communications in mathematical
match
and in computer chemistry

no. 44, October 2001

MATCDY (44) 135 - 146 (2001)

ISSN 0340 - 6253

PATH LAYER MATRIX FOR WEIGHTED GRAPHS

Andrey A. Dobrynin and Leonid S. Mel'nikov*

Sobolev Institute of Mathematics, Siberian Branch of the
Russian Academy of Sciences, Novosibirsk 630090, Russia
E-mail: dobr@math.nsc.ru, omeln@math.nsc.ru

Abstract. The path layer matrix of a weighted graph G counts the numbers of paths with given length coming from vertices of G. If $l_1 < l_2 < \ldots < l_n$ are all possible lengths of paths in G, then the entry (i, j) of this matrix is the number of simple paths having initial vertex v_i and length l_j. This matrix and derived topological indices can be applied for establishing the similarity of molecular graphs. Examples of graphs without cut-vertices having the same path layer matrices are presented.

INTRODUCTION

Matrices associated with graphs are important tools for designing and computing topological indices of molecular graphs. Some of these matrices naturally arise from considering distances between vertices of a graph. The most well-known distances are based on the shortest or longest paths in a graph. Among matrices of this kind we point out the distance matrix, the detour matrix and the (distance) layer matrix. More specific matrices are the path layer matrix, the Wiener matrix, the Cluj matrix, etc.

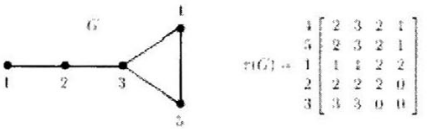

FIGURE 1. Path layer matrix for a simple graph.

Dedicated to Professor A. T. Balaban for his 70th aniversary.

The path layer matrix was introduced for a simple graph with the standard metric. Denote by $p(G)$ the *order* of a graph G, i. e., the number its vertices. The *path layer matrix* of a graph G is the matrix $\tau(G) = \|\tau_{ij}\|$, $i = 1,2,\ldots, p(G)$ and $j = 1,2,\ldots, p(G)-1$, where τ_{ij} is the number of all simple paths with initial vertex v_i that have length j. By ordering the rows of $\tau(G)$ with respect to the decrease of their length (the number of the last nonzero element) and then by lexicographically arranging the rows with the same length, one can obtain a canonical form of $\tau(G)$. An example of canonical path layer matrix for a simple graph is shown in Fig. 1.

The path layer matrix also known as the *path degree sequence* of a graph or the *atomic path code* of a molecule [1, 2]. This matrix and derived invariants have found interesting applications in chemistry for characterization of branching in molecules, for establishing similarity of molecular graphs, and for drug design [1–5]. For trees a path layer matrix coincides with a layer matrix (or distance degree sequence [2]), since any two vertices in a tree are connected by a unique path. Mathematical investigations of this matrix deal with finding a pair of nonisomorphic graphs having some specified property such that both graphs have the same matrix [2, 6–12]. Among these properties we point out the girth, cyclomatic number and planarity of graphs. If such a pair exists, then it is interesting to determine the least order possible for these graphs. In [1] it is asserted that an equality of the path layer matrices for graphs of order $p \le 11$ is a sufficient condition for their isomorphism. The following problem was proposed by Quintas and Slater in [6]: does there exist a pair of connected nonisomorphic r-regular graphs having the same path degree sequences (if the answer is yes, then for each $r \ge 3$, what is the least order $p(r)$ possible for graphs in such a pair)?

Balaban *et al.* presented the first example of such cubic graphs of order 142 [13]. For every $r \ge 3$, r-regular graphs with this property have been constructed in [14]. The order of the graphs is a linear function in r ($p(3) \le 116$ and $p(4) \le 114$). The upper bound for the order of cubic graphs has been improved in [15] ($p(3) \le 62$). The key feature of all similar graphs is that they contain cut-vertices. For a long time, intensive investigations have failed to produce even one pair of nonisomorphic graphs without cut-vertices that have the same path layer matrix. The following question was formulated in [15]: does there exist a pair of nonisomorphic (r-regular) graphs without cut-vertices having the same path layer matrix? Examples of simple 2-connected graphs have been reported in [16] and several families of 2- and 3-

connected graphs with this property have been presented in [17]. In the present paper this approach will be expanded for weighted graphs.

