Simple Construction of Embedding Frequencies of Trees and Rooted Trees

Vladimír Kvasnička* and Jiří Pospíchal

Department of Mathematics, Slovak Technical University, 812 37 Bratislava, Slovakia

Received August 20, 1994[⊗]

A unified approach for the calculation of embedding frequencies of trees as well as rooted trees is suggested. It consists of a simple pruning process when marginal vertices are successively removed. The problem of isomorphism between (rooted) trees is solved by making use of Read's linear canonical code. Simple criteria, based on the concept of valence vectors, for verification whether a (rooted) tree can be a subtree of a (rooted) tree are proved. Tables of embedding frequencies of trees and rooted trees through ten and eight vertices, respectively, are presented. The embedding frequencies of rooted trees are useful for construction of descriptors of molecular graphs when the so-called local properties are studied.

INTRODUCTION

Graph-theoretic *cluster expansion*[1-6] of physical and chemical properties of molecular systems represented by structural formulae belongs to very promising approaches on how to study structure–activity and/or structure–property relationships. This approach is usually based on the so-called *embedding frequencies*[1,2,7-10] which are determined, loosely speaking, as numbers of time a given submolecular cluster (subgraph) appears as a subgraph in the molecular graph (a formal representation of structural formula). Poshusta and McHughes[9] have suggested an algorithm for finding embedding frequencies of subtrees in a given larger tree. They published an extensive list of all embedding frequencies of trees through ten vertices. In the paper[10] they demonstrated efficiency of this approach for a prediction of thermochemical properties of alkanes. The values of these properties may be with high precision expressed as linear combinations of embedding frequencies assigned to few smaller subtrees. Magnetic susceptibility, which can be computed by "additive" graph-theoretical cluster expansion in terms of molecular fragments, can serve as another example of a special type of property.[11] Such properties may be formally considered as *global properties*, a given property is strongly dependent on the whole structure of molecular graphs. Another group of molecular properties is classified as *local properties*. Here, it is possible to assign the given property to an atom (vertex) of structural formula (molecular graph). The very instructive example of these properties are ^{13}C NMR chemical shifts of alkanes; they are unambiguously assigned (up to symmetry) to single carbon atoms of alkane skeletons. In order to correlate the local properties with the structure of molecular graphs we have to know information about the environment of the given vertex. The chosen atom plays a role of the so-called *root*,[12] a vertex distinguished from other vertices of the molecular graph. Therefore, if we would like to use the graph-theoretic cluster expansion for local molecular properties, then the embedding frequencies of rooted subgraphs in the given larger rooted graph should be known. This graph-theoretic problem corresponds to a special case of the original approach of embedding frequencies, the rooted graphs are more restricted forms of trees with one vertex distinguished from other ones.

The purpose of this paper is to outline a unified approach for the calculation of embedding frequencies of trees as well as rooted trees. We shall prove simple criteria, based on the so-called *valence vectors*, whether a given (rooted) tree can be a (rooted) subtree of a larger (rooted) tree. If one of those criteria is not satisfied, then the relationship "to be subtree" is falsified. A simple pruning process consisting of successively removing marginal vertices (with unit valences) is suggested. Its application allows us to construct in a simple way the embedding frequencies of (rooted) trees. An algorithm of the pruning process in a backtrack form is suggested.

Two extensive tables of embedding frequencies are presented. The first table lists embedding frequencies of smaller *trees* through five vertices in larger trees through ten vertices. The trees are identified by Read's linear codes in canonical form,[13] where the subcodes are ordered in a lexicographically nonincreasing sequence. This canonicity of codes ensures their uniqueness for forthcoming applications of embedding frequencies. Unfortunately, in Poshusta and McHughes' papers[9] the codes of trees they used are not presented in the canonical form (e.g., the tree no. 53 coded as 221010200 has the canonical code 220021010), therefore in some cases utilization of their results may involve serious difficulties.

The second table lists embedding frequencies of smaller *rooted trees* through five vertices in larger rooted trees through eight vertices; similarly, as in the first table, rooted trees are coded canonically.

The used code for trees is certainly not the only useful code; a nice example of code, which does not necessarily need separation for vertices with a degree higher then 10, is presented in ref 14. Plans for embedding search of rooted trees were suggested already in ref 15. The very idea of characterization of trees by embeddings even found its implementation in Gordon–Scantlebury topological index, which equals the number of embeddings of propan skeleton, and also a "centric topological index" was devised on the basis of the prunning process.[16]

THEORY

Let us consider a *tree*[12] (connected nonoriented graph without cycles) $T = (V,E)$, where $V = \{v_1,v_2,...,v_N\}$ is a nonempty (i.e., $N \geq 1$) *vertex set* and $E = \{e_1,e_2,...,e_M\}$ is

[⊗] Abstract published in *Advance ACS Abstracts*, November 15, 1994.

0095-2338/95/1635-0121$09.00/0

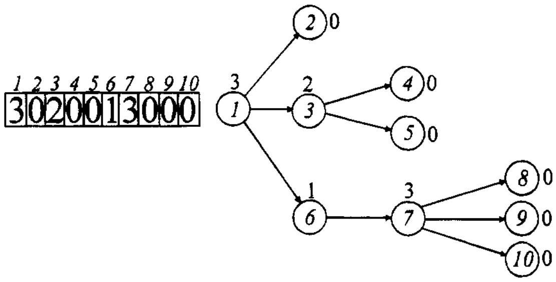

Figure 1. Illustrative example of the linear code of trees, with its entries labeled by integers (placed above them) corresponding to their sequential numbers. Reading the code from the left- to the right-hand side we may simply reconstruct the coded tree. Each entry determines the number of vertices attached from the right hand-side to the corresponding vertex. The vertices of trees are represented by open circles with vertex indices (sequential numbers of code) placed inside.

an *edge set*. The cardinalities of sets V and E are mutually related by

$$M = N - 1 \tag{1}$$

Let $v \in V$ be a vertex of the tree T: the *valence* or *degree* of this vertex, denoted by val(v), is the number of edges that are incident with the vertex v. The sum of valences of all vertices satisfies[12]

$$\sum_{v \in V} \mathrm{val}(v) = 2M = 2(N - 1) \tag{2}$$

A rooted tree[12] has one vertex, called the *root*, which is especially distinguished from the vertices of V. Formally, the rooted tree will be determined as an ordered triple

$$T(v) = (V,E,v) \tag{3}$$

where $v \in V$ is the root of the rooted tree $T(v)$. Two rooted trees $T(v) = (V,E,v)$ and $T'(v') = (V',E',v')$ are *isomorphic*[12] ($T(v) \approx T'(v')$) if and only if there exists a one-to-one mapping $\phi: V \rightarrow V'$ which saves the adjacency of vertices and maps the root of $T(v)$ onto the root of $T'(v')$, i.e., $\phi(v) = v'$.

