Generalization of the Graph Center Concept, and Derived Topological Centric Indexes

D. BONCHEV,*† A. T. BALABAN,‡ and O. MEKENYAN†

Department of Physical Chemistry, Higher School of Chemical Technology, 8010 Burgas, Bulgaria, and Department of Organic Chemistry, The Polytechnic, Bucharest, Roumania

Received June 27, 1979

The concept of topological center is generalized for any connected acyclic or cyclic graph on the basis of the topological distance matrix using four criteria in a specified order. The center point(s) provide(s): (1) the smallest distance to any other vertex; (2) the smallest sum of distances to all other vertexes; (3) the smallest number of times the maximum distance occurs in the distance code, and (4) constancy on repeating criteria (1)–(3) to the pseudocenter graph containing only the vertexes selected by the previous criteria and their incident edges. Two partition schemes of the graph vertexes are proposed: the classical partitions, based on the above four criteria in turn, and the generalized partitions based primarily on the distance to the polycenter vertex (or vertexes, respectively) and secondarily on the above criteria. Topological indexes based on each of these partitions are devised, making use either of a quadratic formula (8) or of Shannon's information function (9). The most sophisticated topological indexes also take into account the order of numbers in the partition by means of Muirhead's scheme. The new centric indexes may be used for topological correlations, for estimating the topological shape of isomeric chemical structures, and for use in coding and computer processing of chemical information.

I. GENERALIZATION OF THE CONCEPT OF GRAPH CENTER

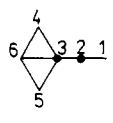
1. First Criterion of Minimum Maximal Distance. As known,[1-3] the concept of the topological center of a graph is based on topological distances between the graph vertexes. By the classical definition, in a graph on p vertexes, the center vertexes have the smallest maximal distance to the other vertexes; i.e., if vertex i belongs to the center vertexes,

$$d_{ij}^{\max} = \min \text{ for } j = 1,2, \ldots, p \qquad (1)$$

For acyclic graphs, a classical result of early graph theory[1-3] discovered independently by Jordan and Sylvester is that any tree has a unique *center* which is either one vertex or two adjacent vertexes; in the latter case, the center is called a *bicenter*. For cyclic graphs, however, we shall show that the classical definition functions only as a first criterion, and that, to make it operational, additional criteria must be introduced.

The determination of the graph center(s) is easily carried out on the basis of the topological distance matrix **D**. This is[1-3] a square $p \times p$ matrix symmetrical in respect to its main diagonal. The topological distance d_{ij} between two vertexes i and j is equal to the number of bonds (edges) connecting the two vertexes along the shortest path between them. In the present paper, all graphs will be assumed to be connected, so that there is a path between any two vertexes. Then, by definition, d_{ij} is an integer which in **D** appears at the intersection of row/column i/j and j/i. When i and j are neighbor vertexes, we have $d_{ij} = 1$.

Example:

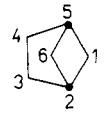


	1	2	3	4	5	6	$d_{ij}^{\max}$
1	0	1	2	3	3	3	3
2	1	0	1	2	2	2	2
3	2	1	0	1	1	1	2
4	3	2	1	0	2	1	3
5	3	2	1	2	0	1	3
6	3	2	1	1	1	0	3

58

Vertexes 2 and 3 have a maximum distance to any other vertex j, $d_{2j}^{\max} = d_{3j}^{\max} = 2$, while for the other four vertexes, $d^{\max} = 3$. Thus, following the classical definition, vertexes 2 and 3 are centers of this graph. Here and in the following, the centers will be colored in black. The maximum distance of the center(s) is called the *radius* of the graph, $r = d_{\text{center}}^{\max}$; e.g., for the above graph $r = 2$.

2. Second Criterion of Minimum Distance Sum (Rank). The classical definition of graph center is insufficient for many cyclic graphs. This classical definition leads in most cases to a large number of center vertexes, thus reducing sharply the efficiency of the notion of graph center. An illustration of this weakness of the classical definition is provided by the graph below, numbered **61** in Table I. The rows and columns are numbered as in the previous example, though the numbering is no longer shown.



$\mathbf{D} =$

						$d^{\max}$	d_r
0	1	2	2	1	2	2	8
1	0	1	2	2	1	2	7
2	1	0	1	2	2	2	8
2	2	1	0	1	2	2	8
1	2	2	1	0	1	2	7
2	1	2	2	1	0	2	8

61

According to the classical definition, all vertexes of this graph should be regarded as centers since they have the same $d^{\max} = 2$. Clearly, this definition of graph center leads to a high degeneracy which does not agree with our intuitive notion of a center.

One reaches the conclusion that additional criteria for graph center are necessary, in order to formalize better the intuitive ideas for the center of cyclic graphs and to reduce the number of center vertexes as much as possible. Taking the classical definition only as a first criterion, we suggest four criteria for the determination of graph center, which are to be applied in the specified order; i.e., the three additional criteria are each subordinate to the preceding one(s).

We proceed from the idea that the graph center should contain the innermost point(s) in the graph, i.e., that a center vertex should be less distant to all other vertexes than a vertex which does not belong to the center. This idea is easily expressed quantitatively by the requirement that the sum of distances of a center vertex to all other vertexes in the graph should be minimum. This sum of distances for vertex i is equal to the sum of all entries in the row column i of the distance matrix **D**, and is called the *total distance rank* of vertex i, $d_{r,i}$. Therefore, the second criterion for graph center is:

$$d_{r,i} = \sum_{j=1}^{p} d_{ij} = \min \qquad (2)$$

† Higher School of Chemical Technology
‡ The Polytechnic.

