Indicators for Atoms Included in Cycles

István Lukovits

Central Research Institute for Chemistry, Hungarian Academy of Sciences,
H-1525 Budapest, P.O. Box 17, Hungary

Received August 1, 1995[⊗]

A numerical method to identify vertices (atoms) involved in cycles has been presented.

INTRODUCTION

There is considerable interest in cyclic substructures of molecular graphs. The branching index, proposed by Randić[1] was used by Kier and Hall[2] to define the "connectivity index". The connectivity index was used as an independent variable in quantitative structure−property relationship (QSPR) studies. In Kier's and Hall's approach various substructures of the molecule like "stars", "circuits" (i.e., cycles), etc. were handled separately, and the connectivity index was obtained for such substructures and regressed against the dependent variable.

The detour index was proposed recently by Amić and Trinajstić[3] as a new graph invariant that may by used in QSPR studies. The detour index is identical with the Wiener index[4] in acyclic structures, but in cyclic structures, instead of the shortest path, the longest path between vertices i and j has to be considered. It was found[5] that the detour index combined with the Wiener index is quite efficient if a series consisting of cycle-containing and acyclic molecules is considered. It has to be noted that most graph invariants proposed so far are not efficient in this respect (molecular connectivity[2] and in some cases the Wiener index[6] are exceptions); therefore most of them were used in series composed of acyclic molecules only. The usefulness of the detour index is deflated by the fact that to date no method (but inspection) is available to compute this index. Once the vertices making up cycles were known, this task would also be easier to solve.

The conjugated circuit model also uses the concept of cycles to determine the number of all possible benzene, naphthalene, and anthracene like substructures in conjugated benzenoid hydrocarbons.[7,8]

Visual identification of various cycles in molecular graphs is a relatively simple task if there are not too many cycles to distinguish. Once, however, the structural formula of a molecule (that is its corresponding graph) is encoded by using the adjacency matrix, even the identification of a single cycle may become a rather time consuming undertaking. In order to solve this problem, the adjacency matrix **A** might be investigated. Matrix **A** consists of N rows and N columns, where N denotes the number of vertices (atoms) of graph. $A_{i,j} = 1$ if vertices i and j are connected by a chemical bond (edge), and $A_{i,j} = 0$ otherwise ($A_{i,i} = 0$). If $A_{i,j} = 1$ the vertices i and j are said to be adjacent. (The notions "graph" and "structural formula", "vertex" and "atom", and "edge" and "chemical bond" will be used interchangeably, hereafter.) In order to identify a triangle (i.e., a cycle made up of three vertices) all products $A_{i,j}A_{j,k}A_{k,i}$ have to be considered, a nonvanishing product indicates that vertices i, j, and k constitute a triangle. By using this method the time needed to identify cycles would increase rapidly with increasing size of the cycles and also with increasing values of N.

The "ring perception problem" has been addressed by many researchers; a good review on this topics is the paper by Downs et al.[9] listing 24 various numerical methods that were suggested to settle this question. The papers published since this review appeared include algorithms used to determine the smallest set of smallest rings[10−12] and the "chain message algorithm" by Balducci and Pearlman.[13] All these algorithms can be used to find the cycle basis of a graph and consequently to construct all cycles involved in that graph.

The aim of the present paper was to introduce an efficient method to identify a vertex contained in one or more cycles. For this purpose indicator $\Delta^{(n)}_i$ will be defined. $\Delta^{(n)}_i/2$ determines the number of n-cycles in which vertex i is involved, where $3 \leq n \leq 6$. The value of the indicator $\Delta^{(n)}_i$ depends only on the number of cycles that contain vertex i, but it does not depend on the actual position of i in the graph. $\Delta^{(n)}_i$ is zero for vertices which are not involved in any cycle. To obtain $\Delta^{(n)}_i$ the powers of the adjacency matrix will be used. The method is efficient if the number of operations is taken into account, and the time of computations is proportional to N^2. Only single edges between vertices were assumed. The indicator test may be carried out for any graph containing triangles, squares, and pentagons. Graphs containing hexagons can also be investigated, but the method will fail if a vertex appears in a hexagon and in two or more triangles or in a hexagon and in a square. This problem will be addressed in a subsequent paper because of its complexity. Note that the method is not capable of identifying the cycles themselves, but once the vertices making up the various cycles are known, the second step of the identification procedure could also be solved. The indicator is a kind of a "local vertex invariant".[14,15] On the other hand, the indicator resembles in some respect the configuration index proposed by Klin et al.[16]