Carbon skeletons of alkanes may be represented by trees, which are composed of vertices with valences bounded by $1 \le \mathrm{val}(v) \le 4$, for $\forall v \in V$. The vertices with unit valences are called *marginal vertices* or *endpoints*. In our forthcoming considerations we shall restrict ourselves to trees which correspond to alkanes, i.e., to trees with vertex valences equal at most to four. For (rooted) trees each subtree must be the so-called *induced subgraph*.[12] This means that a verification of $T' \subseteq T$ (T' is a subtree of T) is reduced to a finding of the subset $\tilde{V} \subseteq V(T)$ so that a subtree induced by $\tilde{V}$ is isomorphic to the tree T'. Moreover, if $T' \subseteq T$ and $|V(T')| = |V(T)|$, then trees T and T' are isomorphic.

An isomorphism between two (rooted) trees may be simply verified by making use of the linear code extensively studied by Read[13] and Knop et al.,[17,18] see Figure 1. This type of linear code is endowed by the very important property that two (rooted) trees are isomorphic if and only if their codes are identical.

We assign to each (rooted) tree T an ordered four-tuple (called the *valence vector*) composed of nonnegative integer entries

$$v_T = (v_1, v_2, v_3, v_4) \tag{4}$$

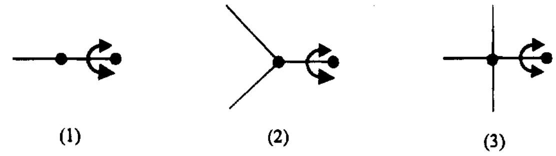

Figure 2. All three possible events that can appear when a marginal vertex is removed from a tree. The corresponding changes of valence vectors are specified by (7) and/or (8).

where v_i (for $i = 1,2,3,4$) is the number of vertices of T with valence equal to i. The entries of (4) satisfy the following two conditions (see eqs 1 and 2)

$$\sum_{i=1}^{4} v_i = N \tag{5a}$$

$$\sum_{i=1}^{4} i v_i = 2M = 2(N - 1) \tag{5b}$$

These two conditions together are not only necessary but also sufficient for the valence vector v_T to be *graphable*,[12] that is there exists a (rooted) tree T with the valence vector $v_T = (v_1, v_2, v_3, v_4)$.

Let us have two (rooted) trees T' and T that are related by $T' \subseteq T$; their valence vectors are denoted by $v_{T'} = (v_1', v_2', v_3', v_4')$ and $v_T = (v_1, v_2, v_3, v_4)$, respectively, then

$$\sum_{i=p}^{4} (v_i' - v_i) \le 0 \tag{6a}$$

$$\sum_{i=1}^{p} (v_i' - v_i) \le 0 \tag{6b}$$

for $p = 1,2,3,4$.

Applying these requirements we may simply verify whether a (rooted) tree T' can be a subtree of T. If there exists $1 \le p \le 4$ so that one of (or both) inequalities (16) are not satisfied, then T' could not be a subtree of T. Its proof can be done in a simple way as follows. Let us consider a marginal vertex (except the root for rooted trees) of T: removing this marginal vertex from T we arrive at a new (rooted) tree T' with valence vector determined by three alternative ways (see Figure 2)

$$\begin{aligned} &\text{case 1:} \quad v_{T'} = (v_1, v_2 - 1, v_3, v_4) \\ &\text{case 2:} \quad v_{T'} = (v_1 - 1, v_2 + 1, v_3 - 1, v_4) \\ &\text{case 3:} \quad v_{T'} = (v_1 - 1, v_2, v_3 + 1, v_4 - 1) \end{aligned} \tag{7}$$

The difference $\Delta v = v_{T'} - v_T$ is determined by

$$\begin{aligned} &\text{case 1:} \quad \Delta v = (0, -1, 0, 0) \\ &\text{case 2:} \quad \Delta v = (-1, 1, -1, 0) \\ &\text{case 3:} \quad \Delta v = (-1, 0, 1, -1) \end{aligned} \tag{8}$$

It is easy to verify that for all three cases the requirements (6a,b) are satisfied. But this was proved only for trees $T' \subset T$ such that T' is created from T removing a marginal vertex. For a general pair $T' \subset T$ these properties should be also satisfied, since the tree T' may be formed from T by a

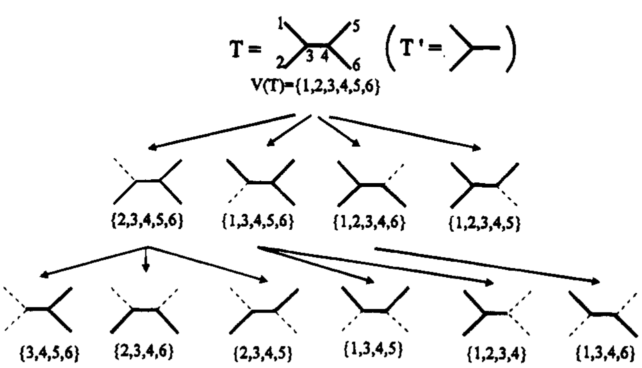

Figure 3. Outline of the pruning process for finding the embedding frequency n(T,T'), both trees are displayed at the top of the figure. First, we remove successively all marginal vertices of T, to obtain four subtrees that are displayed together with their vertex subsets in the second row of figure. Second, the pruning process is repeated for the subtrees obtained in the previous step. The algorithm produces only subtrees induced by different vertex subsets. Only two subtrees (the first and the last but one) satisfy conditions (6a,b) and are isomorphic to T', so that the embedding frequency of T' in T is n(T,T') = 2.

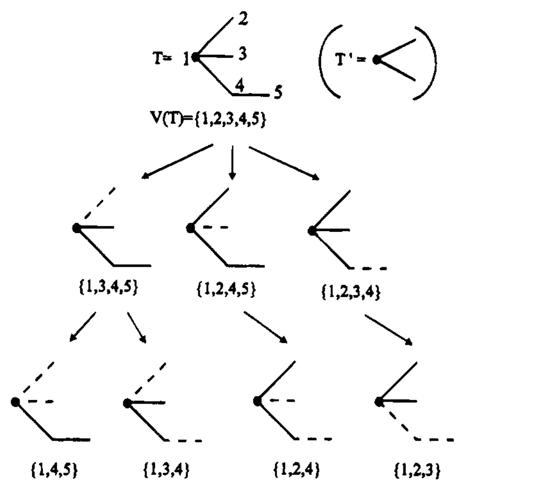

Figure 4. Outline of the pruning process for finding the embedding frequency n(T,T') of the rooted subtree T' in the rooted tree T. The approach is the same as in Figure 3, but the root is never removed in the process of pruning. From the third row of figure we see that the embedding frequency is n(T,T') = 3.

successive removing of marginal vertices. We get a sequence $T_0 = T' \subset T_1 \subset ... \subset T_n = T$ so that for each pair $T_{i-1} \subset T_i$ (for i = 1,2,...,n) the above conditions are satisfied, hence they should be also satisfied for $T' \subset T$, as was to be proved.