DEFINITIONS

A simple graph $G(V,E)$ is called a *weighted graph* if each edge e is assigned a non-negative real number $w(e)$, called the weight of e. An unweighted graph can be regarded as a weighted graph in which each edge e is assigned weight $w(e)=1$. Let P be a simple path of a graph. Then the length of this path, $l(P)$, is defined by the following equality:

$$l(P)=\sum_{e\in E(P)} w(e).$$

FIGURE 2. Path layer matrix for an weighted graph.

Let $l_1 < l_2 < \ldots < l_n$ be all possible lengths of simple paths in G. The *path layer matrix* of a weighted graph G is the matrix $\tau w(G)=\|\tau w_{ij}\|$, $i=1,2,\ldots,p(G)$ and $j=1,2,\ldots,n$, where τw_{ij} is the number of simple paths with initial vertex v_i that have length l_j. A canonical form of τw is the same as for τ. The weighted graph G and its matrix $\tau w(G)$ is depicted in Fig. 2. It is clear that if $w(e)$ is a constant for all edges then $\tau w(G)=w(e)\tau(G)$. The number of mutually distinct lengths in a weighted graph can be very large in real applications. In order to reduce the number of matrix's columns, we can define an equivalence relation on paths or consider paths of bounded length.

STRUCTURAL SIMILARITY OF GRAPHS

Because the matrix $tw(G)$ takes into account weights of graph's edges, it more precisely reflects chemical features of the corresponding structure than the matrix $\tau(G)$. In order to illustrate the use of path layer matrix for establishing similarity among weighted graphs, we consider molecular graphs of antihistamines diagrammed in Fig. 3 [18]. Every graph G_i, $i = 1,2,\ldots,10$, is represented by a sequence $A(G_i) = A_i = (a_1, a_2, \ldots, a_n)$, where a_j is a half-sum of the j-th column of $\tau(G_i)$ or $tw(G_i)$ and n being determined by the longest path among all graphs. This sequence is also known as the *path length distribution* of a graph.

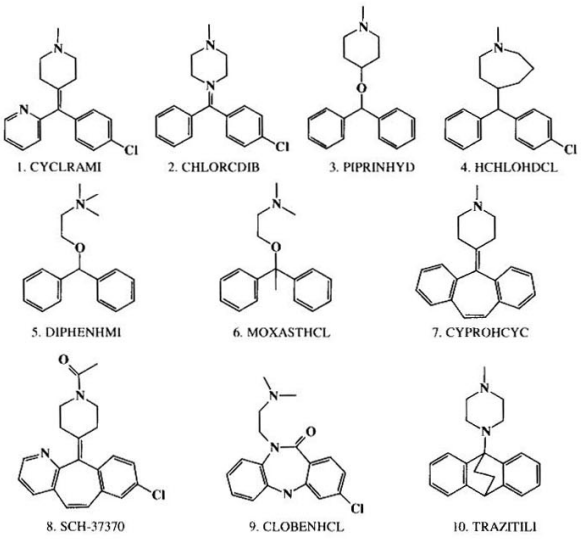

FIGURE 3. Molecular graphs of antihistamines.

At first, skeletons of these graphs will be processed. Path length distributions for the unweighted molecular graphs are collected in Table 1 (the longest path has length 20).

A measure of similarity follows from the count of differences in the number of paths of distinct length in the structures considered. Every sequence can be regarded as a vector in n-dimensional Euclidean space. Those vectors are similar which lie in the same region of the space. As a measure of similarity, we take the distance between two points A and B defined by two sequences:

$$D(A,B) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \ldots + (a_n - b_n)^2}\,.$$