0095-2338/80/1620-0106$01.00/0

Table I. All Possible Graphs with $p \leq 6$ and Vertex Degree ≤ 4 with Their Polycenter as Black Points

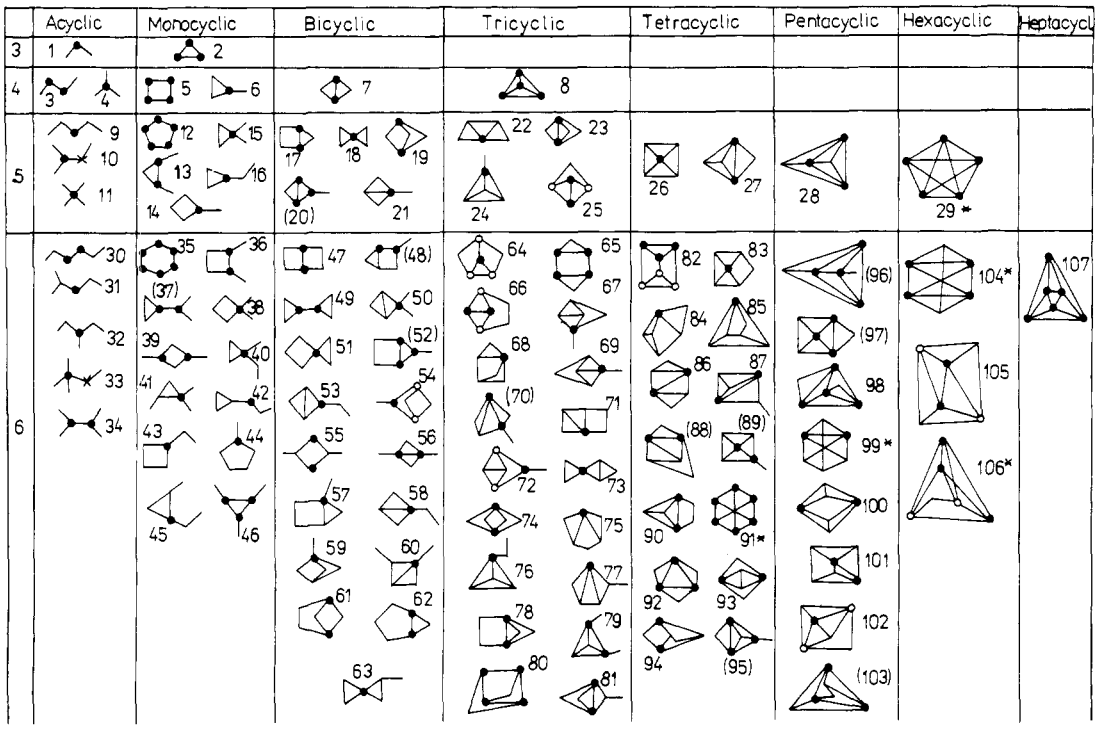


Table II. Denominations Used in the Present Paper for Graph Centers and Shells Based on the Four Criteria Discussed in the Text, as Well as for the Partitions and Topological Indexes Derived from the Distance Matrix

		shells		partitions of p vertexes		topological indexes[a]	
criterion	center	classical (CS)	generalized (GS)	classical (CP)	generalized (GP)	classical (CI)	generalized (GI)
1. min-max distance to any other vertex	classical center	radial shells RCS	RGS	radial partitions RCP	RGP	radial indexes RCI	RGI
2. min distance sum (rank)	rank center	DRCS	DRGS	DRCP	DRGP	DRCI	DRGI
		distance rank subshells		distance rank partitions		distance rank indexes	
3. min frequency of max distance	pseudocenter	distance code subshells DCCS	DCGS=PGS	distance code partitions DCCP	DCGP=CGP	distance code indexes DCCI	DCGI=CGI
4. iteration of criteria 1-3 to constancy of pseudograph	polycenter	polycenter subshells PCS	-	complete partition CCP	-	complete indexes CCI	-
taking the order of nos. in partitions into account	-	-	-	Muirhead partition MCP	MGP	Muirhead indexes MCI	MGI

[a] Each of the topological indexes can be quadratic (8) or logarithmic or informational (9).

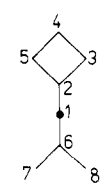
On dividing the total distance rank d_r by the number p of vertexes in the graph, one obtains the mean distance rank, which is

$$\bar{d}_{r,i} = d_{r,i}/p \qquad (3)$$

and which, when minimal, also illustrates the idea that the center point(s) in a graph must have the smallest average distance to any other vertex. Expressions 2 and 3 are equivalent for a given graph, and serve for discriminating among vertexes i which satisfy the first criterion (1).

The total or mean distance ranks of the graph vertexes are easily determined with the help of matrix **D**. For the preceding graph **61**, the total distance rank is presented in the second vertical column following the matrix. The second criterion indicates vertexes 2 and 5 as graph centers since their total distance rank, 7, is smaller than that of the other vertexes, 8. The result thus obtained contains fewer vertexes, and agrees better with intuition, than what the classical definition (first criterion) indicated.

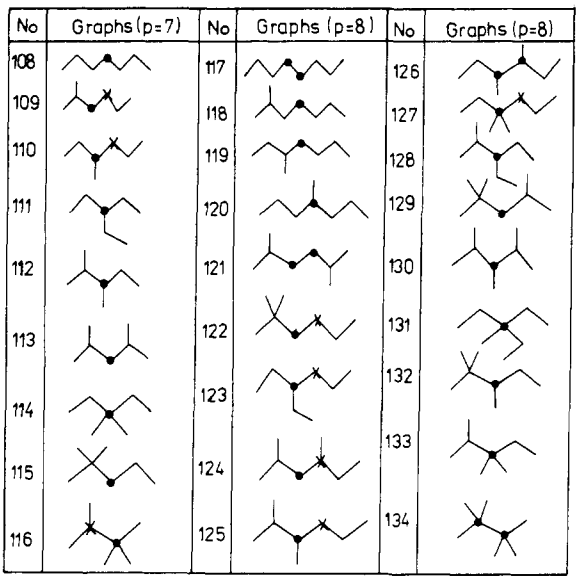
3. Third Criterion of Minimum Number of Times the Largest Distance Occurs. The definition of graph centers still needs additional requirements in order to reduce further the number of vertexes and to provide a more convenient topological index as will be shown in section II. When criteria 1 and 2 fail to discriminate among several vertexes, one can introduce a third criterion by comparing the *distance codes* of these vertexes. We use the term *vertex distance code* for the shorthand notation of all entries in a row or column corresponding to a certain vertex; the entries are arranged in nondecreasing order, and, to save space, when a distance appears a number of times this number is written as a superscript. An example is shown.



