Generalized Spectral Moments of the Iterated Line Graphs Sequence. A Novel Approach to QSPR Studies

Ernesto Estrada

Department of Pharmacy, Faculty of Chemistry and Pharmacy, Universidad Central de Las Villas,
Santa Clara 54830, Villa Clara, Cuba

The spectral moments of the iterated line graph sequence (ILGS) are defined and expressed in terms of structural fragments of the molecular graphs, i.e., embedding frequencies. Several mathematical expressions to find some embedding frequencies of trees are given in terms of the spectral moments of the ILGS. This iterative method was permitted to obtain 12 embedding frequencies of alkanes with up to nine carbon atoms. These embedding frequencies and the spectral moments of the ILGS are used to describe seven important physicochemical properties of alkanes. It was proven that the spectral moments of the ILGS produce high quality QSPR models. This approach showed significant improvements in the statistical parameters of the regressions in comparison to the models obtained with the embedding frequencies. Most of the regression models obtained with the spectral moments of the ILGS can be considered as the best QSPR models reported up to now to describe these physicochemical properties of alkanes.

1. INTRODUCTION

The development of quantitative models to predict physicochemical properties of organic molecules has attracted the attention of chemists for a long time. The most primitive models were based on the additive constitutive features of some of these properties, and, in these cases, to carry out the prediction was only necessary for counting the atom types and specific atomic groups in the molecular structure.[1−4] These simple models have been very useful in chemistry, and some of them are even in use due to their good predictive features, such as the Pascal additive scheme to predict diamagnetic susceptibilities.[5] The construction of these quantitative models was based on empirical schemes supported for a great quantity of experimental data.

One of the characteristic features of the development of chemical sciences has been the increased use of mathematical tools in the development of predictive models. The use of quantum mechanics in chemistry, known as quantum chemistry, represented an important step forward in this direction. However, the use of quantum theory in chemistry has been far removed from Dirac's vision of solving chemical problems "from first principles" and its justification, as Dewar stated, "like that of any other empirical method, lies solely in its practical value".[6] Another important approach which is part of the theoretical chemistry is the application of graphs to represent chemical objects, such as molecules, reactions, synthetic routes, and so forth.[7,8] This approach has always been based on a phenomenological method of research, that is, "inventing formulae which seem to reproduce the experiments".[9] This statement of Heisenberg encloses the main objective of the search of graph theoretical representation of molecules and the models based on them to reproduce the properties which are of interest to chemists.

The structural "transparency" of the graph theoretical approaches in chemistry, i.e., the direct interpretation of results in terms of structural features of molecules, has been admitted as one of the main advantages of this approach in chemistry.[10,11] However, many of the quantitative models obtained by using graph theoretical descriptors, i.e., topological indices, do not have "transparency" in structural terms, and some of them, despite of their accuracy, have been recognized as "obscure".[12]

Thence, it is desirable to develop an approach that combines the simplicity in the interpretation of results that the additive constitutive models have, with the advantages that rigorous mathematical approaches offer. An attempt in this direction is the graph theoretical *cluster expansion* of physical and chemical properties.[13] It is based on the determination of *embedding frequencies* calculated by means of different algorithmic approaches. The use of the spectral moments of the edge adjacency matrix by Estrada[14−16] and its relationship to the embedding frequencies has represented an alternative to the calculation of these quantities, especially difficult in complex molecules, or when the number of them is very large.

The purpose of the present work is to propose an alternative approach to solve the problem of quantitative structure−property relationships (QSPR) studies maintaining a direct structural interpretation of the results. This approach is based on the calculation of spectral moments of adjacency matrices of iterated line graphs. It permits the description of physical properties very accurately and the expression of the results in terms of structural fragments similar to the cluster expansion method. The results obtained here are compared to those obtained by using the embedding frequencies, showing similar or better results in all cases. We also prove that the present method permits the calculation of a significant number of embedding frequencies of trees, mathematically.
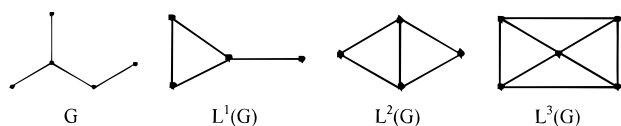
Figure 1. Molecular graph (G) representing the molecule of 2-methylbutane and its first, second, and third line graph.

2. SPECTRAL MOMENTS OF THE ITERATED LINE GRAPH SEQUENCE

Let $G = (V, E)$ be a (molecular) graph, in which $Y = \{v_i/i = 1...n\}$ is the vertex set and $E = \{e_i/i = 1...m\}$ is the edge set, and let $L(G)$ be an operator that transforms the edges e_i in the vertexes of a new graph $G = L(G)$, the line graph of G;[17,18] two vertexes of G' are adjacent if and only if the corresponding two edges in G are adjacent, i.e., incident to the same vertex. The successive application of L to the graph G produces an iterated line graphs sequence (ILGS) $L^1(G), L^2(G), ..., L^k(G)$, which corresponds to the first line graph of G, the line graph of the line graph of G, and so forth.[19,20] In Figure 1 we illustrate the first three iterated line graphs for a hydrogen-depleted tree representing the molecule of 2-methylbutane.