TABLE 1. Path length distributions for skeletons of molecular graphs of Fig. 3.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
A_1 =	(23,	31,	40,	52,	68,	66,	76,	76,	64,	56,	40,	12,	0,	0,	0,	0,	0,	0,	0,	0)
A_1 =	(23,	31,	40,	52,	68,	66,	76,	76,	64,	56,	40,	12,	0,	0,	0,	0,	0,	0,	0,	0)
A_2 =	(23,	31,	40,	52,	68,	66,	76,	76,	64,	56,	40,	12,	0,	0,	0,	0,	0,	0,	0,	0)
A_3 =	(23,	30,	38,	48,	62,	56,	62,	68,	64,	52,	40,	28,	8,	0,	0,	0,	0,	0,	0,	0)
A_4 =	(24,	32,	41,	53,	68,	72,	78,	80,	72,	62,	50,	28,	8,	0,	0,	0,	0,	0,	0,	0)
A_5 =	(21,	28,	31,	38,	45,	42,	48,	44,	36,	28,	20,	4,	0,	0,	0,	0,	0,	0,	0,	0)
A_6 =	(21,	28,	35,	42,	49,	46,	48,	40,	32,	24,	16,	4,	0,	0,	0,	0,	0,	0,	0,	0)
A_7 =	(25,	35,	49,	68,	94,	105,	122,	140,	146,	144,	124,	114,	100,	91,	72,	56,	40,	32,	24,	8)
A_8 =	(28,	40,	55,	75,	103,	121,	145,	170,	186,	189,	173,	155,	136,	128,	105,	98,	62,	56,	44,	30)
A_9 =	(24,	34,	45,	61,	82,	97,	107,	117,	123,	125,	94,	76,	68,	58,	32,	17,	14,	7,	2,	0)
A_{10} =	(27,	41,	64,	97,	144,	174,	218,	248,	274,	262,	246,	200,	174,	140,	84,	40,	32,	24,	8,	0)

The following distance matrix D_r contains rounded distances between all unweighted structures shown in Fig. 3:

$$D_r = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \end{array} \begin{array}{c} \begin{array}{cccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{array} \\ \left[\begin{array}{cccccccccc} 0 & 0 & 27 & 24 & 72 & 75 & 266 & 393 & 169 & 543 \\ 0 & 0 & 27 & 24 & 72 & 75 & 266 & 393 & 169 & 543 \\ 27 & 27 & 0 & 32 & 62 & 66 & 266 & 394 & 171 & 546 \\ 24 & 24 & 32 & 0 & 89 & 93 & 247 & 375 & 150 & 522 \\ 72 & 72 & 62 & 89 & 0 & 11 & 317 & 444 & 226 & 603 \\ 75 & 75 & 66 & 93 & 11 & 0 & 321 & 447 & 230 & 607 \\ 266 & 266 & 266 & 247 & 317 & 321 & 0 & 131 & 107 & 301 \\ 393 & 393 & 394 & 375 & 444 & 447 & 131 & 0 & 234 & 215 \\ 169 & 169 & 171 & 150 & 226 & 230 & 107 & 234 & 0 & 378 \\ 543 & 543 & 546 & 522 & 603 & 607 & 301 & 215 & 378 & 0 \end{array}\right] \end{array}.$$

To find groups of similar structures, the taxonomy algorithm and the corresponding computer program were used [19, 20]. All structures have been divided into two taxons $T_1 = \{1,2,3,4,5,6\}$ and $T_2 = \{7,8,9,10\}$. It should be noted that these groups can be recognized by inspecting the path length distributions. The structures of the second group have many long paths.

Consider now the weighted molecular graphs. We shall use the standard lengths of chemical bonds, i. e., 0.154 for the bond $C-C$, 0.143 for $C-O$, 0.122 for $C=N$, etc. Let P_1 and P_2 be two paths with lengths $l_1 < l_2$ coming from the same vertex. In order to reduce the number of all different lengths, it is assumed that if $l_2 - l_1 < \delta$ then P_1 and P_2 are equivalent and have equal length l_1. At first, a common table of all different lengths for all structures is defined. Then the path layer matrices and path length distributions are calculated for all graphs. Table 2 contains a part of path length distributions of structures shown in Fig. 3. The number of all distinct lengths is equal to 130 for $\delta = 0.01$ and the longest path has length 2.94.

TABLE 2. Path length distributions for molecular graphs of Fig. 3.

	.12	.13	.15	.18	.28	.29	.31	.33	.41	.42	.44	.46	.56	.57	.58
A_1 = (	1,	6,	15,	1,	3,	18,	9,	1,	2,	8,	22,	8,	5,	20,	5, ...
A_2 = (	0,	6,	16,	1,	0,	20,	10,	1,	0,	6,	28,	6,	0,	19,	4, ...
A_3 = (	0,	8,	15,	0,	0,	22,	8,	0,	0,	8,	24,	6,	0,	18,	11, ...
A_4 = (	0,	6,	17,	1,	0,	17,	14,	1,	0,	6,	18,	17,	0,	15,	0, ...
A_5 = (	0,	8,	13,	0,	0,	24,	4,	0,	0,	8,	21,	2,	0,	18,	11, ...
A_6 = (	0,	8,	13,	0,	0,	22,	6,	0,	0,	8,	23,	4,	0,	18,	11, ...
A_7 = (	0,	8,	17,	0,	0,	25,	10,	0,	0,	10,	33,	6,	0,	35,	2, ...
A_8 = (	2,	7,	18,	1,	5,	22,	12,	1,	4,	10,	33,	8,	6,	33,	5, ...
A_9 = (	1,	6,	16,	1,	5,	20,	8,	1,	3,	12,	26,	4,	4,	36,	5, ...
A_{10} = (	0,	6,	21,	0,	0,	22,	19,	0,	0,	6,	41,	17,	0,	22,	5, ...