We turn now our attention to the construction of vertex subsets $\tilde{V} \subseteq V(T)$, which induce subtrees isomorphic to the prescribed (rooted) tree T'. The following *pruning process* forms all possibilities that will give the subset $\tilde{V}$. This vertex subset may be constructed by successive removal of a marginal vertex (except the root for rooted trees) from the current vertex subset, and a new vertex subset is formed. By further use of the described backtrack algorithm, it is ensured that this new subset has not been already formed in the previous steps. If the cardinality of the current subset $\tilde{V}$ is equal to the cardinality of the vertex set of the subtree T', then we construct the canonical code of a subtree $T(\tilde{V})$ induced by $\tilde{V}$. The equality of codes of T' and $T(\tilde{V})$ implies that these (rooted) trees are isomorphic. The pruning process can be substantially accelerated by applying requirements (6). If one of them is not satisfied, then the current branch of pruning is stopped due to the fact that the property $T' \subseteq T(\tilde{V})$ could not be fulfilled. Simple illustrative examples of the pruning processes for trees as well as rooted trees are displayed in Figures 3 and 4, respectively.

The *embedding frequency*[9] of a (rooted) subtree T' in a (rooted) tree T, denoted by n(T,T'), may now be defined as the number of appearances of vertex subsets $\tilde{V} \subseteq V(T)$ in the pruning process such that the (rooted) subtree induced by $\tilde{V}$ is isomorphic to the (rooted) subtree T'. The embedding frequencies of all trees up to ten vertices have been tabulated by Poshusta and McHughes.[9] The above outlined simple algorithm for constructing embedding frequencies based on the pruning process has been checked by Poshusta's results, with perfect agreement having been achieved. A pseudo-Pascal code of the pruning algorithm specified above in a form of the backtrack searching looks as follows.

Pruning algorithm

```
1   V0:=V(T); U1:={marginal vertices of T}; k:=1;
2   C0:={V0}; for i:=1 to |V(T)|-1 do Ci:=∅;
3   while k>0 do
4   if |Uk|>0 then
5   begin v:=an element of Uk;
6       Uk:=Uk\{v}; A:=Vk-1\{v};
7       begin Ck:=Ck∪{A}; Vk:=A; k:=k+1;
8           Uk:= Uk-1∪ {vertex originally con-
                        nected with v, if it is
                        now marginal}
9       end;
10  end else k:=k-1;
```

The first two lines initialize the algorithm. The integer variable k determines the so-called depth of searching. The set V_k is a current vertex subset and the set U_k is composed of marginal vertices of the subtree induced by V_{k-1} such that these marginal vertices have not been used yet in previous levels in the backtrack searching. The class C_k is composed of all possible different vertex subsets that have been constructed for the depth k. The system of inequalities (6a,b) may be applied in line 7, i.e., class C_k is enlarged only if the subgraph satisfies these inequalities. If one of them is not satisfied, then the current branch of the search tree is stopped, and the while cycle (line 3) is repeated for a new marginal vertex v.

Let us calculate the embedding frequency n(T,T'), where $T' \subseteq T$ (this property may be verified by conditions (6a,b)), $|V(T)| = N$, $|V(T')| = N'$, and $N' \leq N$. For $k = N - N'$ we construct all (rooted) trees that are induced by vertex subsets of C_k, then the embedding frequency $n(T,T')$ is equal to the number of all induced (rooted) trees that are isomorphic with T'.

In Table 1 are listed all embedding frequencies of trees through ten vertices and in Table 2 all embedding frequencies of rooted trees through eight vertices. The embedding frequencies of all trees through ten vertices are calculated for all possible eight subtrees through five vertices, see Figure 5. Similarly, the embedding frequencies of all rooted trees through eight vertices are calculated for all possible 17 rooted subtrees (see Figure 6) through five vertices; though we have omitted the rooted subtree composed of one vertex (i.e., root), its embedding frequency for each rooted tree automatically being equal to 1.