The spectral moment of structural matrices has been the objective of different studies for a long time. The spectral moments of the vertex adjacency matrix **A** have found great application in both physical chemistry of solid state[21−23] and in theoretical chemistry of conjugated molecules,[24−27] especially benzenoid hydrocarbons, acyclic polyenes, and phenylenes. Karwowski et al.[28] have studied the spectral moments of the Heisenberg−Hamiltonian matrix for an N-electron system. They have found a way to generate specific eigenvalues from the density distribution moments. On the other hand, Estrada has studied the spectral moments of the edge adjacency matrix, e.g. a topological bond matrix of acyclic,[14] cyclic,[16] and weighted graphs.[15] These moments have been expressed in terms of the different subgraphs of the molecular graphs, and they have been applied to the study of physicochemical properties of alkanes,[14] to the boiling points of alkyl halides,[15] to the antifungal activity of benzyl alcohols,[15] to the boiling points of cycloalkanes,[16] to the diamagnetic susceptibilities of alkanes and alkyl halides,[29] and to the sedative/hypnotic activity of a heterogeneous data set.[30]

The general theory of the spectral moments and its extension to the ILGS is explained in the following:

DEFINITION: Let $A(L^k(G))$ be the adjacency matrix of the kth line graph of the graph G, and then the spectral moments of the kth line graph are defined as follows:

$$\mu_r(L^k(G)) = tr[A(L^k(G))]^r$$

As expressed in that equation, the rth spectral moment of the kth iterated line graph is equal to the trace of the rth power of the adjacency matrix of the kth iterated line graph of G.

The spectral moments of the adjacency matrix of the first line graph have been expressed in terms of the different fragments which are subgraphs of the molecular graph. The first six of these spectral moments are given below:[14]

$$\mu_0(L^1(G)) = |F_1| \tag{1}$$

$$\mu_2(L^1(G)) = 2|F_2| \tag{2}$$

$$\mu_3(L^1(G)) = 6|F_3| \tag{3}$$

$$\mu_4(L^1(G)) = 2|F_2| + 12|F_3| + 24|F_4| + 4|F_5| \tag{4}$$

$$\mu_5(L^1(G)) = 30|F_3| + 120|F_4| + 10|F_6| \tag{5}$$

$$\mu_6(L^1(G)) = 2|F_2| + 60|F_3| + 480|F_4| + 12|F_5| + \\ 24|F_6| + 6|F_7| + 36|F_8| + 24|F_9| \tag{6}$$

Now, we will expand this approach to the spectral moments of the ILGS. The following are the expressions of these spectral moments in terms of the different fragments contained in acyclic graphs:

$$\mu_0(L^2(G)) = |F_2| \tag{7}$$

$$\mu_2(L^2(G)) = 6|F_3| + 2|F_5| \tag{8}$$

$$\mu_3(L^2(G)) = 6|F_3| + 6|F_6| + 24|F_4| \tag{9}$$

$$\mu_4(L^2(G)) = 18|F_3| + 216|F_4| + 2|F_5| + 28|F_6| + \\ 4|F_7| + 24|F_8| + 24|F_9| \tag{10}$$

$$\mu_5(L^2(G)) = 30|F_3| + 840|F_4| + 70|F_6| + 360|F_8| + \\ 200|F_9| + 10|F_{10}| + 30|F_{11}| + 120|F_{12}| \tag{11}$$

$$\mu_6(L^2(G)) = 66|F_3| + 3960|F_4| + 2|F_5| + 228|F_6| + \\ 12|F_7| + 2208|F_8| + 1056|F_9| + 48|F_{10}| + 126|F_{11}| + \\ 1944|F_{12}| + 6|F_{13}| + 36|F_{14}| + 276|F_{15}| + 84|F_{16}| + \\ 24|F_{17}| + 720|F_{18}| + 12|F_{19}| \tag{12}$$

$$\mu_0(L^3(G)) = 3|F_3| + |F_5| \tag{13}$$

$$\mu_2(L^3(G)) = 6|F_3| + 48|F_4| + 10|F_6| + 2|F_7| \tag{14}$$

$$\mu_3(L^3(G)) = 6|F_3| + 168|F_4| + 18|F_6| + 78|F_8| + \\ 48|F_9| + 6|F_{10}| + 12|F_{11}| \tag{15}$$