Rounded distances between all structures form the matrix D_{2w} (path length distributions were used without truncating):

$$D_{\tau w} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \end{array} \begin{array}{c} \begin{array}{cccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{array} \\ \begin{bmatrix} 0 & 51 & 54 & 86 & 39 & 39 & 175 & 198 & 89 & 238 \\ 51 & 0 & 39 & 71 & 57 & 56 & 177 & 205 & 104 & 303 \\ 54 & 39 & 0 & 76 & 52 & 53 & 176 & 205 & 105 & 307 \\ 86 & 71 & 76 & 0 & 89 & 86 & 213 & 235 & 137 & 295 \\ 39 & 57 & 52 & 89 & 0 & 10 & 192 & 218 & 112 & 343 \\ 39 & 56 & 53 & 86 & 10 & 0 & 194 & 221 & 115 & 343 \\ 175 & 177 & 176 & 213 & 192 & 194 & 0 & 96 & 129 & 282 \\ 198 & 205 & 205 & 235 & 218 & 221 & 96 & 0 & 140 & 299 \\ 89 & 104 & 105 & 137 & 112 & 115 & 129 & 140 & 0 & 300 \\ 328 & 303 & 307 & 295 & 343 & 343 & 282 & 299 & 300 & 0 \end{bmatrix} \end{array}.$$

Using the taxonomy program, the following four taxons have been obtained: $T_1 = \{1,2,3,4\}$, $T_2 = \{5,6\}$, $T_3 = \{7,8,9\}$ and $T_4 = \{10\}$. Comparing the above results of taxonomy based on the matrices τ and τw, we can conclude that the use of τw has allowed to distinguish structural similarity of molecular graphs more precisely.

A computer program for calculating the matrix τw has been developed. Its control parameters include type of weights (unit, integer or chemical), maximal difference between lengths of equivalent paths, restrictions on the longest path with respect of length or/and the number of edges.

GRAPHS WITH THE SAME PATH LAYER MATRIX

Consider a weighted graph T with three chosen vertices a, b and c shown in Fig. 4a. The graph T is said to be *admissible* if its automorphism group, $Aut(T)$, satisfies the following conditions:

]

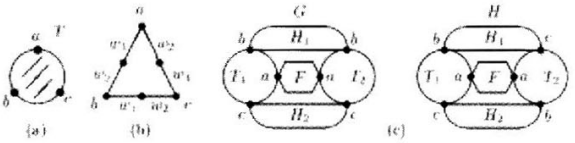

FIGURE 4. Admissible graphs and main construction.

(a) there are automorphisms that cyclically interchange vertices a, b and c in the graph T ($a \to b \to c \to a$), i. e., $Aut(T)$ contains the cyclic subgroup of order 3;

(b) there are no automorphisms fixing the vertex a and interchanging vertices b and c, i. e., $Aut(T)$ does not contain the dihedral subgroup.

An example of an admissible graph is depicted in Fig. 4b ($w_1 \neq w_2$).

Let T_1, T_2 be admissible graphs and H, F be arbitrary weighted graphs. Let graphs G_1 and G_2 be obtained from these graphs as shown in Fig. 4c (copies of H are attached with T_1 and T_2 by the same way).

THEOREM. *Graphs G_1 and G_2 are not isomorphic and $\tau w(G_1) = \tau w(G_2)$.*

It is not hard to verify that conditions (a) and (b) imply nonisomorphism between G_1 and G_2. Denote by $\tau w_G(v)$ the row of $\tau w(G)$ corresponding to the vertex $v \in V(G)$.