Table 1. Embedding Frequencies of All Trees through Ten Vertices

				embedding frequencies								
no.	N	no.[a]	code	1	2	3	4	5	6	7	8	alkane[b]
1	1	1	0	1	0	0	0	0	0	0	0	C1
2	2	1	10	2	1	0	0	0	0	0	0	C2
3	3	1	200	3	2	1	0	0	0	0	0	C3
4	4	1	2010	4	3	2	1	0	0	0	0	C4
5	4	2	3000	4	3	3	0	1	0	0	0	2M-C3
6	5	1	21010	5	4	3	2	0	1	0	0	C5
7	5	2	30010	5	4	4	2	1	0	1	0	2M-C4
8	5	3	40000	5	4	6	0	4	0	0	1	22MM-C3
9	6	1	210110	6	5	4	3	0	2	0	0	C6
10	6	2	210200	6	5	5	3	1	2	1	0	2M-C5
11	6	3	301010	6	5	5	4	1	1	2	0	3M-C5
12	6	4	300200	6	5	6	4	2	0	4	0	23MM-C4
13	6	5	400010	6	5	7	3	4	0	3	1	22MM-C4
14	7	1	2110110	7	6	5	4	0	3	0	0	C7
15	7	2	2110200	7	6	6	4	1	3	1	0	2M-C6
16	7	3	2200200	7	6	7	4	2	4	2	0	24MM-C5
17	7	4	3010110	7	6	6	5	1	3	2	0	3M-C6
18	7	5	3101010	7	6	6	6	1	3	3	0	3E-C5
19	7	6	3010200	7	6	7	6	2	2	5	0	23MM-C5
20	7	7	4000110	7	6	8	4	4	3	3	1	22MM-C5
21	7	8	4001010	7	6	8	6	4	1	6	1	33MM-C5
22	7	9	4000200	7	6	9	6	5	0	9	1	223MMM-C4
23	8	1	21101110	8	7	6	5	0	4	0	0	C8
24	8	2	21101200	8	7	7	5	1	4	1	0	2M-C7
25	8	3	22001200	8	7	8	5	2	4	2	0	25MM-C6
26	8	4	21102010	8	7	7	6	1	4	2	0	3M-C7
27	8	5	30110110	8	7	7	6	1	5	2	0	4M-C7
28	8	6	22002010	8	7	8	6	2	5	3	0	24MM-C6
29	8	7	31010110	8	7	7	7	1	5	3	0	3E-C6
30	8	8	30110200	8	7	8	7	2	4	5	0	23MM-C6
31	8	9	30102010	8	7	8	8	2	4	6	0	34MM-C6
32	8	10	31010200	8	7	8	8	2	5	6	0	32EM-C5
33	8	11	30200200	8	7	9	8	3	4	8	0	234MMM-C5
34	8	12	21103000	8	7	9	5	4	4	3	1	22MM-C6
35	8	13	22003000	8	7	10	5	5	6	4	1	224MMM-C5
36	8	14	40010110	8	7	9	7	4	4	6	1	33MM-C6
37	8	15	40101010	8	7	9	9	4	3	9	1	33EM-C5
38	8	16	30103000	8	7	10	8	5	3	10	1	223MMM-C5
39	8	17	40010200	8	7	10	9	5	2	12	1	233MMM-C5
40	8	18	40003000	8	7	12	9	8	0	18	2	2233MMMM-C4
41	9	1	211101110	9	8	7	6	0	5	0	0	C9
42	9	2	211101200	9	8	8	6	1	5	1	0	2M-C8
43	9	3	212001200	9	8	9	6	2	5	2	0	26MM-C7
44	9	4	211102010	9	8	8	7	1	5	2	0	3M-C8
45	9	5	212002010	9	8	9	7	2	5	3	0	25MM-C7
46	9	6	301101110	9	8	8	7	1	6	2	0	4M-C8
47	9	7	220102010	9	8	9	8	2	6	4	0	35MM-C7
48	9	8	301101200	9	8	9	7	2	7	3	0	24MM-C7
49	9	9	310101110	9	8	8	8	1	6	3	0	3E-C7
50	9	10	310110110	9	8	8	8	1	7	3	0	4E-C7
51	9	11	310101200	9	8	9	8	2	7	4	0	42EM-C6
52	9	12	302001110	9	8	9	8	2	5	5	0	23MM-C7
53	9	13	302001200	9	8	10	8	3	6	6	0	235MMM-C6
54	9	14	301102010	9	8	9	9	2	6	6	0	34MM-C7
55	9	15	310110200	9	8	9	9	2	7	6	0	32EM-C6
56	9	16	310102010	9	8	9	10	2	7	7	0	34EM-C6
57	9	17	302002010	9	8	10	10	3	6	9	0	234MMM-C6
58	9	18	310200200	9	8	10	10	3	8	9	0	243MME-C5
59	9	19	211103000	9	8	10	6	4	5	3	1	22MM-C7
60	9	20	212003000	9	8	11	6	5	5	4	1	225MMM-C6
61	9	21	220103000	9	8	11	7	5	7	5	1	224MMM-C6
62	9	22	230003000	9	8	13	6	8	9	6	2	2244MMMM-C5
63	9	23	400101110	9	8	10	8	4	5	6	1	33MM-C7
64	9	24	400110110	9	8	10	8	4	7	6	1	44MM-C7
65	9	25	400101200	9	8	11	8	5	7	7	1	244MMM-C6
66	9	26	401010110	9	8	10	10	4	6	9	1	33EM-C6
67	9	27	410101010	9	8	10	12	4	6	12	1	33EE-C5
68	9	28	301103000	9	8	11	9	5	5	10	1	223MMM-C6
69	9	29	310103000	9	8	11	10	5	7	11	1	223MME-C5
70	9	30	302003000	9	8	12	10	6	6	13	1	2234MMMM-C5
71	9	31	400110200	9	8	11	10	5	5	12	1	233MMM-C6
72	9	32	400102010	9	8	11	11	5	5	13	1	334MMM-C6
73	9	33	401010200	9	8	11	12	5	5	15	1	233MME-C5
74	9	34	400200200	9	8	12	12	6	4	18	1	2334MMMM-C5
75	9	35	400103000	9	8	13	12	8	3	21	2	2233MMMM-C5
76	10	1	2111011110	10	9	8	7	0	6	0	0	C10
77	10	2	2111011200	10	9	9	7	1	6	1	0	2M-C9
78	10	3	2120011200	10	9	10	7	2	6	2	0	27MM-C8
79	10	4	2111012010	10	9	9	8	1	6	2	0	3M-C9
80	10	5	2120012010	10	9	10	8	2	6	3	0	26MM-C8
81	10	6	2201012010	10	9	10	9	2	6	4	0	36MM-C8
82	10	7	2111020110	10	9	9	8	1	7	2	0	4M-C9
83	10	8	3011101110	10	9	9	8	1	7	2	0	5M-C9
84	10	9	2120020110	10	9	10	8	2	7	3	0	25MM-C8
85	10	10	3011101200	10	9	10	8	2	8	3	0	24MM-C8
86	10	11	2201020110	10	9	10	9	2	8	4	0	35MM-C8
87	10	12	3012001200	10	9	11	8	3	9	4	0	246MMM-C7
88	10	13	2111021010	10	9	9	9	1	7	3	0	3E-C8
89	10	14	2120021010	10	9	10	9	2	7	4	0	52EM-C7
90	10	15	3101101110	10	9	9	9	1	8	3	0	4E-C8
91	10	16	3110110110	10	9	9	9	1	9	3	0	4P-C7
92	10	17	2201021010	10	9	10	10	2	8	5	0	35EM-C7
93	10	18	3101101200	10	9	10	9	2	9	4	0	42EM-C7
94	10	19	2111020200	10	9	10	9	2	6	5	0	23MM-C8
95	10	20	2120020200	10	9	11	9	3	6	6	0	236MMM-C7
96	10	21	2201020200	10	9	11	10	3	7	7	0	235MMM-C7
97	10	22	3011102010	10	9	10	10	2	7	6	0	34MM-C8
98	10	23	3011020110	10	9	10	10	2	8	6	0	32EM-C7
99	10	24	3012002010	10	9	11	10	3	8	7	0	245MMM-C7
100	10	25	3102001110	10	9	10	10	2	8	6	0	45MM-C8
101	10	26	3110110200	10	9	10	10	2	9	6	0	4IP-C7
102	10	27	3102001200	10	9	11	10	3	9	7	0	253MME-C6
103	10	28	3011021010	10	9	10	11	2	9	7	0	34EM-C7
104	10	29	3101102010	10	9	10	11	2	9	7	0	43EM-C7
105	10	30	3101021010	10	9	10	12	2	10	8	0	34EE-C6
106	10	31	3011020200	10	9	11	11	3	8	9	0	234MMM-C7
107	10	32	3020102010	10	9	11	12	3	8	10	0	345MMM-C7
108	10	33	3020021010	10	9	11	12	3	9	10	0	234MME-C6
109	10	34	3020020200	10	9	12	12	4	8	12	0	2345MMMM-C6
110	10	35	3110200200	10	9	11	11	3	10	9	0	32IPM-C6
111	10	36	3102002010	10	9	11	12	3	10	10	0	243MME-C6
112	10	37	3200200200	10	9	12	12	4	12	12	0	243MMIP-C5
113	10	38	2111013000	10	9	11	7	4	6	3	1	22MM-C8
114	10	39	2120013000	10	9	12	7	5	6	4	1	226MMM-C7
115	10	40	2201013000	10	9	12	8	5	6	5	1	225MMM-C7
116	10	41	2300013000	10	9	14	7	8	6	6	2	2255MMMM-C6
117	10	42	2300020110	10	9	12	8	5	9	5	1	224MMM-C7
118	10	43	2300021010	10	9	12	9	5	9	6	1	224MME-C6
119	10	44	2300020200	10	9	13	9	6	8	8	1	2245MMMM-C6
120	10	45	2111030010	10	9	11	9	4	6	6	1	33MM-C8
121	10	46	2120030010	10	9	12	9	5	6	7	1	255MMM-C7
122	10	47	4001101110	10	9	11	9	4	8	6	1	44MM-C8
123	10	48	2201030010	10	9	12	10	5	8	8	1	335MMM-C7
124	10	49	4001101200	10	9	12	9	5	10	7	1	244MMM-C7
125	10	50	2300030010	10	9	14	9	8	10	9	2	2244MMMM-C6
126	10	51	4010101110	10	9	11	11	4	7	9	1	33EM-C7
127	10	52	4010110110	10	9	11	11	4	9	9	1	44EM-C7
128	10	53	4010101200	10	9	12	11	5	9	10	1	244MME-C6
129	10	54	4101010110	10	9	11	13	4	9	12	1	33EE-C6
130	10	55	3011103000	10	9	12	10	5	6	10	1	223MMM-C7
131	10	56	3012003000	10	9	13	10	6	7	11	1	2235MMMM-C6
132	10	57	3101103000	10	9	12	11	5	9	11	1	223MME-C6
133	10	58	3020103000	10	9	13	12	6	8	14	1	2234MMMM-C6
134	10	59	3102003000	10	9	13	12	6	11	14	1	3224EMMM-C5
135	10	60	3030003000	10	9	15	12	9	9	18	2	22344MMMMM-C5
136	10	61	4002001110	10	9	12	11	5	6	12	1	233MMM-C7
137	10	62	4002001200	10	9	13	11	6	8	13	1	2335MMMM-C6
138	10	63	3011030010	10	9	12	12	5	7	13	1	334MMM-C7
139	10	64	4001102010	10	9	12	12	5	8	13	1	344MMM-C7
140	10	65	3101030010	10	9	12	13	5	9	14	1	334MME-C6
141	10	66	3020030010	10	9	13	13	6	8	16	1	2344MMMM-C6
142	10	67	4010110200	10	9	12	13	5	8	15	1	233MME-C6
143	10	68	4010102010	10	9	12	14	5	8	16	1	343MME-C6
144	10	69	4101010200	10	9	12	15	5	9	18	1	332EEM-C5
145	10	70	4002002010	10	9	13	14	6	7	19	1	2334MMMM-C6
146	10	71	4010200200	10	9	13	15	6	8	21	1	3234EMMM-C5
147	10	72	4001103000	10	9	14	13	8	6	21	2	2233MMMM-C6
148	10	73	4001030010	10	9	14	15	8	6	24	2	3344MMMM-C6
149	10	74	4010103000	10	9	14	15	8	7	24	2	3223EMMM-C5
150	10	75	4002003000	10	9	15	15	9	6	27	2	22334MMMMM-C5