$$\mu_4(L^3(G)) = 18|F_3| + 1368|F_4| + 110|F_6| + 2|F_7| + \\ 780|F_8| + 440|F_9| + 36|F_{10}| + 84|F_{11}| + 384|F_{12}| + \\ 4|F_{13}| + 24|F_{14}| + 96|F_{15}| + 48|F_{16}| + 96|F_{17}| \tag{16}$$

$$\mu_0(L^4(G)) = 3|F_3| + 24|F_4| + 5|F_6| + |F_7| \tag{17}$$

$$\mu_2(L^4(G)) = 6|F_3| + 336|F_4| + 30|F_6| + 150|F_8| + \\ 96|F_9| + 14|F_{10}| + 22|F_{11}| + 2|F_{13}| \tag{18}$$

$$\mu_3(L^4(G)) = 6|F_3| + 1608|F_4| + 66|F_6| + 1482|F_8| + \\ 672|F_9| + 42|F_{10}| + 120|F_{11}| + 1464|F_{12}| + 168|F_{14}| + \\ 348|F_{15}| + 96|F_{16}| + 72|F_{17}| + 36|F_{19}| + 6|F_{20}| + \\ 18|F_{21}| + 216|F_{22}| + 144|F_{23}| + 288|F_{24}| \tag{19}$$
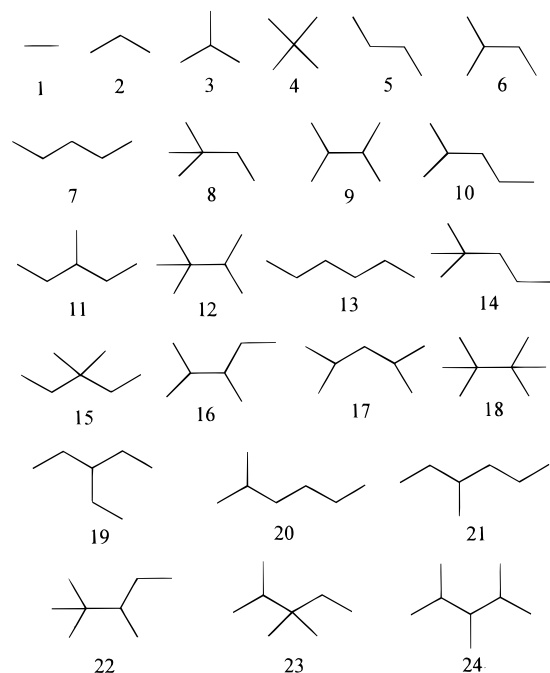
Figure 2. Some structural fragments (embedding frequencies), F_i ($i = 1, 2, ..., 24$) contained in the spectral moments of the iterated line graph sequence (ILGS).

In Figure 2 we illustrate the molecular graphs of the different fragments contained in expressions 1−19. The spectral moments of the molecular graphs, first, second, and third line graphs of the 74 C_2-C_9 alkanes have been computed, and they are given as Supporting Information.

3. CALCULATION OF EMBEDDING FREQUENCIES OF TREES

The so-called embedding frequencies are, in nonformal terms, the number of times a given substructure (cluster) appears as a subgraph in the molecular graph. These embedding frequencies have been designed here as $|F_i|$, and they are implicitly contained in the different orders spectral moments of the iterated line graph sequence. Embedding frequencies are the basis of the method of graph theoretic cluster expansion of physical and chemical properties, first introduced in the work of Smolenskii.[31] Gordon and Kennedy[32,33] proposed the idea of expressing all measurable parameters of chemical systems in terms of a linear combination of graph-theoretic invariants. Further development of cluster expansion method by Klein,[13] Poshusta and McHughes,[34,35] and more recently by Kvasnicka[36] produces an increasing interest in the calculation of embedding frequencies by different methods. All the methods developed to compute such frequencies are based on algorithmic approaches to the problem. For instance, Kvasnicka and Pospíchal developed a unified approach based on a pruning process when marginal vertexes are successively removed, solving the problem of isomorphism between trees by using the Read's linear canonical code.[36]

What we propose here is a simple method based on the spectral moments of the ILGS permitting the computation of some embedding frequencies of trees. In the present work, we will compute only the embedding frequencies of 12 subtrees. As a matter of comparison, in the paper of

Kvasnicka and Pospíchal they found the embedding frequencies of eight subtrees; all of them will be considered here.