In the next section the indicator test will be introduced, and several numerical examples will be discussed. Readers who are not familiar with chemical graph theory might want to consult a textbook on this topics.[17]

THE INDICATOR METHOD

The powers of the adjacency matrix **A** may be used to enumerate all walks (a walk is a sequence of vertices and edges, the same vertices and edges may occur more than

[⊗] Abstract published in *Advance ACS Abstracts,* January 1, 1996.

0095-2338/96/1636-0065$12.00/0

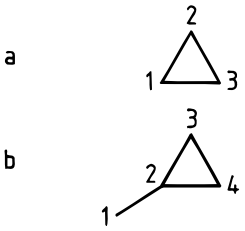

Figure 1.

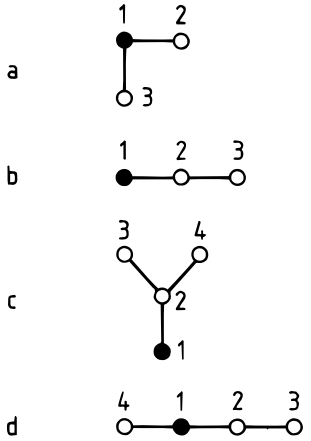

Figure 2.

once) between any pair of vertices *i*,*j*. $(A^n)_{ij}$, the *ij*th entry of matrix $\mathbf{A}^n$ is equal to the total number of walks between vertices *i* and *j*. A walk will be denoted by listing sequentially all vertices which are visited, and it will be assumed that consecutive vertices are adjacent. *ijkmim* therefore denotes a walk starting at vertex *i*, and encountering vertices *j*, *k*, *m*, *i* (again), and *m* in turn. An *n*-cycle (i.e., a cycle containing *n* vertices) will be identified by simply listing a self-returning walk around that cycle. Therefore 1231 denotes a self-returning walk of length three for a graph (a triangle) and the triangle itself (Figure 1a). The notion "length of a walk" denotes the number of vertices visited during the walk minus one. The length of walk 1231 is three. Each *n*-cycle gives rise to two walks of length *n*, from and to a particular atom. The walk may be a clockwise or a counterclockwise walk around that cycle. For atom 1 in Figure 1a these are 1231 and 1321.

$(A^2)_{i,i}$ is the number of the first neighbors of vertex *i*. In general $(A^n)_{i,i}$ is the total number of self-returning walks of length *n* at vertex *i*. Note that if *n* is odd and the graph is acyclic, then $(A^n)_{i,i} = 0$, since between any pair of vertices *i* and *j*, there is only one path. Therefore if a walk starts from vertex *i* and returns to vertex *i* all edges will be passed at least twice (or an integral multiple of two), and the length of the walk must be an even number.

From these facts it follows at once that $\Delta^{(3)}_i/2 = (A^3)_{i,i}/2$ is equal to the number of **triangles** in which vertex *i* is involved. $\#_i = (A^3)_{i,i}/2$ denotes the number of triangles containing vertex *i*.