Distance matrix $\underline{D}$	d^{max}	d_r	distance code
0 1 2 3 2 1 2 2	3	13	$1^2\,2^4\,3^1$
1 0 1 2 1 2 3 3	3	13	$1^3\,2^2\,3^2$
2 1 0 1 2 3 4 4	4	17	$1^2\,2^2\,3^1\,4^2$
3 2 1 0 1 4 5 5	5	21	$1^2\,2^1\,3^1\,4^1\,5^2$
2 1 2 1 0 3 4 4	4	17	$1^2\,2^2\,3^1\,4^2$
1 2 3 4 3 0 1 1	4	15	$1^3\,2^1\,3^2\,4^1$
2 3 4 5 4 1 0 2	5	21	$1^1\,2^2\,3^1\,4^2\,5^1$
2 3 4 5 4 1 2 0	5	21	$1^1\,2^2\,3^1\,4^2\,5^1$

The vertex distance code is similar to the Randić atomic

Table III. Centers of Acyclic Graphs Having Seven and Eight Vertexes



code[4-6] written in the same concise manner but based on all possible paths (and not, as in the present case, on distances which are the shortest paths only) between the graph vertexes. It is easily seen that the two codes are identical for acyclic graphs (trees).[7]

In the above example, the first two criteria indicate both vertexes 1 and 2 as graph centers, but vertex 1 has a different distance code from vertex 2. Proceeding again from the intuitive idea that a graph center should contain the innermost point(s) of the graph, if two or more vertexes have the same values for d^{max} and d_r but different distance codes, then those of them will belong to the graph center, in which the largest distance occurs the fewest times. In the above example, the largest distance $d^{max} = 3$ occurs twice for vertex 2, but only once for vertex 1; therefore, vertex 1 will be the center according to the third criterion.

If the largest distance occurs the same number of times when comparing different distance codes, one should compare stepwise the next lower distance, choosing as center the point(s) for which the occurrence of this second, or third, etc., maximum distance is lower; let two vertexes, 1 and 2, have different distance codes but not be discernible by the first two criteria. We note the distance codes:

$$\text{vertex 1: } d_1^{a_1},d_2^{b_1},d_3^{c_1},\ldots,(d_{max}-2)^{k_1},(d_{max}-1)^{l_1},d_{max}^{m_1}$$

$$\text{vertex 2: } d_1^{a_2},d_2^{b_2},d_3^{c_2},\ldots,(d_{max}-2)^{k_2},(d_{max}-1)^{l_2},d_{max}^{m_2} \quad (4)$$

Of course, equality of criterion 2 requires that

$$d_1a_1 + d_2b_1 + \ldots + d_{max}m_1 = d_1a_2 + d_2b_2 + \ldots + d_{max}m_2 \quad (5)$$

Some of the distances (but not d_{max}) may occur only for one of the two vertexes to be compared, provided that relation 5 is valid.

Criterion 3 states that in the above case (4) vertex 1 belongs to the graph center if:

$$m_1 < m_2 \quad (6)$$

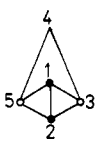
$$\text{or, when } m_1 = m_2, \text{ if } l_1 < l_2 \quad (6')$$

$$\text{or, when } m_1 = m_2 \text{ and } l_1 = l_2, \text{ if } k_1 < k_2 \quad (6'')$$

etc.

4. Fourth Criterion: Iterations of Criteria 1-3. By applying in the given order criteria 1, 2, and 3, we obtain a number of vertexes which constitute what we term the *pseudocenter* of the graph. To obtain the final reduction of the number of vertexes in the center, the fourth criterion is applied. It consists in deleting from the initial given graph all vertexes but the pseudocenter vertexes, and all edges but those whose both endpoints are pseudocenter vertexes, resulting in a *pseudocenter graph*; criteria 1-3 are applied in the same order to the pseudocenter graph. The operation is repeated for the pseudocenter graph, and so on, until the pseudocenter remains unchanged after two consecutive applications of criteria 1-4. The result is the kind of center we have looked for. To differentiate it from the classical center we shall call it the *polycenter* of the graph. For use in the subsequent discussion, we shall denote the kind of center resulting from the application of each of the criteria 1-4 by the names indicated in Table II.

An example for application of criterion 4 follows. In the



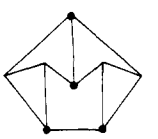
	1 2 3 4 5	d^{max}	d_r	Dist.code	Pseudograph d^{max}
D=	0 1 1 2 1	2	5	$1^3\,2^1$	1
	1 0 1 2 1	2	5	$1^3\,2^1$	1
	1 1 0 1 2	2	5	$1^3\,2^1$	2
	2 2 1 0 1	2	6	$1^2\,2^2$	-
25	1 1 2 1 0	2	5	$1^3\,2^1$	2

example, the pseudograph contains four of the five initial vertexes of the graph; application of criterion 4 eliminates from the pseudograph vertexes 3 and 5, leaving the polycenter which contains vertexes 1 and 2.

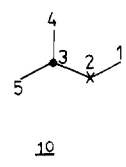
In practice, the pseudocenter is identical with the polycenter so that criterion 4 is no longer necessary, whenever criteria 1-3 result in a pseudocenter graph formed by:

- one vertex (**6**) or several isolated (nonadjacent) vertexes (**19**)
- A cyclic graph C_n formed by an n-membered ring of vertexes with degree 2, where $n \geq 3$ (**5**), or several isolated cyclic graphs $c_n, c_m, \ldots$
- a complete graph K_n, where each vertex has degree $n - 1$ and is adjacent to all $n - 1$ other vertexes, or several isolated graphs $K_n, K_m, \ldots$.

An example for the former case of one graph K_2 which is a pair of adjacent vertexes is **7**. An example for the former case of one graph $K_3 = C_3$ is **20**. An example for the former case of one graph K_4 is **8**, where the tetrahedron graph coincides with its pseudocenter and polycenter because it has four equivalent vertexes (i.e., in the same orbit, in graph-theoretical language). In all above examples, the numbers indicate graphs with their polycenters (as black points) and pseudocenters (as white points) in Table I. An example for the latter case of two isolated K_2 graphs in the polycenter is shown below.