The embedding frequencies for the first fragments are easily calculated from the spectral moment of the ILGS; the procedure is iterative as we need the knowledge of the number of some fragments to calculate others in subsequent steps of the method:

$$|F_1| = \mu_0(L^1(G)) \tag{20}$$

$$|F_2| = 1/2\mu_2(L^1(G)) \tag{21}$$

$$|F_3| = 1/6\mu_3(L^1(G)) \tag{22}$$

$$|F_5| = 1/2[\mu_2(L^1(G)) - 6|F_3|] \tag{23}$$

$$|F_5| = 1/24[\mu_4(L^1(G)) - 2|F_2| - 12|F_3| - 4|F_5|] \tag{24}$$

$$|F_6| = 1/6[\mu_3(L^2(G)) - 6|F_3| - 24|F_4|] \tag{25}$$

$$|F_7| = 1/2[\mu_2(L^3(G)) - 6|F_3| - 10|F_6| - 48|F_4|] \tag{26}$$

The computation of $|F_8|$ and $|F_9|$ requires the solution of the following system of two equations having two unknown variables

$$24|F_8| + 24|F_9| = A$$
$$36|F_8| + 24|F_9| = B$$

where A and B are given by the following expressions:

$$A = \mu_4(L^2(G)) - 2|F_5| - 18|F_3| - 4|F_7| - 28|F_6| - 216|F_4|$$

$$B = \mu_6(L^1(G)) - 2|F_2| - 60|F_3| - 480|F_4| - 12|F_5| - 24|F_6| - 6|F_7|$$

and we obtain the expressions for calculating $|F_8|$ and $|F_9|$ which are given below:

$$|F_8| = 1/12[\mu_6(L^1(G)) - \mu_4(L^2(G)) - 2|F_2| - 42|F_3| - 264|F_4| - 10|F_5| - 4|F_6| - 2|F_7|] \tag{27}$$

$$|F_9| = 1/24[\mu_4(L^2(G)) - 2|F_5| - 18|F_3| - 4|F_7| - 28|F_6| - 216|F_4| - 24|F_8|] \tag{28}$$

To obtain similar expressions for the calculation of $|F_{10}|$, $|F_{11}|$, and $|F_{12}|$ we need to solve the following system of three equations

$$14|F_{10}| + 14|F_{11}| + 84|F_{12}| = C$$
$$10|F_{10}| + 30|F_{11}| + 120|F_{12}| = D$$
$$6|F_{10}| + 12|F_{12}| = E$$

Table 1. Statistical Parameters of the Regression Models Obtained by Using the Embedding Frequencies of Trees to Describe Seven Physicochemical Properties of Alkanes

fragment	BP	MV	MR	HV	TC	PC	ST
$|F_1|$	59.163	15.463	4.792	11.267	167.57	−4.866	10.230
$|F_2|$	−18.536	1.702		−4.149	−89.55		−6.744
$|F_3|$	6.145		0.028		40.62		3.374
$|F_4|$		−2.019	−0.160	−2.006	−32.94	1.054	−1.605
$|F_5|$		−1.782		3.526		0.475	
$|F_6|$			0.018	2.197	35.22		2.161
$|F_7|$	−9.906	0.799	0.019	−0.155	−10.95	0.757	−0.335
$|F_8|$	−0.802	0.207	−0.044	−2.102	−32.33	0.476	−2.195
$|F_9|$	−1.102			−1.510	−32.25	0.581	−1.836
$|F_{10}|$	3.849	−0.519	−0.0015	0.059	4.46	−0.171	0.170
$|F_{11}|$	2.755	−0.093			2.91	−0.356	0.046
$|F_{12}|$	2.655	−0.181	0.020	1.011	22.41	−1.439	1.418
intercept	−138.235	51.135	63.859	−1.784	−132.25	50.791	−0.320
R	0.9976	0.9997	0.99997	0.9996	0.9947	0.9833	0.9906
s	3.39	0.433	0.043	0.159	6.392	0.841	0.288
F	1506	11955	128499	7683	526	207	268

where C, D, and E are as follows:

$$C = \mu_7(L^1(G)) - 126|F_3| - 1680|F_4| - 84|F_6| - 210|F_8| - 112|F_9|$$

$$D = \mu_5(L^2(G)) - 30|F_3| - 840|F_4| - 70|F_6| - 360|F_8| - 200|F_9|$$

$$E = \mu_3(L^3(G)) - 6|F_3| - 168|F_4| - 18|F_6| - 78|F_8| - 48|F_9|$$

When this system of equations is solved, we obtain the following expressions for the number of fragments $|F_{10}|$, $|F_{11}|$, and $|F_{12}|$:

$$|F_{11}| = (1/18)[E + (3/5)(D - (10/7)C)] \qquad (29)$$

$$|F_{10}| = -(1/10)[D - (10/7)C - 10|F_{11}|] \qquad (30)$$

$$|F_{12}| = (1/84)[C - 14|F_{10}| - 14|F_{11}|] \qquad (31)$$

The embedding frequencies $|F_1| \rightarrow |F_{12}|$ have been calculated here for all alkanes up to nine carbon atoms. The calculations were carried out by implementing formulas 20−31 on a Microsoft Excel workbook, and the results were compared to those reported in the literature. The complete list of these embedding frequencies should be requested from the author by e-mail. The number of embedding frequencies calculated here has been limited by the order of the line graph obtained in the ILGS. However, it is possible that if the number of iterations is increased in the ILGS, maybe a greater number of embedding frequencies can be obtained.