[a] Serial number of trees with fixed number of vertices (see second column). [b] Names of alkanes are constructed after refs 14 and 29.

Table 2. Embedding Frequencies of All Rooted Trees through Eight Vertices

				embedding frequencies															
no.	N	no.[a]	code	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	2	1	10	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	3	1	110	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	3	2	200	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
5	4	1	1110	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
6	4	2	1200	1	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0
7	4	3	2010	2	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0
8	4	4	3000	3	0	3	0	0	0	1	0	0	0	0	0	0	0	0	0
9	5	1	11110	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0
10	5	2	11200	1	1	0	2	0	0	0	0	1	0	0	0	0	0	0	0
11	5	3	12010	1	2	0	1	1	0	0	0	0	1	0	0	0	0	0	0
12	5	4	13000	1	3	0	0	3	0	0	0	0	0	1	0	0	0	0	0
13	5	5	20110	2	1	1	1	0	1	0	0	0	0	0	1	0	0	0	0
14	5	6	20200	2	2	1	0	1	2	0	0	0	0	0	0	1	0	0	0
15	5	7	21010	2	2	1	0	0	2	0	0	0	0	0	0	0	1	0	0
16	5	8	30010	3	1	3	0	0	2	1	0	0	0	0	0	0	0	1	0
17	5	9	40000	4	0	6	0	0	0	4	0	0	0	0	0	0	0	0	1
18	6	1	111110	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0
19	6	2	111200	1	1	0	1	0	0	0	2	0	0	0	0	0	0	0	0
20	6	3	112010	1	1	0	2	0	0	0	1	1	0	0	0	0	0	0	0
21	6	4	113000	1	1	0	3	0	0	0	0	3	0	0	0	0	0	0	0
22	6	5	120110	1	2	0	1	1	0	0	1	0	1	0	0	0	0	0	0
23	6	6	120200	1	2	0	2	1	0	0	0	1	2	0	0	0	0	0	0
24	6	7	121010	1	2	0	2	1	0	0	0	0	2	0	0	0	0	0	0
25	6	8	130010	1	3	0	1	3	0	0	0	0	2	1	0	0	0	0	0
26	6	9	201110	2	1	1	1	0	1	0	1	0	0	0	1	0	0	0	0
27	6	10	201200	2	1	1	2	0	1	0	0	1	0	0	2	0	0	0	0
28	6	11	202010	2	2	1	1	1	2	0	0	0	1	0	1	1	0	0	0
29	6	12	203000	2	3	1	0	3	3	0	0	0	0	1	0	3	0	0	0
30	6	13	210110	2	2	1	1	0	2	0	0	0	0	0	1	0	1	0	0
31	6	14	210200	2	3	1	0	1	3	0	0	0	0	0	0	1	2	0	0
32	6	15	300110	3	1	3	1	0	2	1	0	0	0	0	2	0	0	1	0
33	6	16	300200	3	2	3	0	1	4	1	0	0	0	0	0	2	0	2	0
34	6	17	301010	3	2	3	0	0	4	1	0	0	0	0	0	0	1	2	0
35	6	18	400010	4	1	6	0	0	3	4	0	0	0	0	0	0	0	3	1
36	7	1	1111110	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0
37	7	2	1111200	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0
38	7	3	1112010	1	1	0	1	0	0	0	2	0	0	0	0	0	0	0	0
39	7	4	1113000	1	1	0	1	0	0	0	3	0	0	0	0	0	0	0	0
40	7	5	1120110	1	1	0	2	0	0	0	1	1	0	0	0	0	0	0	0
41	7	6	1120200	1	1	0	2	0	0	0	2	1	0	0	0	0	0	0	0
42	7	7	1121010	1	1	0	2	0	0	0	2	1	0	0	0	0	0	0	0
43	7	8	1130010	1	1	0	3	0	0	0	1	3	0	0	0	0	0	0	0
44	7	9	1201110	1	2	0	1	1	0	0	1	0	1	0	0	0	0	0	0
45	7	10	1201200	1	2	0	1	1	0	0	2	0	1	0	0	0	0	0	0
46	7	11	1202010	1	2	0	2	1	0	0	1	1	2	0	0	0	0	0	0
47	7	12	1203000	1	2	0	3	1	0	0	0	3	3	0	0	0	0	0	0
48	7	13	1210110	1	2	0	2	1	0	0	1	0	2	0	0	0	0	0	0
49	7	14	1210200	1	2	0	3	1	0	0	0	1	3	0	0	0	0	0	0
50	7	15	1300110	1	3	0	1	3	0	0	1	0	2	1	0	0	0	0	0
51	7	16	1300200	1	3	0	2	3	0	0	0	1	4	1	0	0	0	0	0
52	7	17	1301010	1	3	0	2	3	0	0	0	0	4	1	0	0	0	0	0
53	7	18	2011110	2	1	1	1	0	1	0	1	0	0	0	1	0	0	0	0
54	7	19	2011200	2	1	1	1	0	1	0	2	0	0	0	1	0	0	0	0
55	7	20	2012010	2	1	1	2	0	1	0	1	1	0	0	2	0	0	0	0
56	7	21	2013000	2	1	1	3	0	1	0	0	3	0	0	3	0	0	0	0
57	7	22	2020110	2	2	1	1	1	2	0	1	0	1	0	1	1	0	0	0
58	7	23	2020200	2	2	1	2	1	2	0	0	1	2	0	2	1	0	0	0
59	7	24	2021010	2	2	1	2	1	2	0	0	0	2	0	2	1	0	0	0
60	7	25	2030010	2	3	1	1	3	3	0	0	0	2	1	1	3	0	0	0
61	7	26	2101110	2	2	1	1	0	2	0	1	0	0	0	1	0	1	0	0
62	7	27	2101200	2	2	1	2	0	2	0	0	1	0	0	2	0	1	0	0
63	7	28	2102010	2	3	1	1	1	3	0	0	0	1	0	1	1	2	0	0
64	7	29	2103000	2	4	1	0	3	4	0	0	0	0	1	0	3	3	0	0
65	7	30	2110110	2	2	1	2	0	2	0	0	0	0	0	2	0	1	0	0
66	7	31	2110200	2	3	1	1	1	3	0	0	0	0	0	1	1	2	0	0
67	7	32	2200200	2	4	1	0	2	4	0	0	0	0	0	0	2	4	0	0
68	7	33	3001110	3	1	3	1	0	2	1	1	0	0	0	2	0	0	1	0
69	7	34	3001200	3	1	3	2	0	2	1	0	1	0	0	4	0	0	1	0
70	7	35	3002010	3	2	3	1	1	4	1	0	0	1	0	2	2	0	2	0
71	7	36	3003000	3	3	3	0	3	6	1	0	0	0	1	0	6	0	3	0
72	7	37	3010110	3	2	3	1	0	4	1	0	0	0	0	2	0	1	2	0
73	7	38	3010200	3	3	3	0	1	6	1	0	0	0	0	0	2	2	3	0

