

CCA-1682

YU ISSN 0011 1643

UDC 541

Original Scientific Paper

**Molecular ID Numbers<sup>†\*</sup>**

*Klaus Szymanski, Wolfgang R. Müller, Jan V. Knop,  
and Nenad Trinajstić\*\**

*Computing Centre, The University of Düsseldorf, 4000 Düsseldorf,  
Federal Republic of Germany*

and

*\*\*The Rugjer Bošković Institute, P.O.B. 1016, 41001 Zagreb, Croatia, Yugoslavia*

Received September 2, 1985

A search for counterexamples of Randić's prime molecular ID-numbers for alkane-trees with up to 20 carbon atoms produced only one pair of non-isomorphic trees with the same prime ID-number. Thus, it is shown that the prime ID-number, although a highly discriminating molecular descriptor, is not unique.

Randić has recently proposed a novel structural descriptor, called the molecular identification (ID) number, as a potentially useful label for identifying molecular graphs (chemical structures) by a real number.<sup>1</sup> He defined the ID-number as follows.

The mapping from the edge-set  $E$  of the graph to the real numbers  $R$

$$g: E \rightarrow R \quad (1)$$

is defined by

$$g(e) = g(u, v) = (\deg u \cdot \deg v)^{-1/2} \quad (2)$$

where  $e$  stands for the edge connecting the vertex  $u$  to the vertex  $v$

$$e = (u, v) \in E \quad (3)$$

and  $\deg u$  denotes the degree of the vertex  $u$  in the graph (structure)  $G$ , i.e. the number of edges incident with  $u$ .<sup>2</sup> Therefore, the  $g(e)$ 's represent the connectivity weights. Let  $p = e_1 e_2 \dots e_m$  ( $m > 0$ ) be a path in  $G$ , then the mapping  $g$  may be extended to the set of all paths in  $G$  by

$$g^* = \prod_{i=1}^m g(e_i) \quad (4)$$

The ID-number of  $G$  is then finally defined as

$$\text{ID}(G) = N + \sum_p g^*(p) \quad (5)$$

<sup>†</sup> Reported in part at the IUPAC International Symposium on the Applications of Mathematical Concepts to Chemistry (Dubrovnik, September 2—5, 1985).

\* Dedicated to the memory of the late Professor Andrej Ažman.

where  $N$  is the number of vertices in  $G$ . The summation in (5) is taken over all different paths in  $G$ .

The ID-number appears to be an attractive topological (or more correctly, graph theoretical) descriptor which is relatively easy to derive from many classes of structures<sup>3</sup> and which has structural significance. It is of interest to find out if the ID-numbers are unique. Randić invited search for graphs that have non-unique ID.<sup>1</sup>

As soon as Randić's work became known to us we started a search for examples of trees that have identical ID-numbers.<sup>4</sup> Since we are in a position to generate all molecular trees (and subsets of trees) with given numbers of vertices,<sup>5-8</sup> it was a natural decision to look for such examples in the field of trees. An added incentive was that the special structure of trees allows a very efficient method for computing the ID-numbers (20.000 computer operations for a single alkane). We computed the ID-numbers for all alkane-trees<sup>9</sup> with up to 20 vertices (atoms). The following result was obtained. In the field of 618,050 alkanes (all alkanes with up to 20 carbon atoms) we found only 124 duplicates and 1 triplicate; i. e. 124 pairs and 1 triple of non-isomorphic alkanes having exactly the same ID-number. The smallest degenerate pair has 15 vertices. There is only one pair of non-isomorphic trees with the same ID-number among 4,347 different isomeric alkane-trees with 15 vertices. These two trees are shown in Figure 1.

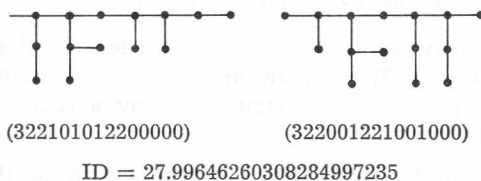


Figure 1. The smallest pair of alkane-trees with the same ID-number. Below each tree its  $N$ -tuple representation is given.<sup>5</sup>

Our result was communicated to Randić before the paper<sup>4</sup> containing it was submitted for publication. Randić, stimulated by our work,<sup>11</sup> immediately produced two interesting results.