4. APPLICATIONS TO QUANTITATIVE STRUCTURE−PROPERTY STUDIES

It is a well-known fact that the embedding frequencies are very useful graph-theoretical descriptors in the development of quantitative structure−property and structure−activity relationships. The main advantage of this approach, i.e., the cluster expansion, is concerned not only with the statistical quality of the QSPR and QSAR models obtained but also with the direct structural interpretation of the results. The last is maybe the more attractive quality of the use of graph-theoretic methods in chemistry. These qualities are also observed when the spectral moments of the edge adjacency matrix have been used in QSPR and QSAR studies.

To show the possibilities of the spectral moments of the ILGS approach to describe physicochemical properties of organic compounds we selected seven of these properties reported for a data set of 74 alkanes. This data set, first compiled by Needham et al.,[39] has been widely used in the validation of QSPR applications of novel graph-theoretic approaches. The properties studied here are boiling points (BP), molar volume at 20 °C (MV), molar refractivity at 20 °C (MR), heat of vaporization at 25 °C (HV), critical temperature (TC), critical pressure (PC), and superficial tension at 20 °C (ST).

In the first part of this study we developed QSPR models to describe these physicochemical properties using the embedding frequencies calculated in the previous section. The best linear regression models obtained by the stepwise search are given in Table 1. The statistical parameters of the regression models, also given in Table 1, are the regression coefficient R, the standard deviation of the regression s, and the Fisher ration F.

The number of variables included in the QSPR models given in Table 1 are different for the different physicochemical properties studied. This is a consequence of the nonsignificant contribution that some of these embedding frequencies do to the respective experimental property. For instance, the inclusion of the $|F_4|$, $|F_5|$ and $|F_6|$ in the model describing the boiling point of alkanes produces no significant improvement in the quality of the QSPR model as measured by the statistical parameters used. All these models, with the unique exception of the model describing critical pressure, have correlation coefficients greater than 0.99. This requisite has been claimed as an indicator of the quality which is desired for QSPR models.

To apply the spectral moments of the ILGS in describing the seven physicochemical properties of alkanes studied, we calculated the first 10 spectral moments of the first, second, third, and fourth line graph of the molecular graphs. Consequently, the total number of variables, i.e., spectral moments, to be used is 40. However, when the regression models were to be developed, we limited the number of variables to be included in them up to 12. This number was

Table 2. Spectral Moments of the Iterated Line Graph Sequence (ILGS) Included in the Quantitative Models Describing Seven Physicochemical Properties of Alkanes

	BP	MV	MR	HV	TC	PC	TC
1	$\mu_0(L^1(G))$	$\mu_0(L^1(G))$	$\mu_0(L^1(G))$	$\mu_0(L^1(G))$	$\mu_0(L^1(G))$	$\mu_2(L^1(G))$	$\mu_0(L^1(G))$
2	$\mu_5(L^1(G))$	$\mu_2(L^1(G))$	$\mu_2(L^1(G))$	$\mu_2(L^1(G))$	$\mu_3(L^1(G))$	$\mu_5(L^1(G))$	$\mu_3(L^1(G))$
3	$\mu_6(L^1(G))$	$\mu_3(L^1(G))$	$\mu_4(L^1(G))$	$\mu_3(L^1(G))$	$\mu_5(L^1(G))$	$\mu_7(L^1(G))$	$\mu_6(L^1(G))$
4	$\mu_7(L^1(G))$	$\mu_5(L^1(G))$	$\mu_8(L^1(G))$	$\mu_4(L^1(G))$	$\mu_7(L^1(G))$	$\mu_8(L^1(G))$	$\mu_{10}(L^1(G))$
5	$\mu_9(L^1(G))$	$\mu_{10}(L^1(G))$	$\mu_9(L^1(G))$	$\mu_8(L^1(G))$	$\mu_8(L^1(G))$	$\mu_9(L^1(G))$	$\mu_2(L^1(G))$
6	$\mu_4(L^1(G))$	$\mu_6(L^1(G))$	$\mu_{10}(L^1(G))$	$\mu_9(L^1(G))$	$\mu_2(L^1(G))$	$\mu_2(L^1(G))$	$\mu_4(L^1(G))$
7	$\mu_7(L^1(G))$	$\mu_7(L^1(G))$	$\mu_5(L^1(G))$	$\mu_3(L^1(G))$	$\mu_4(L^1(G))$	$\mu_3(L^1(G))$	$\mu_{10}(L^1(G))$
8	$\mu_3(L^1(G))$	$\mu_3(L^1(G))$	$\mu_3(L^1(G))$	$\mu_4(L^1(G))$	$\mu_5(L^1(G))$	$\mu_4(L^1(G))$	$\mu_2(L^1(G))$
9	$\mu_4(L^1(G))$	$\mu_5(L^1(G))$	$\mu_4(L^1(G))$	$\mu_7(L^1(G))$	$\mu_3(L^1(G))$	$\mu_6(L^1(G))$	$\mu_3(L^1(G))$
10	$\mu_3(L^1(G))$	$\mu_7(L^1(G))$	$\mu_6(L^1(G))$	$\mu_{10}(L^1(G))$	$\mu_7(L^1(G))$	$\mu_8(L^1(G))$	$\mu_7(L^1(G))$
11	$\mu_4(L^1(G))$	$\mu_3(L^1(G))$	$\mu_2(L^1(G))$	$\mu_{10}(L^1(G))$	$\mu_6(L^1(G))$	$\mu_2(L^1(G))$	$\mu_8(L^1(G))$
12	$\mu_6(L^1(G))$	$\mu_6(L^1(G))$	$\mu_5(L^1(G))$			$\mu_{10}(L^1(G))$	$\mu_{10}(L^1(G))$