Table 2 (Continued)

				embedding frequencies															
no.	*N*	no.[a]	code	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
74	7	39	3101010	3	3	3	0	0	6	1	0	0	0	0	0	0	3	3	0
75	7	40	4000110	4	1	6	1	0	3	4	0	0	0	0	3	0	0	3	1
76	7	41	4000200	4	2	6	0	1	6	4	0	0	0	0	0	3	0	6	1
77	7	42	4001010	4	2	6	0	0	6	4	0	0	0	0	0	0	1	6	1
78	8	1	11111110	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0
79	8	2	11111200	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0
80	8	3	11112010	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0
81	8	4	11113000	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0
82	8	5	11120110	1	1	0	1	0	0	0	2	0	0	0	0	0	0	0	0
83	8	6	11120200	1	1	0	1	0	0	0	2	0	0	0	0	0	0	0	0
84	8	7	11121010	1	1	0	1	0	0	0	2	0	0	0	0	0	0	0	0
85	8	8	11130010	1	1	0	1	0	0	0	3	0	0	0	0	0	0	0	0
86	8	9	11201110	1	1	0	2	0	0	0	1	1	0	0	0	0	0	0	0
87	8	10	11201200	1	1	0	2	0	0	0	1	1	0	0	0	0	0	0	0
88	8	11	11202010	1	1	0	2	0	0	0	2	1	0	0	0	0	0	0	0
89	8	12	11203000	1	1	0	2	0	0	0	3	1	0	0	0	0	0	0	0
90	8	13	11210110	1	1	0	2	0	0	0	2	1	0	0	0	0	0	0	0
91	8	14	11210200	1	1	0	2	0	0	0	3	1	0	0	0	0	0	0	0
92	8	15	11300110	1	1	0	3	0	0	0	1	3	0	0	0	0	0	0	0
93	8	16	11300200	1	1	0	3	0	0	0	2	3	0	0	0	0	0	0	0
94	8	17	11301010	1	1	0	3	0	0	0	2	3	0	0	0	0	0	0	0
95	8	18	12011110	1	2	0	1	1	0	0	1	0	1	0	0	0	0	0	0
96	8	19	12011200	1	2	0	1	1	0	0	1	0	1	0	0	0	0	0	0
97	8	20	12012010	1	2	0	1	1	0	0	2	0	1	0	0	0	0	0	0
98	8	21	12013000	1	2	0	1	1	0	0	3	0	1	0	0	0	0	0	0
99	8	22	12020110	1	2	0	2	1	0	0	1	1	2	0	0	0	0	0	0
100	8	23	12020200	1	2	0	2	1	0	0	2	1	2	0	0	0	0	0	0
101	8	24	12021010	1	2	0	2	1	0	0	2	1	2	0	0	0	0	0	0
102	8	25	12030010	1	2	0	3	1	0	0	1	3	3	0	0	0	0	0	0
103	8	26	12101110	1	2	0	2	1	0	0	1	0	2	0	0	0	0	0	0
104	8	27	12101200	1	2	0	2	1	0	0	2	0	2	0	0	0	0	0	0
105	8	28	12102010	1	2	0	3	1	0	0	1	1	3	0	0	0	0	0	0
106	8	29	12103000	1	2	0	4	1	0	0	0	3	4	0	0	0	0	0	0
107	8	30	12110110	1	2	0	2	1	0	0	2	0	2	0	0	0	0	0	0
108	8	31	12110200	1	2	0	3	1	0	0	1	1	3	0	0	0	0	0	0
109	8	32	12200200	1	2	0	4	1	0	0	0	2	4	0	0	0	0	0	0
110	8	33	13001110	1	3	0	1	3	0	0	1	0	2	1	0	0	0	0	0
111	8	34	13001200	1	3	0	1	3	0	0	2	0	2	1	0	0	0	0	0
112	8	35	13002010	1	3	0	2	3	0	0	1	1	4	1	0	0	0	0	0
113	8	36	13003000	1	3	0	3	3	0	0	0	3	6	1	0	0	0	0	0
114	8	37	13010110	1	3	0	2	3	0	0	1	0	4	1	0	0	0	0	0
115	8	38	13010200	1	3	0	3	3	0	0	0	1	6	1	0	0	0	0	0
116	8	39	13101010	1	3	0	3	3	0	0	0	0	6	1	0	0	0	0	0
117	8	40	20111110	2	1	1	1	0	1	0	1	0	0	0	1	0	0	0	0
118	8	41	20111200	2	1	1	1	0	1	0	1	0	0	0	1	0	0	0	0
119	8	42	20112010	2	1	1	1	0	1	0	2	0	0	0	1	0	0	0	0
120	8	43	20113000	2	1	1	1	0	1	0	3	0	0	0	1	0	0	0	0
121	8	44	20120110	2	1	1	2	0	1	0	1	1	0	0	2	0	0	0	0
122	8	45	20120200	2	1	1	2	0	1	0	2	1	0	0	2	0	0	0	0
123	8	46	20121010	2	1	1	2	0	1	0	2	1	0	0	2	0	0	0	0
124	8	47	20130010	2	1	1	3	0	1	0	1	3	0	0	3	0	0	0	0
125	8	48	20201110	2	2	1	1	1	2	0	1	0	1	0	1	1	0	0	0
126	8	49	20201200	2	2	1	1	1	2	0	2	0	1	0	1	1	0	0	0
127	8	50	20202010	2	2	1	2	1	2	0	1	1	2	0	2	1	0	0	0
128	8	51	20203000	2	2	1	3	1	2	0	0	3	3	0	3	1	0	0	0
129	8	52	20210110	2	2	1	2	1	2	0	1	0	2	0	2	1	0	0	0
130	8	53	20210200	2	2	1	3	1	2	0	0	1	3	0	3	1	0	0	0
131	8	54	20300110	2	3	1	1	3	3	0	1	0	2	1	1	3	0	0	0
132	8	55	20300200	2	3	1	2	3	3	0	0	1	4	1	2	3	0	0	0
133	8	56	20301010	2	3	1	2	3	3	0	0	0	4	1	2	3	0	0	0
134	8	57	21011110	2	2	1	1	0	2	0	1	0	0	0	1	0	1	0	0
135	8	58	21011200	2	2	1	1	0	2	0	2	0	0	0	1	0	1	0	0
136	8	59	21012010	2	2	1	2	0	2	0	1	1	0	0	2	0	1	0	0
137	8	60	21013000	2	2	1	3	0	2	0	0	3	0	0	3	0	1	0	0
138	8	61	21020110	2	3	1	1	1	3	0	1	0	1	0	1	1	2	0	0
139	8	62	21020200	2	3	1	2	1	3	0	0	1	2	0	2	1	2	0	0
140	8	63	21021010	2	3	1	2	1	3	0	0	0	2	0	2	1	2	0	0
141	8	64	21030010	2	4	1	1	3	4	0	0	0	2	1	1	3	3	0	0
142	8	65	21101110	2	2	1	2	0	2	0	1	0	0	0	2	0	1	0	0
143	8	66	21101200	2	2	1	3	0	2	0	0	1	0	0	3	0	1	0	0
144	8	67	21102010	2	3	1	2	1	3	0	0	0	1	0	2	1	2	0	0
145	8	68	21103000	2	4	1	1	3	4	0	0	0	0	1	1	3	3	0	0
146	8	69	22001110	2	3	1	1	1	3	0	1	0	0	0	1	1	2	0	0