In order to determine the number of **squares** (4-cycles) in which vertex *i* is involved, entry $(A^4)_{i,i}$ has to be investigated. All first neighbors will contribute to $(A^4)_{i,i}$, because 12121 and 12131 type walks (Figure 2a) and a 12321 type walk (Figure 2b) are possible. In general if there are $(A^2)_{i,i}$ first neighbors, the value of $(A^2)_{i,i} + (A^2)_{i,i}[(A^2)_{i,i} - 1]$ has to be subtracted from $(A^4)_{i,i}$. Further the sum of non-self-returning paths $\sum'(A^2)_{i,j}$ ($\sum'$ indicates that the summation has be done for all $j \neq i$) of length two has to be subtracted. The numerical procedure for a structure depicted in Figure 3a is demonstrated in Table 1. Indicator $\Delta^{(4)}_i = 2$ for vertices 1−4, which are contained in the cycle, and $\Delta^{(4)}_i = 0$, for vertices 5 and 6, which lie outside the cycle. Note that the presence of triangles cannot influence the value of the indicator $\Delta^{(4)}_i$, as entry $(A^3)_{i,i}$ is not used to obtain $\Delta^{(4)}_i$. This can be illustrated by considering the structure depicted in Figure 3b and Table 2. Vertices 1 and 2 appear in five, vertices 3 and 4 appear in four, and vertex 5 appears in two squares, namely 52315 and 52415 (Figure 3b).

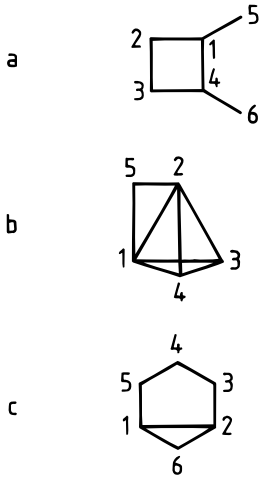

Figure 3.

Table 1. Simple Indicator Test Performed for Vertices of Graph a (Figure 3)

vertex/contribution	1	2	3	4	5	6
$(A^4)_{i,i}$	14	9	9	14	3	3
$-(A^2)_{i,i}$	3	2	2	3	1	1
$-(A^2)_{i,i}[(A^2)_{i,i}-1]$	6	2	2	6	0	0
$-\sum'(A^2)_{i,j}$	3	3	3	3	2	2
$\Delta^{(4)}_i$	2	2	2	2	0	0

Table 2. Combined Indicator Test Performed for Vertices of Graph b (Figure 3)

vertex/contribution	1	2	3	4	5
triangle					
$(A^3)_{i,i} \equiv \Delta^{(3)}_i$	8	8	6	6	2
square					
$(A^4)_{i,i}$	34	34	25	25	14
$-(A^2)_{i,i}$	4	4	3	3	2
$-(A^2)_{i,i}[(A^2)_{i,i}-1]$	12	12	6	6	2
$-\sum'(A^2)_{i,j}$	8	8	8	8	6
$\Delta^{(4)}_i$	10	10	8	8	4
pentagon					
$(A^5)_{i,i}$	102	102	76	76	34
$-10\#_i$	40	40	30	30	10
$-4N_i$	32	32	12	12	0
$-2T_i$	6	6	10	10	12
$-2t_i$	20	20	20	20	8
$\Delta^{(5)}_i$	4	4	4	4	4

The next test can be used to decide whether vertex *i* is contained in a **pentagon** (5-cycle). Each triangle in which vertex i is involved will contribute to $(A^5)_{i,i}$. There are ten possibilities to walk around a triangle and to return to the origin (Figure 1a) in five steps: 123121, 123231, 123131, 131321, 132321, 132121, 121321, 131231, 132131, and