Table 3. Statistical Parameters of the Regression Models Obtained by Using Spectral Moments of the Iterated Line Graph Sequence to Describe Seven Physicochemical Properties of Alkanes

	BP	MV	MR	HV	TC	PC	ST
1	83.305	40.467	4.145	17.360	64.132	-3.366	0.819
2	-0.26	-12.268	8.39×10^{-2}	-5.574	13.532	0.819	-0.455
3	-2.693	5.329	2.80×10^{-1}	1.769	-13.155	-0.259	-1.20×10^{-2}
4	0.967	-0.134	-3.25×10^{-3}	-0.194	3.119	7.46×10^{-2}	1.04×10^{-4}
5	-1.23×10^{-2}	-4.28×10^{-4}	4.40×10^{-4}	-3.35×10^{-3}	-0.888	-5.55×10^{-3}	0.622
6	0.299	7.26×10^{-3}	4.75×10^{-1}	1.61×10^{-4}	7.90×10^{-2}	-0.441	3.13×10^{-2}
7	-2.66×10^{-3}	-8.31×10^{-4}	2.71×10^{-3}	0.170	-0.288	0.617	3.48×10^{-7}
8	2.364	-7.16×10^{-2}	-2.06×10^{-2}	1.56×10^{-2}	4.79×10^{-2}	-7.71×10^{-2}	-0.237
9	-0.212	-2.09×10^{-3}	-9.09×10^{-4}	5.91×10^{-5}	-0.474	6.93×10^{-3}	3.70×10^{-2}
10	-0.244	2.28×10^{-5}	1.98×10^{-5}	-7.74×10^{-7}	-1.45×10^{-4}	-1.63×10^{-4}	-1.00×10^{-5}
11	1.88×10^{-2}	1.12×10^{-2}	1.78×10^{-2}	1.16×10^{-9}	1.94×10^{-5}	-1.17×10^{-2}	3.99×10^{-9}
12	-1.19×10^{-5}	-2.09×10^{-6}	-3.86×10^{-6}			2.95×10^{-12}	-4.38×10^{-12}
intercept	-203.04	26.553	6.879	-6.482	-35.884	48.561	11.673
R	0.9993	0.9999	0.9999	0.9996	0.9995	0.9949	0.9913
s	1.642	0.307	0.039	0.158	2.069	0.480	0.280
F	3643	17864	105284	7092	4654	490	260

selected because it was the maximum number of embedding frequencies used in the present study. The spectral moments included in the QSPR models developed for the seven properties studied are given in Table 2.

The coefficients for each variable in the models as well as the statistical parameters of the regression equations are given in Table 3. As can be observed in this table, all models found have regression coefficients greater than 0.99. There are significant improvements in the quality of the regression models, as measured by the standard deviation of the regression, when these models are compared to those obtained with embedding frequencies. The most important improvements are observed for the critical temperature, for which the standard deviation is decreased three times with respect to the model using embedding frequencies. Two other very significant improvements are reported by the boiling points and critical pressure with reduction of more than two times in their respective standard deviations.