The new definition of graph center (which we have termed polycenter) applies equally well to cyclic and acyclic graphs, and in all cases results in a restricted number of vertexes in the polycenter (unless the initial graph is a cyclic or a complete graph, when no reduction of the number of vertexes is possible because they are all equivalent). As in the classical definition, the polycenter for acyclic graphs is either a vertex or a pair of adjacent vertexes by virtue of criterion 1 which is identical with the classical definition. However, in some cases, acyclic graphs which in the classical definition possess a bicenter result in the new definition with one vertex as the polycenter. Such is the case of graph **10** and of several other acyclic graphs from Tables I and III. In such cases, Tables I and III indicate by



10

Matrix D	d^{max}	d_r
0 1 2 3 3	3	
1 0 1 2 2	2	6
2 1 0 1 1	2	5
3 2 1 0 2	3	
3 2 1 2 0	3	

a cross the vertex of the bicenter which was eliminated by criteria 2–4.

It must be emphasized that in applying criteria 1–4, a subsequent criterion is used only for discriminating among vertexes which give the same values for the preceding criteria, and never overrules a preceding criterion. In other words, criterion 1 is always strongest, criterion 2 is stronger than 3 and 4, and criterion 3 is stronger than criterion 4.

Table I presents all possible graphs[1] with three to six vertexes with degrees ≤4, indicating by black points the polycenter and, where the fourth criterion had to be applied, the pseudocenter vertexes as white points (more exactly, the white points indicate the vertexes of the pseudocenter graph which are deleted by virtue of the fourth criterion leaving the polycenter vertexes). In Table I the graphs are arranged according to the number of vertexes and to the number of their cycles. The latter is the *cyclomatic number* $\mu = q - p + 1$, where q and p are the numbers of edges and vertexes in the graph.[1] For acyclic graphs, a cross indicates a bicenter vertex which does not belong to the polycenter. An asterisk following the number of a graph indicates a nonplanar graph, i.e., a graph which cannot be embedded in a plane without having crossing edges; so far no constitutional graph is known in chemistry which is nonplanar.

As seen in Table I, most graphs have equivalent vertexes in the polycenter, i.e., vertexes belonging to the same orbit. However, a few graphs whose number (notation) is enclosed in brackets in Table I, have nonequivalent vertexes in the polycenter. For such graphs, a fifth criterion may be applied separating the vertexes into orbits of equivalent vertexes, but this will be discussed in a separate paper.[8]

II. TOPOLOGICAL INDEXES BASED ON THE NEW CONCEPT OF CENTER

One may construct on the basis of the new definition of graph center (polycenter) two partition schemes for graph vertexes, as will be shown below, eliminating stepwise those vertexes which lie at the periphery of the graph, as well as their incident edges, the polycenter vertexes being eliminated last. The numbers of vertexes deleted at each step form a sequence (partition of p) which may be used for obtaining a topological index for the given graph.

The graph is therefore considered as consisting of a number of shells and subshells, named according to Table II, and giving rise to the two series of partitions presented in Table II: more distant shells surround the inner shells, and the polycenter forms the nucleus of the graph. Any elimination procedure starts with the outermost shell and ends with the nucleus of the graph, but the intermediate subshells may be defined in more than one way.

For constructing a topological index on the basis of a partition, we shall proceed by analogy with the index of Gutman et al.[9] on the basis of the partition of vertex degrees, and more specifically by close analogy with the centric index; recently, on the basis of the classical definition of graph center, the first centric index was suggested by Balaban[10] for acyclic graphs, on the basis of the so-called pruning (lopping) partition of vertexes. All vertexes of degree one in the tree and the edges incident to them are simultaneously deleted at each step of the elimination procedure. The numbers δ_i of vertexes deleted at each step i form the pruning sequence. For instance, the tree presented below has a pruning sequence 5,2,1 (sequence S).

The centrix index B is defined by Balaban[10] as the sum of the squares of partition numbers δ_i:

$$B = \sum_i \delta_i^2 \tag{7}$$

and for the above acyclic graph we have $B = 1^2 + 2^2 + 5^2 = 30$.

The pruning partition cannot be directly applied in the case of cyclic graphs. Instead, one has to make use of the partitions defined below, and one can calculate a centric index C_r by an equation analogous to (7):

$$C_r = \sum_i r_i^2 \tag{8}$$

In eq 8, r_i is the number of vertexes belonging to the same shell or subshell, and deleted at any single step when proceeding toward the polycenter.

Each partition scheme can be treated also by means of Information Theory,[11-13] defining an informational index (information content) of this partition. Recently, a number of information indexes were proposed for chemical structures,[14-17] most of them based on different graph invariants by Bonchev et al. Such an informational centric index can be determined by using an equation derived from Shannon's information content of a sequence of numbers r_i as in (8):

$$I_{c,r} = \sum_i r_i \log_2 \sum_i r_i - \sum_i r_i \log_2 r_i \tag{9}$$

The logarithm in base 2 is taken for calculating the index in bits.[14,15]

1. Topological Indexes Based on "Classical" Radial Partitions. The first of the above four criteria provides an ordering of vertexes into *classical radial shells*, according to their maximum distance to any other vertex. Thus a *classical radial partition* of vertexes is induced which may be used to obtain topological indexes according to relations 8 and 9. Their discriminating power is, however, small.

The second criterion subdivides further each group of vertexes indicated by the previous criterion, according to the distance rank, leading to a *distance rank classical partition.* A further splitting of groups of vertexes is induced on the basis of the third criterion leading to a *distance code classical partition.*

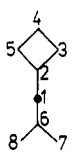
Finally, by means of the fourth criterion, a *complete classical partition* is obtained which by relations 8 and 9 yields the corresponding highly discriminating topological indexes.

The above partitions are illustrated in Table IV with the help of five examples: the first four have as polycenter one vertex, and the fifth has a polycenter containing two adjacent vertexes; i.e., its polycenter graph is a complete graph K_2. The increasing discriminating ability of the partitions is illustrated by the first four examples, where each of the four criteria discriminates respectively among groups of vertexes indicated by the previous criterion.