Table 3. Combined Indicator Test Performed for Vertices of Graph c (Figure 3)

vertex/contribution	1	2	3	4	5	6
			triangle			
$(A^3)_{i,i} \equiv \Delta^{(3)}{}_i$	2	2	0	0	0	2
			square			
$(A^4)_{i,i}$	13	13	7	6	7	8
$-(A^2)_{i,i}$	3	3	2	2	2	2
$-(A^2)_{i,i}[(A^2)_{i,i}-1]$	6	6	2	2	2	2
$-\sum'(A^2)_{i,j}$	4	4	3	2	3	4
$\Delta^{(4)}{}_i$	0	0	0	0	0	0
			pentagon			
$(A^5)_{i,i}$	18	18	4	2	4	14
$-10\#_i$	10	10	0	0	0	10
$-4N_i$	4	4	0	0	0	0
$-2T_i$	0	0	2	0	2	0
$-2t_i$	2	2	0	0	0	4
$\Delta^{(5)}{}_i$	2	2	2	2	2	0
			hexagon			
$(A^6)_{i,i}$	66	66	31	24	31	42
$-\sum' p^{(3)}{}_{i,j}$	3	3	4	4	4	4
$-(A^2)_{i,i}$	3	3	2	2	2	2
$-3\sum'(A^2)_{i,j}$	12	12	9	6	9	12
$-2Y_i$	2	2	2	0	2	4
$-6C^{(2)}{}_i$	18	18	6	6	6	6
$-6C^{(3)}{}_i$	6	6	0	0	0	0
$-\sum'(A^2)_{i,j}[(A^2)_{i,i}-1]$	16	16	6	4	6	8
$-4\#_i$	4	4	0	0	0	4
$\Delta^{(6)}{}_i$	2	2	2	2	2	2

121231. Therefore $\#_i$ has to be multiplied by ten and subtracted from $(A^5)_{i,i}$.

A cyclopropylmethane substructure contributes additional self-returning walks of length five for all its vertices (Figure 1b); for the methane vertex 1:123421 and 124321; for the substituted cyclopropane vertex 2: 234212, 243212, 212342 and 212432; for each of the nonsubstituted cyclopropane vertices 3 and 4: 342123, 321243 and 421234, 432124.

To recognize a vertex in position 2 (Figure 1b) two necessary conditions have to be fulfilled: (1) $A_{1,2} = 1$, i.e., vertices i and j have to be adjacent and (2) $(A^3)_{2,2} - (A^2)_{1,2} = N_2 \neq 0$, which is the number of all triangles containing vertex 2 but do not involve vertex 1. The two necessary conditions are together sufficient to recognize a vertex in position 2 (Figure 1b). Variable $N_i = 1$ if vertex i is of type 2 and $N_i = 0$, otherwise. The contribution of vertex i of type 2 to $(A^5)_{i,i}$ is $4N_i$. Similarly a vertex in position 1 has to fulfill two necessary conditions: (1) $A_{1,2} = 1$; and (2) $(A^3)_{1,1} - (A^2)_{1,2} = T_1 \neq 0$. Variable $T_i = 1$ if vertex i is of type 1 and $T_i = 0$, otherwise. The contribution of vertex i of type 1 to $(A^5)_{i,i}$ is $2T_i$ (Tables 2 and 3).

In order to identify a vertex i in position 3 or 4 (Figure 1b) three necessary conditions have to be fulfilled: (1) $A_{i,k}A_{k,1} = 1$ for some $k \in N$ (in our example $k = 2$, and subscript 1 indicates a vertex of type 1); (2) $(A^3)_{i,i} \neq 0$; and (3) $(A^3)_{k,k} \neq 0$, where subscript k indicates a vertex of type 2. The contribution of vertices of type 3 (or 4) to $(A^5)_{i,i}$ is $2t_i$ where $t_i = 1$ if vertex i is of type 3 or 4, and $t_i = 0$, otherwise (note that N_i, T_i and t_i are indicators, themselves). After all subtractions indicator $\Delta^{(5)}{}_i$ will be obtained.

The actual calculations done for pentagons (Figure 3b,c) are illustrated in Tables 2 and 3. The structure depicted in Figure 3b contains two pentagons: 523415 and 524315 (Table 2); and the structure depicted in Figure 3c contains (Table 3) one triangle (1261) and a pentagon (123451).