It is not unimportant the fact that the present approach permits a direct structural interpretation of the results in terms of structural features of molecules. The substitution of eqs 1−19 or their similar for other spectral moments in the regression models reported to describe the physicochemical properties of alkanes produces QSPR models which directly relate to the experimental properties of the molecular structure of alkanes in a similar way as the models obtained with the embedding frequencies. This kind of comparison will be done in a forthcoming paper studying the sum of ^{13}C NMR chemical shifts in alkanes.[40]

5. CONCLUDING REMARKS

Two important conclusions can be obtained from the present study. First, it has been proven that significant structural information is contained in the iterated line graph sequence, especially when the spectral moments of them are used as structural invariant. This finding, together with previous results reported by us, open novel possibilities in the search of graph-theoretical invariants which can be used in the development of high quality QSPR and QSAR models with structural interpretation. On the other hand, we have illustrated how this novel invariant, i.e., spectral moments of the ILGS, can be used in the generation of some embedding frequencies of trees, which are regularly obtained from algorithmic manipulation of a graph. The possible extension of this approach to find embedding frequencies of a greater number of subtrees, and maybe of cyclic fragments, by using higher order line graphs in the ILGS, is an open question that requires future studies.

The second but not least important conclusion is related to the use of spectral moments of the ILGS in QSPR studies. We have shown that this approach produces high quality QSPR models to describe seven important physicochemical properties of alkanes. These models are, in most cases, the best QSPR models obtained for such properties by using graph-theoretic descriptors. However, the most important fact from the practical point of view is that the use of the spectral moments of the ILGS represents an alternative to the use of embedding frequencies of trees as molecular descriptors in QSPR and QSAR studies. Spectral moments

Spectral Moments of the Iterated Line Graphs Sequence

J. Chem. Inf. Comput. Sci., Vol. 39, No. 1, 1999 **95**

are easy to compute structural invariants, and, as we have proven here, they produce similar or better QSPR models than the embedding frequencies, with no loss of structural information in the interpretation of results.

We have shown here, one more time, that the study of algebraic properties of graph-theoretic representation of molecules represents an important and not completely explored source of molecular descriptors. The route of "inventing formulas which seem to reproduce the experiments", always followed in phenomenological chemical graph theory, has proven to be not only an important alternative in QSPR and QSAR studies but also an efficient way of doing so, making the complex simple, the simple customary, and the customary pleasant.

Supporting Information Available: Tables of the spectral moments of the graphs, first, second, and third line graphs of the 74 C_2–C_9 alkanes (8 pages). See any current masthead page for ordering information.