The proof of the theorem is similar as for the case of simple graphs and it is based on the condition (a) and the structure of graphs G_1 and G_2 [17]. This result is proven by constructing two bijection:

1) $\phi : V(G_1) \to V(G_2)$ such that $\tau w_{G_1}(v) = \tau w_{G_2}(\phi(v))$;

2) $f : \wp_{G_1}(v) \to \wp_{G_2}$, where $\wp_G(v)$ is the set of all simple paths of fixed length beginning at the vertex v in G. For details see [17].

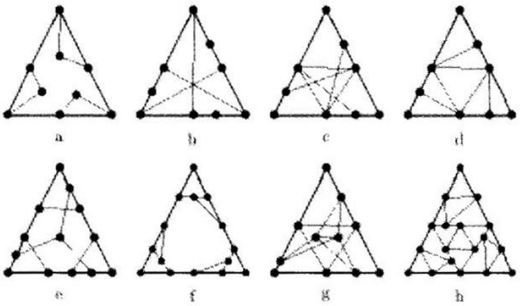

FIGURE 5. Admissible graphs for constructing regular graphs.

In order to construct graphs with given properties and the same path layer matrices, it suffices to select suitable graphs T_1, T_2 and H, F. In the simplest case, $T_1 \cong T_2$ and $H \cong F \cong P_2$, where P_2 is the path with two vertices. Now we recall some results and open questions for unweighted graphs of small degree.

PROPOSITION 1 [17]. *For every* $p \geq 18$, *there are nonisomorphic graphs of order* p *having the same path layer matrices.*

Graphs of order 18 are constructed from admissible graphs T on 9 vertices shown in Fig. 5a–d. Additional properties of T provide the corresponding properties of the obtained graphs: 2- or 3-connectedness, planarity or nonplanarity. It should be noted that the resulting graphs have not cut-vertices. To find minimal graphs, we have to answer on the following question.

QUESTION 1 [17]. *Does there exist a pair of graphs of order* p, $12 \leq p \leq 17$, *having the same path layer matrices* ?

It is known that cubic graphs are of great interest in chemistry and graph theory. For this class of graphs we have the following:

PROPOSITION 2 [17]. *For every even* $p \geq 30$ ($p \geq 26$), *there are nonisomorphic planar* (*nonplanar*) *cubic graphs of order* p *having the same path layer matrices.*

3-connected planar cubic graphs of order 30 (of order 26 for nonplanar graphs) can be constructed from the admissible graphs T on 15 and 13 vertices depicted in Fig. 5e,f. To seek minimal cubic graphs, we should examine only three families of graphs.

QUESTION 2 [17]. *Does there exist a pair of cubic graphs of order* $p = 20, 22, 24$ *having the same path layer matrices* ?

Simple regular graphs of degree 4 as well as cubic graphs include molecular graphs of many important classes of chemical compounds.

PROPOSITION 3 [17]. *For every* $p \geq 51$ ($p \geq 34$), *there are nonisomorphic planar* (*nonplanar*) *4-regular graphs of order* p *having the same path layer matrices.*

The corresponding admissible graphs T are shown in Fig. 5g,h.

QUESTION 3 [18]. *Does there exist a pair of 4-regular planar (nonplanar) graphs of order* $p \leq 50$ ($p \leq 33$), *having the same path layer matrices* ?

It seems that the minimal weighted graph having properties (a) and (b) is depicted in Fig. 4b. The molecular graph of benzene is the minimal example of an admissible multigraph

T (see Fig. 6). A pair of minimal multigraphs with the same path layer matrices is shown in Fig. 7. More complicated multigraphs are presented in Fig. 8. This implies the following

PROPOSITION 4. *For every* $p \geq 12$, *there are nonisomorphic planar weighted graphs (multigraphs) of order* p *having the same path layer matrices.*

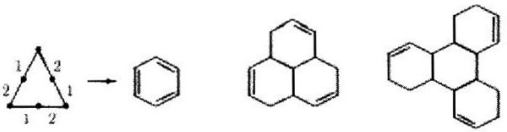

FIGURE 6. Examples of admissible molecular graphs.

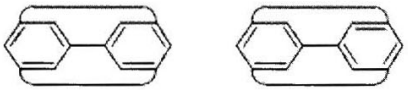

FIGURE 7. Minimal multigraphs with the same path layer matrix.