2. Topological Indexes Based on "Generalized" Radial Partitions. A different approach for obtaining a partition of graph vertexes is based specifically on the idea of graph center and of the distance between a given vertex and the center. Since in a cyclic graph we have a polycenter, the average distance between the given vertex and the polycenter vertexes must be considered; this is

$$d_c = \frac{1}{v} \sum_i^v d_{c,i} \tag{10}$$

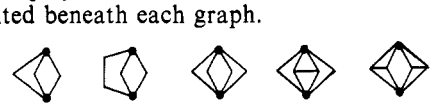
Table IV. Examples of Cyclic Graphs Whose Polycenters Are Found by Means of Criteria 1-4, and of Derived Classical or Generalized Partitions

graph	distance matrix	criteria and classical partition				generalized partition	
		1	2	3	4	d_c	deletion order
	0 1 1 2 2	2	6	$1^2 2^2$	-	1	a′
	1 0 1 2 2	2	6	$1^2 2^2$	-	1	a″
	1 1 0 1 1	1	4	1^4	-	0	b
	2 2 1 0 1	2	6	$1^2 2^2$	-	1	a‴
18	2 2 1 1 0	2	6	$1^2 2^2$	-	1	a⁗
	partitions	4;1	4;1	4;1	4;1	4;1	4;1
	0 1 2 2 3 2	3	10	$1^1 2^3 3^1$	-	1	b
	1 0 1 1 2 1	2	6	$1^4 2^1$	-	0	e
	2 1 0 1 2 2	2	8	$1^2 2^3$	-	1	c′
	2 1 1 0 1 2	2	7	$1^3 2^2$	-	1	d
	3 2 2 1 0 1	3	9	$1^2 2^2 3^1$	-	2	a
57	2 1 2 1 1 0	2	8	$1^2 2^3$	-	1	c″
	partitions	2;4		1,1;2,1,1		1;4;1	1;1,2,1;1
	0 1 2 3 2 1 2 2	3	13	$1^2 2^4 3^1$	-	0	f
	1 0 1 2 1 2 3 3	3	13	$1^3 2^2 3^2$	-	1	e
	2 1 0 1 2 3 4 4	4	17	$1^2 2^2 3^1 4^2$	-	2	c′
	3 2 1 0 1 4 5 5	5	21	$1^2 2^1 3^1 4^1 5^2$	-	3	a
	2 1 2 1 0 3 4 4	4	17	$1^2 2^2 3^1 4^2$	-	2	c″
	1 2 3 4 3 0 1 1	4	15	$1^3 2^1 3^2 4^1$	-	1	d
	2 3 4 5 4 1 0 2	5	21	$1^1 2^2 3^1 4^2 5^1$	-	2	b′
	2 3 4 5 4 1 2 0	5	21	$1^1 2^2 3^1 4^2 5^1$	-	2	b″
	partitions	3;3;2	3;2,1;2	1,2;2,1;1,1		1;4;2;1	1;2,2;1,1;1
	0 1 2 1 2 1	2	7	$1^3 2^2$	2	1	b′
	1 0 1 2 2 2	2	8	$1^2 2^3$	-	2	a′
	2 1 0 1 1 2	2	7	$1^3 2^2$	2	1	b″
	1 2 1 0 1 2	2	7	$1^3 2^2$	1	0	c
	2 2 1 1 0 1	2	7	$1^3 2^2$	2	1	b‴
64	1 2 2 2 1 0	2	8	$1^2 2^3$	-	2	a″
	partitions	6	2,4	2,4	2,3,1	2;3;1	2;3;1
	0 1 1 2 2 3	3	9	$1^2 2^2 3^1$	-	1.5	b
	1 0 1 1 2 3	3	8	$1^3 2^1 3^1$	-	1.5	c
	1 1 0 2 1 2	2	7	$1^3 2^2$	-	0.5	e′
	2 1 2 0 1 2	2	8	$1^2 2^3$	-	1.5	d
	2 2 1 1 0 1	2	7	$1^3 2^2$	-	0.5	e″
48	3 3 2 2 1 0	3	11	$1^1 2^2 3^2$	-	1.5	a
	partitions	3;3		1,1,1;1,2		4;2	1,1,1,1;2

where v is the number of vertexes in the polycenter, and $d_{c,i}$ is the topological distance between the given vertex and a vertex i in the polycenter; the sum is taken over all v vertexes in the polycenter. Unlike topological distances encountered so far, d_c no longer must be an integer; as seen from Table IV, in the case of the last graph, d_c is 0.5 for the center vertexes and 1.5 for the other vertexes.

The values of d_c induce a partition of vertexes which will be called the generalized radial partition. By means of relations 8 and 9 this partition can be converted into topological indexes.

The generalized partitions of some graphs from Table I shown below, whose polycenter vertexes are two nonadjacent points, are presented beneath each graph.



	19	61	74	93	100
Radial GP	5	2;4	6	6	6
Distance Rank GP	3,2	2;2,2	4,2	2,2,2	4,2

It may be seen that, in such cases, the generalized radial partition deletes some vertexes which do not belong to the polycenter together with the polycenter vertexes. The more refined partitions, starting with the distance rank generalized partition (see below), can discriminate the polycenter vertexes which are deleted last. Graph **61** is the only case out of 134 examined where this should be done by definition since even the refined procedures fail to discriminate the polycenter from some other vertexes.

The ordering induced by d_c can be compounded with that given by the four criteria discussed above, in order to enhance the discriminating ability of the topological indexes. In practice, however, as observed from Table IV, the generalized radial partition is about as discriminating as the distance rank classical partition, and therefore more discriminating than the classical radial partition. Thus, usually one does not gain much by refining the generalized radial partition with the help of the first criterion (classical ordering after values of d^{max}). We shall therefore subdivide the groups induced by d_c only by compounding with criteria 2 and 3 (criterion 4 is preliminary used on the determination of the polycenter) and obtain therefrom the *distance rank* and *complete generalized partitions*, respectively. These are illustrated in Table IV. Each of these generalized partitions yields according to relations 8 and 9 topological indexes. In Table IV, letters a, b, c, . . ., denote the order in which vertices are removed stepwise proceeding from the outermost shell toward the nucleus of the graph; when more than one vertex is to be deleted in the same step, i.e., is contained in the same shell or subshell, the same letter is used in the form a′, a″, ... For examples 1, 2, and 4 in Table IV, the distance rank and complete generalized partitions are identical; therefore, the complete partition is only given in the last column.