Table 2 (Continued)

				embedding frequencies															
no.	*N*	no.[a]	code	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
147	8	70	22001200	2	3	1	2	1	3	0	0	1	0	0	2	1	2	0	0
148	8	71	22002010	2	4	1	1	2	4	0	0	0	1	0	1	2	4	0	0
149	8	72	22003000	2	5	1	0	4	5	0	0	0	0	1	0	4	6	0	0
150	8	73	30011110	3	1	3	1	0	2	1	1	0	0	0	2	0	0	1	0
151	8	74	30011200	3	1	3	1	0	2	1	2	0	0	0	2	0	0	1	0
152	8	75	30012010	3	1	3	2	0	2	1	1	1	0	0	4	0	0	1	0
153	8	76	30013000	3	1	3	3	0	2	1	0	3	0	0	6	0	0	1	0
154	8	77	30020110	3	2	3	1	1	4	1	1	0	1	0	2	2	0	2	0
155	8	78	30020200	3	2	3	2	1	4	1	0	1	2	0	4	2	0	2	0
156	8	79	30021010	3	2	3	2	1	4	1	0	0	2	0	4	2	0	2	0
157	8	80	30030010	3	3	3	1	3	6	1	0	0	2	1	2	6	0	3	0
158	8	81	30101110	3	2	3	1	0	4	1	1	0	0	0	2	0	1	2	0
159	8	82	30101200	3	2	3	2	0	4	1	0	1	0	0	4	0	1	2	0
160	8	83	30102010	3	3	3	1	1	6	1	0	0	1	0	2	2	2	3	0
161	8	84	30103000	3	4	3	0	3	8	1	0	0	0	1	0	6	3	4	0
162	8	85	30110110	3	2	3	2	0	4	1	0	0	0	0	4	0	1	2	0
163	8	86	30110200	3	3	3	1	1	6	1	0	0	0	0	2	2	2	3	0
164	8	87	30200200	3	4	3	0	2	8	1	0	0	0	0	0	4	4	4	0
165	8	88	31010110	3	3	3	1	0	6	1	0	0	0	0	2	0	3	3	0
166	8	89	31010200	3	4	3	0	1	8	1	0	0	0	0	0	2	5	4	0
167	8	90	40001110	4	1	6	1	0	3	4	1	0	0	0	3	0	0	3	1
168	8	91	40001200	4	1	6	2	0	3	4	0	1	0	0	6	0	0	3	1
169	8	92	40002010	4	2	6	1	1	6	4	0	0	1	0	3	3	0	6	1
170	8	93	40003000	4	3	6	0	3	9	4	0	0	0	1	0	9	0	9	1
171	8	94	40010110	4	2	6	1	0	6	4	0	0	0	0	3	0	1	6	1
172	8	95	40010200	4	3	6	0	1	9	4	0	0	0	0	0	3	2	9	1
173	8	96	40101010	4	3	6	0	0	9	4	0	0	0	0	0	0	3	9	1

[a] Serial number of rooted trees with fixed number of vertices (see second column).

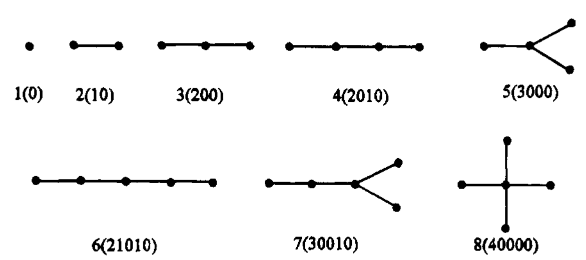

Figure 5. All possible eight subtrees through five vertices. In Table 1 are given embedding frequencies that correspond to these subtrees.

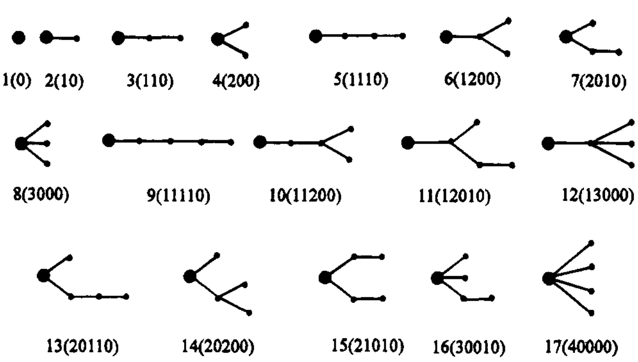

Figure 6. All possible rooted subtrees through five vertices. In Table 2 are given only embedding frequencies that correspond to rooted subtrees labeled by 2−17, whereas the first rooted subtree composed of one vertex (root) is not considered, its embedding frequency being automatically equal to one for any rooted tree.

