k & \text{if } r = s \\
  0 & \text{if } r \neq s 
\end{cases} \tag{12}
\]

where \( d_r \) is the degree of the \( r \)-th vertex. The elements of the matrix \( A^k \) are related to the structure of the corresponding graph in the following manner:

\[
(A^k)_{rs} = \begin{cases} 
  \text{number of self-returning walks of length } k, \\
  \text{starting and ending at the } r \text{-th vertex } \quad (\text{if } r = s) \\
  \text{number of walks of length } k, \text{ starting from } \\
  \text{vertex } r \text{ and ending at vertex } s \quad (\text{if } r \neq s)
\end{cases} \tag{13}
\]

It is evident from (12) that

\[
\text{Tr}(D^3) = \sum_{r=1}^{n} d_r^3. \tag{14a}
\]

That the trace of \( DA \) is zero is shown by the following observations.

\[
\text{Tr}(DA) = \sum_{r=1}^{n} (DA)_{rr} = \sum_{r=1}^{n} D_r A_r D_r. \tag{14b}
\]

Now, \( D_r \) and \( D_r \) are non-zero only if \( r = i \) and \( r = j \). But then \( A_{ij} = 0 \) and, consequently, \( D_r A_r D_r \) is zero for all possible combinations of the indices \( r,j. \) Thus,

\[
\text{Tr}(DA) = 0. \tag{14b}
\]
In the same way it follows that
\[ \text{Tr}(D^2A) = 0 \quad \text{and} \quad \text{Tr}(A D^2) = 0. \] ... (14c)

In order to compute the trace of $DA^2$ we start with
\[ \text{Tr}(D^3A) = \sum_{i=1}^{n} D_{ii}(A^2)_{ir}. \]

Then, because of (12),
\[ \text{Tr}(D A^3) = \sum_{i=1}^{n} d_{i}(A^2)_{ir}. \]

Bearing in mind (13), we see that $(A^2)_{ir}$ is equal to the number of first neighbours of the vertex $r$, i.e.,
\[ (A^2)_{ir} = d_i. \]

Therefore,
\[ \text{Tr}(D A^3) = \sum_{i=1}^{n} d_i^2. \] ... (14d)

Similarly,
\[ \text{Tr}(A D A) = \sum_{i=1}^{n} A_{ir} A_{ir} d_i. \]

The trace of $A D A$ is written as
\[ \text{Tr}(A D A) = \sum_{i=1}^{n} A_{ir} A_{ir} d_i. \]

For a given value of $i$, $A_{ir}$ and $A_{ir}$ will be non-zero whenever the vertex $r$ is a first neighbour of $i$. Hence, for a fixed value of $i$ there will be $d_i$ non-zero summands $A_{ir} A_{ir} d_i = d_i$. Their sum is, of course, $d_i^2$ and we thus have
\[ \text{Tr}(A D A) = \sum_{i=1}^{n} d_i^2. \] ... (14f)

From (13) it is apparent that $(A^3)_{ir}$ is non-zero only if the vertex $r$ lies on a triangle. For triangle-free graphs, $(A^3)_{ir} = 0$, implying
\[ \text{Tr}(A^3) = 0. \] ... (14g)

Substituting (14a-g) back into (11) we arrive at (3).

Acknowledgement
One of the authors (IG) thanks the National Science Council of the Republic of China for financial support of this research.

References
18 Bonchev D & Trinajstic N, J chem Phys, 67 (1977) 4517.