For all partitions in Table IV, to facilitate orientation in the splitting of groups of vertexes, both for the classical radial partition and for the generalized radial partition, the groups of vertexes are separated by semicolons; subsequent splittings are separated by commas. [No distinction is made between commas and semicolons when calculating topological indexes.]

In all topological indexes defined according to relations 8 and 9 on the basis of all the partitions mentioned above, the order of numbers in the partition does not influence the value of the topological index. Yet the order which reflects the

Table V. Partitions and Topological Indexes Derived from the Distance Matrix for Some Graphs Selected out of 134 Cyclic (up to C_6) and Acyclic (up to C_8) Graphs (Denominations According to Table II)

no.	RCP RCI	DRCP DRCI	MCP MCI	RGP RGI	DRGP DRGI
1	2	3	4	5	6
37	4;2	2,2;2	2,4,6,6,6,6	4;2	2,2;2
	20	12	164	20	12
44	3;3	1,2;2,1	1,3,5,6,6,6	2;3;1	2;1,2;1
	18	10	143	14	10
51	3;3	1,2;2,1	1,3,5,6,6,6	1;4;1	1;2,2;1
	18	10	143	18	10
62	2;4	2;2,2	2,4,6,6,6,6	1;2;1;2	-
	20	12	164	10	-
66	6	2,4	2,4,6,6,6,6	2;2;2	-
	36	20	164	12	-
78	6	4,2	4,6,6,6,6,6	2;2;2	-
	36	20	196	12	-
90	6	2,2,2	2,4,6,6,6,6	2;2;2	-
	36	12	164	12	-
94	2;4	2;2,2	2,4,6,6,6,6	1;3;2	1;1,2;2
	20	12	164	14	10
97	6	1,2,3	1,3,6,6,6,6	3;3	1,2;3
	36	14	154	18	14
102	6	1,2,3	1,3,5,6,6,6	1;4;1	1;2,2;1
	36	14	143	18	10
105	6	2,4	2,4,6,6,6,6	2;2;2	-
	36	20	164	14	-
106	6	1,5	1,3,6,6,6,6	1;2;3	-
	36	26	154	14	-
107	6	-	6,6,6,6,6,6	6	-
	36	-	216	36	-
122	4;2;2	1,3;1,1;1,1	1,4,5,6,7,8,8,8	1;4;2;1	1;4;1,1;1
	24	14	319	22	20
124	3;3;2	1,2;2,1;1,1	1,3,4,5,6,7,8,8	3;3;2	1,2;1,1,1;2
	22	12	264	22	12

Additional partitions and indexes not identical with the preceding ones in Table V:

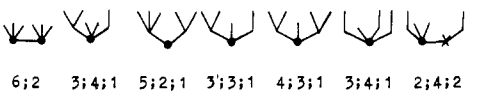
DCCP	no. 125: 1,2;1,1,1;1,1		MGP	no. 44: 2,3,5,6,6,6;	no. 62: 1,3,4,6,6,6
DCCI	10		MGI	146	134
CCP	no. 66: 2,2,2;	no. 102: 1,2,2,1; no. 105: 2,2,2; no. 106: 1,2,3		no. 78: 2,4,6,6,6,6;	no. 94: 1,2,4,6,6,6
CCI	12	10	14	164	129
CGP	no. 122: 1;3,1;1,1;1;	no. 124: 1,2;1,1,1;1,1			
CGI	14	10			

number of vertexes deleted progressively according to one convention or another, on removing the shells of the graphs from the outermost one toward the nucleus of the graph, may in some cases be an additional discriminating factor among isomeric graphs. We shall examine in the next sections how to take the order into account in calculating the topological indexes, first for acyclic graphs, then for cyclic ones.

3. How to Calculate Topological Indexes for Acyclic Graphs Taking into Account Also the Order of Numbers in the Partition. When the centric index B was derived for acyclic graphs according to relation 7, there was no need for taking into account the order of numbers in the pruning partition of vertexes (sequence S) because for trees it can be demonstrated that this sequence is always in nonincreasing order: in a tree, as one proceeds from the endpoints toward the center or bicenter by stepwise deleting the outermost δ_i endpoints, the numbers δ_i in this sequence S cannot increase.

Table III presents the nine isomers of heptane, C_7H_{16}, and the 18 isomers of octane, C_8H_{18}. Their partitions and derived topological indexes are given in Table V (the acyclic graphs with three to six vertexes are presented in Table I). It may be seen that the discriminating power of the pruning partition (and hence of the derived topological indexes according to relations (7) = (8), or (9) by analogy with the indexes discussed in the preceding sections) is low. The classical partitions, as well as the generalized partitions, on the other hand, are more discriminating, but these partitions have the order of numbers as an additional factor which can be used for increasing the differentiating ability of a topological index.

For acyclic graphs, yet another derivation of the classical radial partition of vertexes may be devised, which is similar in principle to the procedure described in the book by Busacker and Saaty[3] and originates with Cayley: the tree is drawn so that its center or bicenter is at the bottom, all vertexes at distance 1 from the center or bicenter lie on the next upper row, and so on, as in the example shown. By counting the

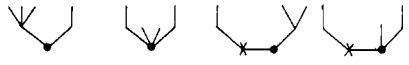


numbers of vertexes on the same horizontal row from top to bottom, one obtains the classical radial partition.

Unlike the lopping (pruning) partition, both the classical radial partition and the generalized radial partition possess additional information in the order of their numbers, because this order is not prescribed from the outset as for the pruning partition. This may be easily seen by examining Table V. All the above three partitions are different from one another. By including further refinements, i.e., on going to distance rank, distance code, or complete partitions, both the classical and the generalized partitions become more and more similar, but even when the numbers in the complete partitions are the same, their order may still be different for the classical and generalized radial partition.