Firstly, he considered all trees that have the same ID-numbers and found that the associated molecular path sequences differ in all cases considered,<sup>12</sup> sometimes already in the leading terms. Hence, he suggested that molecules should be first screened for their ID-numbers and if these happen to be the same, the corresponding molecular path sequences have to be compared to make sure whether or not there is a case of graph isomorphism.

The second result of Randić is more interesting. He redefined his descriptor in terms of weights based on prime numbers.<sup>13</sup> Note that Randić restricted the definition of the ID-number to molecular graphs depicting carbon compounds only, i. e. for each vertex in these graphs the degree is less than or equal to 4. The mapping  $g(e)$ , i. e. the connectivity weights and the prime number weights, are given in the Table.

Since the weights are computed by means of the number one and the first nine prime numbers (2, 3, 5, 7, 11, 13, 17, 19, 23) Randić called this

TABLE

*The Connectivity Weights and the Prime Number Weights*

Bond type (deg $u \cdot$ deg $v$ )	Connectivity weight $g(u, v)$	Prime number weight $g(u, v)$
(1, 1)	$1/\sqrt{1} = 1$	$1/\sqrt{1}$
(1, 2)	$1/\sqrt{2} = (1/2) \sqrt{2}$	$1/\sqrt{2}$
(1, 3)	$1/\sqrt{3} = (1/3) \sqrt{3}$	$1/\sqrt{3}$
(1, 4)	$1/\sqrt{4} = 1/2$	$1/\sqrt{5}$
(2, 2)	$1/\sqrt{4} = 1/2$	$1/\sqrt{7}$
(2, 3)	$1/\sqrt{6} = (1/6) \sqrt{6}$	$1/\sqrt{11}$
(2, 4)	$1/\sqrt{8} = (1/4) \sqrt{2}$	$1/\sqrt{13}$
(3, 3)	$1/\sqrt{9} = 1/3$	$1/\sqrt{17}$
(3, 4)	$1/\sqrt{12} = (1/6) \sqrt{3}$	$1/\sqrt{19}$
(4, 4)	$1/\sqrt{16} = 1/4$	$1/\sqrt{23}$

descriptor the prime ID-number.<sup>13</sup> In parallel to this name, we call the ID-number based on connectivity weights the connectivity ID-number.

For molecular graphs depicting carbon compounds, the connectivity ID-number is an element of

$$F = \mathbb{Q}(\sqrt{2}, \sqrt{3})$$

whilst the prime ID-number is an element of

$$F' = \mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{5}, \sqrt{7}, \sqrt{11}, \sqrt{13}, \sqrt{17}, \sqrt{19}, \sqrt{23})$$

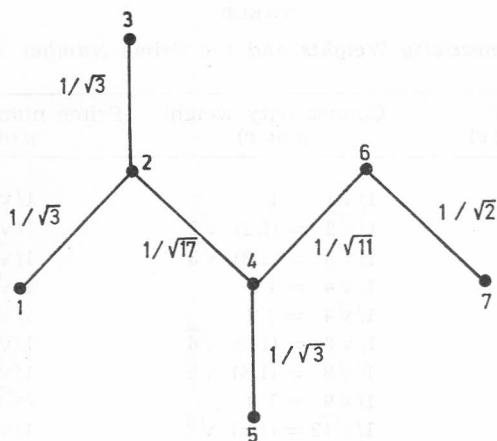
The basis of  $F$  as a vector space over the rational numbers  $\mathbb{Q}$  has the dimension 4 ( $=2^2$ ), whilst the dimension of  $F'$  over  $\mathbb{Q}$  is 512 ( $=2^9$ ).<sup>14</sup> Therefore, it may be expected that the representation of a molecular graph in  $F'$  would have a higher discriminating power.

Randić did not claim that the prime ID-number will be unique, but did invite the search for potential counterexamples.<sup>13</sup> As an example the calculation of the prime ID-number for 2,3-dimethylpentane is shown in Figure 2.