DISCUSSION

The embedding frequencies are very useful graph-theoretical descriptors of molecular graphs for construction of structure−activity and/or structure−property relationships. They reflect in a natural way, closely related to a thinking of chemists, structural features of molecular graphs. Their calculation can be simply done for acyclic molecular graphs mainly due to (i) the possibility to detect isomorphism between trees by Read's linear codes and (ii) the fact that for trees each subtree is automatically induced by a vertex subset. A simple pruning procedure for construction of embedding frequencies of trees as well as rooted trees is described in section 2. It can be considerably accelerated by making use of conditions (6a,b) detecting on the basis of valence vectors whether a smaller tree can be a subtree of a larger tree. If one of the conditions is not satisfied, then the branch of search algorithm is stopped and the method returns to the previous branching point of the search tree. The calculated embedding frequencies of trees and rooted trees are listed in Tables 1 and 2, respectively.

The present method of calculation of embedding frequencies may be simply generalized for general acyclic molecular graphs (i.e., which contain also double and triple bonds and heteroatoms).[19] For instance, if we would like to calculate embedding frequencies of alkynes, then the edges of (rooted) trees should also be specified by the so-called bond multiplicities. Read's linear codes are simply generalized for trees with multiple edges, so that each entry (except of the first entry) should be enlarged by an additional entry (integer), which specifies the multiplicity of the edge; for details see ref 19.

The embedding-frequency approach has been recently successfully used in our papers dedicated to neural network applications to predict some thermochemical properties of alkanes[20,21] and also to classify inductive and resonance effects of functional groups.[22] Recently, similar applications are under way for a broad classes of acyclic molecules with heteroatoms.[23] Our results indicate that neural networks based on the embedding-frequency descriptors provide relationships that are comparable (or even better) than other results obtained by making use of other topological indices as descriptors.[24-26] Neural network predictions of ^{13}C-NMR chemical shifts done by one (V.K.) of the present authors[27,28] were based on descriptors that are closely related to the method of embedding frequencies of rooted trees. Currently, we are repeating these calculations in the present framework, and the preliminary results are very promising.

ACKNOWLEDGMENT

This research was supported in part by '94 Chemical Structure Association Trust Award.

REFERENCES AND NOTES

(1) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.
(2) Gordon, M.; Kennedy, J. W. The Graph-like State of Matter. *J. Chem. Soc. Faraday II* **1973**, *69*, 484–504.
(3) Klein, D. J. Chemical Graph-Theoretic Cluster Expansions *Int. J. Quant. Chem.: Quant. Chem. Symp.* **1986**, *20*, 153–171.
(4) Poshusta, R. D.; Schmalz, T. G.; Klein, D. J. Heisenberg-model Cluster Expansion for Half-filled Hubbard and PPP Models. *Mol. Phys.* **1989**, *66*, 317–332.
(5) Trinajstic, N.; Klein, D. J.; Randic, M. On Some Solved and Unsolved Problems of Chemical Graph Theory. *Int. J. Quant. Chem.: Quant. Chem. Symp.* **1986**, *20*, 699–742.
(6) Rouvray, D. H. Isomer Enumeration Methods. *Chem. Soc. Rev.* **1974**, *3*, 355–372.
(7) Ruskey, F. Listing and Counting Subtrees of a Tree. *SIAM J. Comput.* **1981**, *10*, 141–150.
(8) Alon, N. On the Number of Certain Subgraphs Contained in Graphs with a Given Number of Edges. *Israel J. Math.* **1986**, *53*, 97–120.
(9) Poshusta, R. D.; McHughes, M. C. Embedding Frequencies of Trees. *J. Math. Chem.* **1989**, *3*, 193–215.
(10) McHughes, M. C.; Poshusta, R. D. Graph-theoretic Cluster Expansions. Thermochemical Properties for Alkanes. *J. Math. Chem.* **1990**, *4*, 227–249.
(11) 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.
(12) Harary, F. *Graph Theory*; Addison Wesley: Reading, MA, 1969.
(13) Read, R. C. The Coding of Various Kinds of Unlabeled Trees. In *Graph Theory and Computing*; Read, R. C., Ed.; Academic Press: New York, 1972; pp 153–182.
(14) Balaban, A. T. Topological Indices and Their Uses: A New Approach for the Coding of Alkanes. *J. Mol. Struct. (THEOCHEM)* **1988**, *165*, 243–253.
(15) Poshusta, R. D. Codes for Rooted Trees. In *MATH/CHEM/COMP 1988*; Graovac, A., Ed.; *Studies in Physical and Theoretical Chemistry*; Elsevier: Amsterdam, 1989; Vol. 63, pp 9–24.
(16) Balaban, A. T. Chemical Graphs XXXIV. Five New Topological Indexes for the Branching of the Tree-like Graphs. *Theor. Chim. Acta* **1979**, *53*, 355–375.
(17) Knop, J. V.; Muller, W. R.; Szymanski, K.; Trinajstic, N. *Computer Generation of Certain Classes of Molecules*; SKTH: Zagreb, Croatia, 1985.
(18) Knop, J. V.; Muller, W. R.; Jericevic, Z.; Trinajstic, N. Computer Enumeration and Generation of Trees and Rooted Trees. *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 91–99.
(19) Kvasnička, V.; Pospíchal, J. Constructive Enumeration of Acyclic Molecules. *Collect. Czech. Chem. Comm.* **1991**, *56*, 1777–1802.
(20) Cherqaoui, D.; Villemin, D.; Kvasnička, V. Application of Neural Network Approach for Prediction of some Thermochemical Properties of Alkanes. *Chemometrics and Intell. Lab. Syst.* **1994**, *24*, 117–128.
(21) Kvasnička, V.; Cherqaoui, D.; Villemin, D. Application of High-order Neural Networks for Prediction of Thermochemical Properties of Alkanes. *Collect. Czech. Chem. Comm.*, submitted.
(22) Kvasnička, V.; Sklenák, Š.; Pospíchal, J. Neural Network Classification of Inductive and Resonance Effects of Substituents. *J. Am. Chem. Soc.* **1993**, *115*, 1495–1500.
(23) 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.
(24) Mihalic, Z.; Trinajstic, N. A Graph-Theoretical Approach to Structure-Property Relationships. *J. Chem. Ed.* **1992**, *69*, 701–712.
(25) Randic, M. On Characterization of Molecular Branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609–6614.
(26) Kvasnička, V. An Application of Neural Networks in Chemistry. *Chem. Papers* **1990**, *44*, 775–792.
(27) Kvasnička, V. An Application of Neural Networks in Chemistry. Prediction of ^{13}C NMR Chemical Shifts. *J. Math. Chem.* **1991**, *6*, 63–76.
(28) Kvasnička, V.; Pospíchal, J. *Neural Network Applications in Chemistry*, Lecture Notes in Chemistry Springer Verlag: Berlin, manuscript in preparation.
(29) Balaban, A. T.; Ciubotariu, D.; Medeleanu, M. Topological Indices and Real Number Vertex Invariants Based on Graph Eigenvalues or Eigenvectors. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 517–523.

CI940089Q