How does one take into account also the order of numbers in the partitions when calculating topological indexes based on these partitions? There are two possible answers to this problem. One of them is to try and correlate the inversions

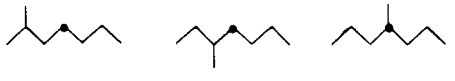
in the nonincreasing order (i.e., the number of times two adjacent digits in the partitions must be permuted so that the partition becomes arranged in nonincreasing order) with the degree of branching of the tree (corresponding to the alkane). By applying the concept of branching[15] to the present problem, it is clear that an inversion is correlated with increased branching (since inversions occur when shorter branches occur close to the center or bicenter of the graph, rather than at the periphery). Therefore, one could conventionally introduce an increment into the topological index for each inversion in the nonincreasing order, say the value 1. A few examples follow along with this proposal.



Classical radial partition	4;2;1	2;4;1	3;2;2	2;3;2
Increment	0	1	0	1
Index (8)	21	22	17	18

The second possibility of taking the order into account is to use the Muirhead reordering procedure.[18,19] Partial sums are formed from each partition, by replacing the second number in the partition by the sum of the first two numbers, then replacing the third number in the sequence by the sum of the first three numbers, and so on. Thus the differently ordered sequences of the same numbers which cannot be discriminated either by the quadratic (8) or the logarithmic functions (9) used in calculating topological indexes, become sequences of different numbers, in increasing order which lead to distinct topological indexes. Thus additional discriminating ability is introduced into the classical or generalized partitions, even beyond that supplied by the complete partitions. We shall call these modified partitions the *Muirhead classical partition*, and the *Muirhead generalized partition*, respectively.

For the three octane isomers presented below, their classical radial partitions are, respectively, (3; 2; 2; 1), (2; 3; 2; 1), and (2; 2; 3; 1). To illustrate the idea of Muirhead partitions starting from these sequences, we have first to make sure that the number of terms in the partition equals their sum, and this is done by adding as many zeroes as necessary (to provide a comparability between all graphs having a certain number of vertices, every structure is referred to the graph in which all N vertices are distinct; hence, the number of terms in the vertex partition should also be N, the missing terms being added as zeroes, e.g., four zeroes in the above example). Then the above three sequences are replaced by the partial sums as indicated above, leading to the Muirhead partitions: (3,5,7,8,8,8,8,8), (2,5,7,8,8,8,8,8), and (2,4,7,8,8,8,8,8).



Classical radial partition:	3;2;2;1	2;3;2;1	2;2;3;1
Increment:	0	1	2
Term partition:	3,2,2,1,0,0,0,0	2,3,2,1,0,0,0,0	2,2,3,1,0,0,0,0
Muirhead partition:	3,5,7,8,8,8,8,8	2,5,7,8,8,8,8,8	2,4,7,8,8,8,8,8
Muirhead index (8):	403	398	389

Usually, however, the Muirhead procedure is only applied to complete classical partitions, or complete generalized partitions, and not to a classical radial partition, as above. As can be seen from Table V for all C_7 and all C_8 isomers the complete Muirhead indexes are the same for both classical and generalized schemes. In 6 out of 27 cases, however, these indexes result from different shells and subshells.

4. Topological Indexes for Cyclic Graphs Calculated from Muirhead Classical and Generalized Partitions. Unlike acyclic graphs, where the pruning partition has a prescribed order of numbers, cyclic graphs, with their classical partition and generalized partition with all their variants presented in Table II, have the order of numbers in the partition as supplementary information. This order can be included in the calculation of the topological indexes by making use of Muirhead partitions, as indicated in the preceding paragraph.

Table V presents partitions and topological indexes for the cyclic and acyclic graphs contained in Tables I and III, denoting each graph by the same number as in these tables. Although more discriminating than quadratic indexes, the information centric indexes are for brevity omitted in Table V.

On going from the radial to the most detailed partition (radial → distance rank → distance code → complete → complete Muirhead) in both classical and generalized schemes partitions do not always differ. The rank partition is usually highly discriminating. In most of the examined cases, the following equalities hold (see Table II for the abreviations): CCP = DCCP = DRCP; CGP = DRGP; MGP = MCP. Because of this we list in Table V only the most distinct partitions and indexes. The others are presented below Table V for these structures only where they differ from the preceding ones. These are distance code partition (8 cases), complete classical partition (6 cases), complete generalized partition (8 cases), and Muirhead generalized partition (16 cases, as compared with Muirhead classical partition).

On increasing the discriminating power of the partition, it may be seen that the vertexes grouped together become more and more equivalent chemically or mathematically: for acyclic structures, the complete partitions provide the maximum possible equivalence of atoms which in graph-theoretical terms is the equivalence in the orbits of the automorphism group of the graph. For cyclic graphs, the same division into equivalence classes requires the introduction of a fifth criterion which will be discussed in a future paper.[8] The few cyclic graphs where this total breakdown into equivalence classes was not achieved in Table V are indicated in Tables I, III, and V by enclosing in brackets the notation of the graph.

III. CONCLUDING REMARKS

A more complete definition of graph center is proposed in the present paper, and called polycenter. The necessity of this generalization originated both from theoretical and practical reasons, the latter including not only the introduction of the new topological indexes described in the preceding sections, but also the "centroid" description of graph structure, as indicated for acyclic graphs by Read[20] or Lederberg et al.[21,22] Such a description is of great importance for a unique coding and computer processing of chemical information, for purposes of chemical nomenclature,[23,24] and for the enumeration of isomers.

The question arises if the present definition of graph center is sufficient. The present definition provides the maximum possible discrimination of vertexes on the basis of the topological distance matrix; therefore, it is sufficient within this limitation. A further generalization can be made on another basis, namely the atomic path code of Randić[4-6] or the requirement of equivalence of vertexes in the polycenter (the equivalence refers to "chemical equivalence" or to orbits of the automorphism group), as will be shown in a future paper.[8] For acyclic graphs, the Randić atomic path code is identical with the distance code used in our approach,[7] so that the present definition provides for acyclic graphs the maximum possible discrimination among vertexes; this conclusion leads to the idea of developing a fast algorithm for the determination

of graph orbits on the basis of the distance matrix. Efforts in this direction, as well as toward applications of the present and further generalizations of the notion of graph center in the fields of chemical nomenclature, computer processing of chemical structures, and synthetic strategy for organic structures are in progress.[25]

ACKNOWLEDGMENT

We thank Dr. M. Randić for his manuscripts prior to publication, and to Mr. A. Zamora for helpful suggestions in revising the paper.