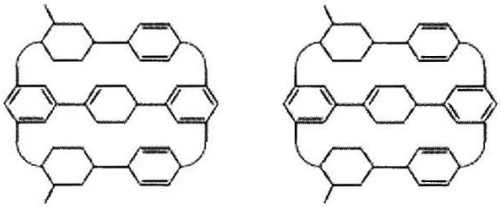

FIGURE 8. Example of molecular graphs.

It is interesting to discover a pair of weighted graphs of such kind with small number of vertices (with or without cut-vertices).

QUESTION 4. *Does there exist a pair of weighted graphs of order* $p \leq 11$ *having the same path layer matrices* ?

Acknowledgements.
The work of this author was supported by grants from the Russian Foundation of Fundamental Research (project codes 99-01-00581 and 00-07-90296) and INTAS (project code 97-1001).

REFERENCES AND NOTES

[1] M. Randić, Characterizations of atoms, molecules, and classes of molecules based on path enumerations. *Commun. Math. Chem.* (*MATCH*), **1979**, *7*, 5–64.

[2] G. S. Bloom, J. W. Kennedy, and L. V. Quintas, Some problems concerning distance and path degree sequences. *Lecture Notes Math.* **1983**, *1018*, 179–190.

[3] M. Randić and C. L. Wilkins, Graph theoretical approach to recognition of structural similarity in molecules. *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 23–31.

[4] M. Randić, G. A. Kraus, and B. Džonova-Jerman-Blažić, Ordering of graphs as an approach to structure-activity studies. In: *Chemical Applications of Topology and Graph Theory*. Eds.: R. B. King, Elsevier, New York, **1983**, 192–205.

[5] M. Randić, Design of molecules with desired properties. In: *Concepts and Application of Molecular Similarity*. Eds.: M. A. Johnson and G. M. Maggiora, John Wiley & Sons, New York, **1990**, 77–145.

[6] L. V. Quintas and P. J. Slater, Pairs of non-isomorphic graphs having the same path degree sequences. *Commun. Math. Chem.* (*MATCH*), **1981**, *12*, 75–86.

[7] F. Y. Halberstam and L. V. Quintas, A note on tables of distance and path degree sequences for cubic graphs. Presented at Silver Jubilee Conference on Combinatorics, University of Waterloo, Waterloo, Ontario, Canada, June 14 – July 2, 1982.

[8] F. Y. Halberstam and L. V. Quintas, Distance and path degree sequences for cubic graphs. Mathematical Department, Pace University, New York, NY, 1982.

[9] P. J. Slater, Counterexamples to Randić 's conjecture on distance degree sequences for trees. *J. Graph Theory*, **1982**, *6*, 89–91.

[10] F. Y. Halberstam and J. Żak, A note on planarity and distance degree sequences. *Lecture Notes Math.*, **1983**, *1018*, 286–289.

[11] V. A. Skorobogatov and A. A. Dobrynin, Metric analysis of graphs. *Commun. Math. Chem.* (*MATCH*), **1988**, *23*, 105–151.

[12] A. A. Dobrynin, Path enumerations and similarity of graphs. *Graph Theory Notes New York*, **1992**, *23*, 7–9.

[13] A. T. Balaban, G. S. Bloom, and L. V. Quintas, unpublished result, 1982.

[14] A. A. Dobrynin, Regular graphs having the same path layer matrix. *J. Graph Theory*, **1990**, *14*, 141–148.

[15] A. A. Dobrynin, Cubic graphs with 62 vertices having the same path layer matrix. *J. Graph Theory*, **1993**, *17*, 1–4.

[16] A. A. Dobrynin, Cubic graphs without cut-vertices having the same path layer matrix, submitted.

[17] L. S. Mel'nikov and A. A. Dobrynin, Constructing 3-connected graphs with the same path layer matrices. *Discretn. Analiz Issled. Oper. Ser. 2*, **2000**, *7*, 60–73 (in Russian).

[18] S. K. Kearsley, S. Sallamack, E. M. Fluder, J. D. Andose, R. T. Mosley, and R. P. Sheridan, Chemical similarity using physiochemical property descriptors. *J. Chem. Inf. Comput. Sci.*, **1996**, *36*, 118–127.

[19] L. I. Makarov, Separability of the weighted graph vertices and the algorithms of their classification, *Optoelectronics, Instrumentation and Data Processing*, **1997**, *5*, 79–85.

[20] L. I. Makarov, Methods and algorithms for predicting the properties of chemical compounds by common fragments of molecular graphs, *J. Struct. Chem.*, **1998**, *39*, 93–102.