We calculated the prime ID-numbers for all alkane-trees with up to 20 vertices. This time we found only one case of degeneracy in the investigated field of alkanes (618,050 alkanes with up to 20 carbon atoms). The smallest (and the only) degenerate pair was found in the case of isomeric alkanes with 20 carbon atoms. This pair is given in Figure 3.

The result obtained shows, as anticipated, that the prime ID-number is a highly discriminating topological descriptor (index), much more discriminative than the connectivity ID-number, but it is still not unique.

Since the prime ID-number is quantity so discriminative, it may prove to be a valuable tool in chemical documentation, including possibly use in non-empirical approach as to quantitative structure-activity relationships (QSAR),<sup>15</sup> because all other currently used topological indices in QSAR are of limited discriminatory power.<sup>16-18</sup> Because of this, research oriented to design a single number as the unique descriptor of a molecule has been vigorously pursued in recent years.<sup>19</sup>



Path	Weights	
(1, 2)	$1/\sqrt{3}$	= 0.57735
(2, 3)	$1/\sqrt{3}$	= 0.57735
(2, 4)	$1/\sqrt{17}$	= 0.24254
(4, 5)	$1/\sqrt{3}$	= 0.57735
(4, 6)	$1/\sqrt{11}$	= 0.30151
(6, 7)	$1/\sqrt{2}$	= 0.70711
(1, 2, 3)	$1/\sqrt{3}$	= 0.33333
(1, 2, 4)	$1/\sqrt{3} \cdot \sqrt{17}$	= 0.14003
(2, 4, 5)	$1/\sqrt{3} \cdot \sqrt{17}$	= 0.14003
(2, 4, 6)	$1/\sqrt{3} \cdot \sqrt{17}$	= 0.07313
(3, 2, 4)	$1/\sqrt{3} \cdot \sqrt{17}$	= 0.14003
(4, 6, 7)	$1/\sqrt{2} \cdot \sqrt{11}$	= 0.21320
(5, 4, 6)	$1/\sqrt{3} \cdot \sqrt{11}$	= 0.17048
(1, 2, 4, 5)	$1/\sqrt{3} \cdot \sqrt{17} \cdot \sqrt{3}$	= 0.08085
(1, 2, 4, 6)	$1/\sqrt{3} \cdot \sqrt{17} \cdot \sqrt{11}$	= 0.04222
(2, 4, 6, 7)	$1/\sqrt{17} \cdot \sqrt{11} \cdot \sqrt{2}$	= 0.05171
(3, 2, 4, 5)	$1/\sqrt{3} \cdot \sqrt{17} \cdot \sqrt{3}$	= 0.08085
(3, 2, 4, 6)	$1/\sqrt{3} \cdot \sqrt{17} \cdot \sqrt{11}$	= 0.04222
(5, 4, 6, 7)	$1/\sqrt{2} \cdot \sqrt{3} \cdot \sqrt{11}$	= 0.12309
(1, 2, 4, 6, 7)	$1/\sqrt{3} \cdot \sqrt{17} \cdot \sqrt{11} \cdot \sqrt{2}$	= 0.02985
(3, 2, 4, 6, 7)	$1/\sqrt{3} \cdot \sqrt{17} \cdot \sqrt{11} \cdot \sqrt{2}$	= 0.02985

$$g^*(p) = 4.67408$$

$$\text{prime ID} = N + \sum_p g^*(p) = 7 + 4.67408 = 11.67408$$

Figure 2. Calculation of the prime ID-number for 2,3-dimethylpentane



### SAŽETAK Molekularni ID brojevi

Klaus Szymanski, Wolfgang R. Müller, Jan V. Knop i Nenad Trinajstić

U potrazi za protuprimjerom za Randićev prim ID broj ispitani su svi alkani do 20 ugljikovih atoma. U tom je skupu alkana pronađen samo jedan par stabala s identičnim prim ID brojevima. Na taj je način pokazano da prim ID broj, iako vrlo selektivan, nije jednoznačan molekularni deskriptor.

Abstract: In the search for a counterexample to Randić's primary ID number, all alkanes with up to 20 carbon atoms were examined. Only one pair of trees was found with identical primary ID numbers. This shows that the primary ID number, although very selective, is not a unique molecular descriptor.