REFERENCES AND NOTES

(1) Partington, J. R. *An Advanced Treatise on Physical Chemistry. Vol. IV*; Longmans, Green and Co.: London, 1953.
(2) Tatevskii, V. M.; Benderskii, V. A.; Yarovoi, S. S. *Rules and Methods for Calculating the Physicochemical Properties of Paraffinic Hydrocarbons*; Pergamon Press: New York, 1961.
(3) Lyman, W. J.; Reehl, W. F.; Rosenblatt, D. H. *Handbook of Chemical Property Estimation Methods*; McGraw-Hill: New York, 1982.
(4) Somayajulu, G. R.; Zwolinski, B. J. Isomeric Variation Procedures for Physicochemical Properties of Alkanes. *J. Phys. Chem.* **1966**, *70*, 3498–3501.
(5) Sellwood, P. W. *Magnetochemistry*; Intersciences Pub.: New York, 1956.
(6) Dewar, M. J. S. Quantum Mechanical Molecular Models. *J. Phys. Chem.* **1985**, *89*, 2145–2150.
(7) Balaban, A. T., Ed.; *Chemical Applications of Graph Theory*; Academic Press: London, 1976.
(8) Trinajstić, N. *Chemical Graph Theory*; CRC Press: Boca Raton, FL, 1992.
(9) Heisenberg, W. Theory, Criticism and a Philosophy. In *From a Life of Physics*; Salam, A., Ed.; World Scientific Pu.: Singapore, 1989.
(10) Randić, M. The Nature of Chemical Structure. *J. Math. Chem.* **1990**, *4*, 157–184.
(11) Randić, M. On Characterization of Chemical Structure. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 672–687.
(12) Mihalić, Z.; Trinajstić, N. A Graph-Theoretical Approach to Structure–Property Relationships. *J. Chem. Educ.* **1992**, *69*, 701–712.
(13) Klein, D. J. Chemical Graph-Theoretic Cluster Expansion. *Int. J. Quantum Chem.: Quantum Chem. Symp.* **1986**, *69*, 701–712.
(14) Estrada, E. Spectral Moments of the Edge adjacency Matrix in Molecular Graphs. 1. Definition and Application to the Prediction of Physical Properties of Alkanes. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 844–849.
(15) Estrada, E. Spectral Moments of the Edge adjacency Matrix in Molecular Graphs. 2. Molecules Containing Heteroatoms and QSAR Applications. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 320–328.
(16) Estrada, E. Spectral Moments of the Edge adjacency Matrix in Molecular Graphs. 3. Molecules Containing Cycles. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 23–27.
(17) Harary, F. *Graph Theory*; Adison-Wesley: Reading, MA, 1969; Chapter 8.
(18) Bertz, S. H. Branching in Graphs and Molecules. *Discr. Appl. Math.* **1988**, *19*, 65–83.
(19) Gutman, I.; Popovic, L.; Estrada, E.; Bertz, S. H. The Line Graph Model. Predicting Physico-Chemical Properties of Alkanes. *ACH Models Chem.* **1998**, *135*, 147–155.
(20) Estrada, E.; Guevara, N.; Gutman, I.; Rodríguez, L. Molecular Connectivity Indices of Iterated Line Graphs. A New Source of Descriptors for QSPR and QSAR Studies. *SAR QSAR Environ. Res.* **1998**, *9*, 229–240.
(21) Burdett, J. K.; Lee, S. Moments and Energies of Solids. *J. Am. Chem. Soc.* **1985**, *107*, 3050–3063.
(22) Burdett, J. K.; Lee, S. The Moments Method and Elemental Structures. *J. Am. Chem. Soc.* **1985**, *107*, 3063–3082.
(23) Lee, S. Second-Moment Scaling and Covalent Crystal Structures. *Acc. Chem. Res.* **1991**, *24*, 249–254.
(24) Jiang, Y.; Tang, A.; Hoffmann, R. Evaluation of Moments and Their Application in Hückel Molecular Orbital Theory. *Theor. Chim. Acta* **1984**, *66*, 183–192.
(25) Jiang, Y.; Qian, X.; Shao, Y. The Evaluation of Moments for Benzenoid Hydrocarbons. *Theor. Chim. Acta* **1995**, *90*, 135–144.
(26) Markovic, S.; Gutman, I. Dependence of Spectral Moments of Benzenoid Hydrocarbons on Molecular Structure. *J. Mol. Struct. (THEOCHEM)* **1991**, *235*, 81–87.
(27) Markovic, S.; Stajkovic, A. The Evaluation of Spectral Moments for Molecular Graphs of Phenylenes. *Theor. Chim. Acta* **1997**, *96*, 256–260.
(28) Karwowski, J.; Bielinska-Waz, D.; Kurkowski, J. Eigenvalues of Model Hamiltonian Matrices from Spectral Density Distribution Moments: The Heisenberg Spin Hamiltonian. *Int. J. Quantum Chem.* **1996**, *60*, 185–193.
(29) Estrada, E. Modelling the Diamagnetic Susceptibility of organic Compounds by a Sub-structural Graph-theoretical Approach. *J. Chem. Soc., Faraday Trans.* **1998**, *94*, 1407–1410.
(30) Estrada, E.; Peña, A.; García-Domenech, R. Designing Sedative/Hypnotic Compounds from a Novel Substructural Graph-Theoretical Approach. *J. Comput.-Aided Mol. Des.* In press.
(31) Smolenskii, E. A. Utilization of the Theory of Graphs fro the Calculation of Structure-Additive Properties of Hydrocarbons. *Zh. Fiz. Khim.* **1964**, *38*, 1288–1291.
(32) Gordon, M.; Kennedy, J. W. The Graph-like State of Matter. Part 2. LCGI Schemes for the Thermodynamics of Alkanes and the Theory of Inductive Inference. *J. Chem. Soc., Faraday Trans. II* **1973**, *69*, 484–504.
(33) Essam, J. W.; Kennedy, J. W.; Gordon, M. The Graph-like State of Matter. Part 8. LCGI Schemes and the Statistical Analysis of Experimental Data. *J. Chem. Soc., Faraday Trans. II* **1977**, 73, 1289–1307.
(34) Poshusta, R. D.; McHughes, M. C. Embedding Frequencies of Trees. *J. Math. Chem.* **1989**, *3*, 193–215.
(35) McHughes, M. C.; Poshusta, R. D. Graph-theoretic Cluster Expansions. Thermochemical Properties of Alkanes. *J. Math. Chem.* **1990**, *4*, 227–249.
(36) Kvasnicka, V.; Pospíchal, J. Simple Construction of Embedding Frequencies of Trees and Rooted Trees. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 121–128.
(37) Schmalz, T. G.; Klein, D. J.; Sandleback, B. L. Chemical Graph-Theoretical Cluster Expansion and Diamagnetic Susceptibility. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 54–57.
(38) Cherqaoui, D.; Villemin, D.; Kvasnicka, V. Application of Neural Network Approach for Prediction of some Thermochemical Properties of Alkanes. *Chemometrics Intell. Lab. Syst.* **1994**, *24*, 117–128.
(39) Needham, D. E.; Wei, I.-C.; Seybold, P. G. Molecular Modeling of the Physical Properties of Alkanes. *J. Am. Chem. Soc.* **1988**, *110*, 4186–4194.
(40) Estrada, E.; de Julián-Ortíz, J. V. Work in progress.