In order to obtain the number of **hexagons** in which vertex i is involved, the entry $(A^6)_{i,i}$ has to be investigated. The number of paths of length three starting at i, for which $A_{i,j} = 0$ and $i \neq j$, is denoted by $\sum' p^{(3)}{}_{i,j}$. This figure has to be subtracted from $(A^6)_{i,i}$. The number of paths of length two, for which $i \neq j$, is equal to $\sum'(A^2)_{i,j}$. Each such path contributes three to $(A^6)_{i,i}$, as can be seen by inspecting the structure shown in Figure 2b, since the following three walks of length six are possible: 1212321, 1232321, and 1232121. The number of first neighbors at vertex i $(A^2)_{i,i}$ has also to be subtracted from $(A^6)_{i,i}$, because the following walk of length six is possible between neighbors (Figure 2a): 1212121. Let Y_i denote the number of first neighbors of vertex i, the degree of which is $n_j \geq 3$. The contribution to $(A^6)_{i,i}$ is equal to the ways two edges can be selected from the $n_j - 1$ outbound edges present, multiplied by 2. For example vertex 2 (Figure 2c) is the first neighbor of vertex 1 and $n_2 = 3$. Therefore $2Y_1 = 2 \times 1 = 2$, and the two possible walks are 1232421 and 1242321.

There are six possible walks if the two first neighbors are visited (Figure 2a): 1212131, 1213121, 1312121, 1213131, 1312131, and 1313121. In general if $C^{(2)}{}_i$ denotes the number of ways two different vertices can be selected from $(A^2)_{i,i}$ neighbors present, $6C^{(2)}{}_i$ has to be subtracted from $(A^6)_{i,i}$. If $C^{(3)}{}_i$ denotes the number of ways three vertices (Figure 2c) can be selected from $(A^2)_{i,i}$ neighbors present, $6C^{(3)}{}_i$ has to be subtracted from $(A^6)_{i,i}$.

A walk of length six can also be completed by starting at a path of length two returning to the origin and then proceeding to a first neighbor or vice versa (Figure 2d). In our example the first possibility is 1232141, and there are $\sum'(A^2)_{i,j}[(A^2)_{i,i}-1]$ such walks. The second possibility is 1412321, and there are $\sum'(A^2)_{i,j}[(A^2)_{i,i}-1]$ such cases. Altogether the value of $2\sum'(A^2)_{i,j}[(A^2)_{i,i}-1]$ has to be subtracted from $(A^6)_{i,i}$.

Finally if a single triangle is present and vertex i is involved in that triangle (i.e., $(A^3)_{i,i}/2 = \#_i = 1$ (and $\#_i = 0$, otherwise), number $4\#_i$ has to be subtracted from $(A^6)_{i,i}$, since there are four walks of length six around a triangle (Figure 1a): 1231231, 1231321, 1321321, and 1321231. After these manipulations we obtain indicator $\Delta^{(6)}{}_i$. It has to be noted that a vertex which is adjacent to a vertex involved in a triangle will not affect $\Delta^{(6)}{}_i$. On the other hand, squares and vertices adjacent to squares as well as connected triangles also have to be considered. The method presented for hexagons can only be used, if no squares are present, and there is only a single triangle (Figure 3c). The method will be improved to include these cases also in a subsequent paper.

The number of elementary operations needed for this algorithm is nN^2 (the number of necessary multiplications to obtain $\mathbf{A}^n$) plus the number of steps needed to calculate all contributions that have to be subtracted from $(A^n)_{i,i}$. For $n \leq 6$ this figure is also a multiple of N^2, but with increasing the number of n the number of substructures to be investigated may increase more rapidly. The second step of the ring perception problem, the assignment of the selected vertices to various cycles has yet to be solved. In this step for each vertex i, $n - 1$ other vertices have to be selected from the pool of $N - 1$ vertices. The selection procedure would involve two steps: (1) Determine whether vertex j (j