REFERENCES AND NOTES

(1) F Harary, "Graph Theory", Addison-Wesley, Reading, Mass., 1969.
(2) D. Konig, "Theorie der endichen und unendlichen Graphen", Akademieverlagsgesellschaft, m.b.H., Leipzig, 1936, p 64.
(3) R. G. Busacker and T. L. Saaty, "Finite Graphs and Networks. An Introduction with Applications", McGraw-Hill, New York, 1965, p 200.
(4) M. Randić, *Chem. Phys. Lett.*, in press.
(5) M. Randić and C. L. Wilkins, *Chem. Phys. Lett.*, **55**, 547 (1978).
(6) M. Randić, G. M. Brissey, R. B. Spencer, and C. L. Wilkins, *Comput. Chem.*, in press.
(7) M. Randić and D. Bonchev, to be submitted for publication.
(8) D. Bonchev, M. Randić, and A. T. Balaban, *Int. J. Quantum Chem.*, in press.
(9) I. Gutman, M. Ruscić, N. Trinajstić, and C. F. Wilcox, Jr., *J. Chem. Phys.*, **62**, 3399 (1975).
(10) A. T. Balaban, *Theor. Chim. Acta* in press.
(11) C. Shannon and W. Weaver, "Mathematical Theory of Communications", University of Illinois, Urbana, Ill., 1949.
(12) L. Brillouin, "Science and Information Theory", Academic Press, New York, 1956.
(13) A. Katz, "Principles of Statistical Mechanics—The Information Theory Approach", Freeman, San Francisco, 1967.
(14) D. Bonchev, D. Kamenski, and V. Kamenska, *Bull. Math. Biol.*, **38**, 119 (1976).
(15) D. Bonchev and N. Trinajstić, *J. Chem. Phys.*, **67**, 4517 (1977).
(16) D. Bonchev and V. Kamenska, *Croat. Chem. Acta*, **51**, 19 (1978).
(17) D. Bonchev, "Proceedings of the Bremen Conference on Mathematical Structures in Chemistry, Bremen, 1978", *Math. Chem.*, **7**, 65 (1979).
(18) R. F. Muirhead, *Proc. Edinburgh Math. Soc.*, **19**, 36 (1901); **21**, 144 (1903); **24**, 45 (1906).
(19) G. H. Hardy, J. E. Littlewood, and G. Polya, "Inequalities", Cambridge University Press, London, 1934, p 44.
(20) R. C. Read and R. S. Miller, "A New System for the Designation of Chemical Compounds for the Purposes of Data Retrieval. I. Acyclic Compounds", Preprint, University of the West Indies, Jamaica, 1969.
(21) J. Lederberg, G. L. Sutherland, B. G. Buchanan, E. A. Feigenbaum, A. V. Robertson, A. M. Duffield, and C. Djerassi, *J. Am. Chem. Soc.*, **91**, 2973 (1969).
(22) L. Masinter, N. S. Sridharan, J. Lederberg, and D. H. Smith, *J. Am. Chem. Soc.*, **96**, 7702 (1974).
(23) A. T. Balaban and F. Harary, *Tetrahedron*, **24**, 2505 (1968); A. T. Balaban, *ibid.*, **25**, 2949 (1969); *Rev. Roum. Chim.*, **18**, 635 (1973); erratum *ibid.*, **19**, 338 (1974); *Math. Chem.*, **2**, 51 (1976).
(24) A. T. Balaban and P. v. R. Schleyer, *Tetrahedron*, **34**, 3599 (1978).
(25) A. T. Balaban and D. Bonchev, *Tetrahedron*, submitted for publication.

Automatic Generation of the Chemical Ringcode from a Connectivity Table

U. KLINGEBIEL

Chemie Information und Dokumentation, Berlin, Federal Republic of Germany

K. SPECHT*

Central Documentation Department, Schering AG, Berlin/Bergkamen, Federal Republic of Germany

Received December 10, 1979

A software package is described which generates automatically the General Chemical Ringcode from the connectivity table used by the IDC (Internationale Dokumentationsgesellschaft für Chemie mbH, Fachinformationszentrum Chemie). The conversion package, written for an IBM 370 computer, comprises, in addition to the main program written in Assembler, 35 subroutines of which 33 are written in Fortran, the remaining two in Assembler. The project was induced by the PDR (Pharma Dokumentationsring, e.V.), and supported by the BMFT (Bundesminister für Forschung und Technologie) under the scope of the program of the Federal Government of Germany for the promotion of Information and Documentation 1974 – 1977.

The target of the project was to reach compatibility between two different chemical documentation systems and between connectivity and fragmentary code records, and to create a powerful screen for atom-by-atom searches to optimize chemical structure retrieval.

The Ringdoc. The Pharma Dokumentationsring e.V. (PDR), a union of 18 European pharmaceutical companies and institutions, uses the Ringcode[1] for the encoding of organic chemical structures. The Ringcode, which was developed about 20 years ago by the founder members, is a fragmentary code. Code fragments are, for example, rings, chains, substitutents, functional groups, types of ring-condensation, positions of heteroatoms in a ring, positions of substituents, etc. Each fragment is associated with a defined position in 27 columns of a punched card. This means that all structural information can be coded in 27 × 12 = 324 positions or fragments of the General Chemical Code. Special codes were developed for the coding of steroids, inorganics, carbohydrates, and peptides. Usually one punched card per structure is sufficient. But in some cases additional cards are necessary, e.g., organic salts (separate coding of anion and cation) and adducts. Figure 1 shows the coding sheet for 3-(β-dimethylaminoethyl)indoline.

THE CONNECTIVITY TABLE (CT)

The CT used by the IDC (also known as BASF, main matrix[2]), which is the source for the Ringcode generation, consists of 200 lines and 26 columns. Figure 2 shows the table of the already-mentioned 3-(β-dimethylaminoethyl)indoline.

0095-2338/80/1620-0113$01.00/0