$\neq i$) is of the same type as i (i.e., whether j is also involved in an n-cycle), and (2) determine whether distance $d_{i,j} \leq [n/2]$, the brackets denote that integral value of the resulting quotient has to be considered. The number of necessary steps needed for this step may also be of order N^2 if the distance matrix is used for the identification procedure. Therefore the efficiency of the present approach can be compared with that of Balducci and Pearlman.[13]

The substructures contributing to $(A^n)_{i,i}$ were found by inspection. With increasing n (i.e., large cycles) this procedure will be more and more complicated. A general scheme to find automatically all substructures that contribute to $(A^n)_{i,i}$ has to be found.

ACKNOWLEDGMENT

The author is indebted to Dr. Ch. Rückert for many helpful suggestions concerning the presentation of this material.

REFERENCES AND NOTES

(1) Randić, M. On Characterization of Molecular Branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609−6615.
(2) Kier, L. B.; Hall, L. H. Molecular Connectivity in Chemistry and Drug Research; Academic Press: New York, 1976; p 48.
(3) Amić, D.; Trinajstić, N. On the Detour Matrix. *Croat. Chem. Acta* **1995**, *68*, 53−62.
(4) (a) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, *69*, 17−20. (b) Wiener, H. Correlation of Heats of Isomerization and Differences in Heats of Vaporization of Isomers among the Paraffin Hydrocarbons. *J. Am. Chem. Soc.* **1947**, *69*, 2636−2638.
(5) Lukovits, I. The Detour Index. *Croat. Chem. Acta* Submitted for publication.
(6) Lukovits, I. The Generalized Wiener Index for Molecules Containing Double Bonds and the Partition Coefficients. *Reports Mol. Theor.* **1990**, *1*, 127−131.
(7) Randić, M. Conjugated Circuits and Resonance Energies of Benzenoid Hydrocarbons. *Chem. Phys. Lett.* **1976**, *38*, 68−70.
(8) Nikolić, S.; Trinajstić, N.; Klein, D. J. The Conjugated-Circuit Model. *Comput. Chem.* **1990**, *14*, 313−322.
(9) Downs, G. M.; Gillet, V. J.; Holliday, J. D.; Lynch, M. F. Review of Ring Perception Algorithms for Chemical Graphs. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 172−187.
(10) Qian, C.; Fisanick, W.; Hartzler, D. E.; Chapman, S. W. Enhanced Algorithm for Finding the Smallest Set of Smallest Rings. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 105−110.
(11) Baumer, L.; Sala, G.; Sello, G. Ring Perception in Organic Structures: A New Algorithm for Finding SSSR. *Comput. Chem.* **1991**, *15*, 293−299.
(12) Fan, B. T.; Panaye, A.; Doucet, J. P.; Barbu, A. Ring Perception. A New Algorithm for Directly Finding The Smallest Set of Smallest Rings for a Connection Table. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 657−662.
(13) Balducci, R.; Pearlman, R. S. Efficient Exact Solution of the Ring Perception Problem. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 822−831.
(14) Balaban, T. S.; Filip P. A.; Ivanciuc, O. Computer Generation of Acyclic Graphs Based on Local Vertex Invariants and Topological Indices - Derived Canonical Labelling and Coding of Trees and Alkanes. *J. Math. Chem.* **1992,** *11*, 79−105.
(15) Balaban, T. A. Using Real Numbers as Vertex Invariants for Third-Generation Topological Indexes. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 23−28.
(16) Klin, M. H.; Tratch, S. S.; Zefirov, N. S. 2D-Configurations and Clique-Cyclic Orientations of the Graphs $L(K_p)$. *Reports Mol. Theor.* **1990**, *1*, 149−163.
(17) Trinajstić, N. *Chemical Graph Theory*, 2nd ed.; CRC Boca Raton, FL, 1992; p 1.

CI950082O