# Topological Organic Chemistry. 13. Transformation of Graph Adjacency Matrixes to Distance Matrixes

Harry P. Schultz†
Department of Chemistry, University of Miami, Coral Gables, Florida 33124

Received March 5, 2000

A sequence of simple arithmetic operations is described that generates distance matrixes from the adjacency matrixes of graphs.

## INTRODUCTION

The two most important matrixes that delineate the labeled chemical graph ($G$) are the adjacency (vertex, $\mathbf{A}(G)$) and the distance (edge count, $\mathbf{D}(G)$) matrixes, wherein $\mathbf{D}_{ij} = 1_{ij}$ if $i = j$, 0 otherwise; $1_{ij}$ is the shortest edge count between vertexes $i$ and $j$. **A** is the adjacency matrix where $\mathbf{A}_{ij} = 1$ if vertexes $i$ and $j$ are adjacent, 0 otherwise.

Mihalić et al.[1] succinctly outlined the history of the **D** matrix. They alluded to its derivation from the **A** matrix, stating that for large, complex graphs a computer must be used. In a related paper Mihalić et al.[2] observed that for small alkanes the **D** matrix is best generated using the powers of the corresponding **A** matrix. Roberts[3] and Brown and Brown[4] have described such operations. Müller et al.[5] were the source of a superior algorithm used for **D** matrix construction from the **A** matrix.

In this paper I describe a rapid, unsophisticated, arithmetic procedure for transforming alkane **A** matrixes into **D** matrixes.

## CALCULATIONS

The IUPAC system of numbering and nomenclature was used in the labeled, hydrogen-suppressed, stem-and-leaf graph of 2-methylbutane, with interatomic carbon−carbon edge counts set at unity, as pictured in Figure 1a. The essence of this method for creating **D** matrixes from **A** matrixes was to identify a preceding, adjacent vertex of known edge count and extend it by one edge count to the immediate succeeding unknown edge count vertex, continuing the operation until all edge count elements of the **D** matrix were in place. A specific example is outlined.

Figure 1b displays the **A** matrix of the 2-methylbutane, with 0's defining the main diagonal, but with traditional matrix element labels in place of the other 0 locations. The first unknown **D** matrix value is in the $a_{13}$ position, from which one drops vertically down to the 0 location on the main diagonal. Horizontal movement left to the first encountered unit digit ($a_{32}$) locates the immediately preceding vertex of known edge count value adjacent to the sought after unknown $a_{13}$ position. Vertical movement back up to the first horizontal vector identifies element $a_{12}$, a vertex of known edge count 1. The desired unknown $a_{13}$ location is adjacent to (one edge count) the known $a_{12}$ element (one edge count). The sum of the $a_{32}$ value (1) and the $a_{12}$ value (1) is 2, the desired value for position $a_{13}$. Because of the symmetry of the **A** and **D** matrixes about their main diagonals, the number 2 for the $a_{13}$ position is also placed in the $a_{31}$ site, as seen in Figure 1c.

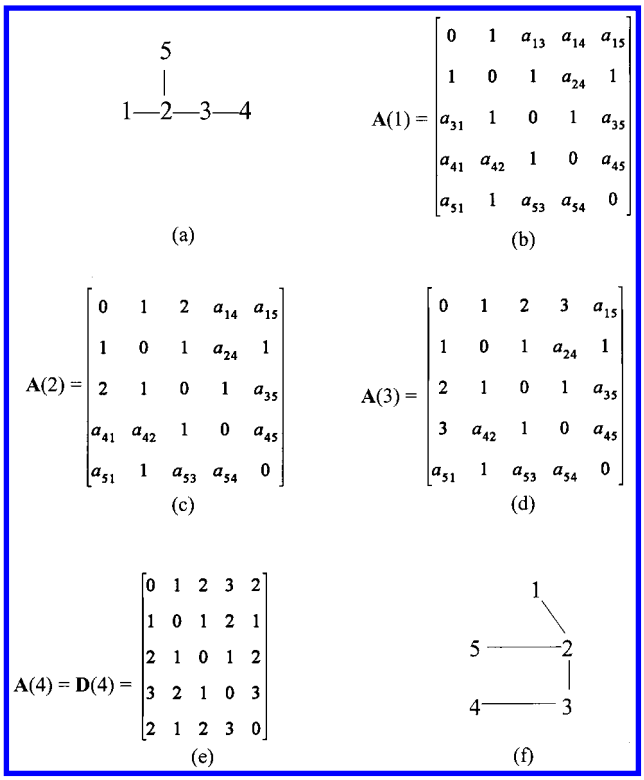


**Figure 1.** (a) 2-Methylbutane graph; (b) **A** matrix; (c) incomplete **D** matrix; (d) incomplete **D** matrix; (e) **D** matrix; (f) 2-methylbutane graph drawn from the **A** matrix.

The $a_{14}$ and $a_{41}$ locations are next determined in the same fashion: (1) $a_{14}$ vertically down to $a_{44}$; (2) $a_{44}$ horizontally to $a_{43}$ of known edge count 1; (3) $a_{43}$ vertically up to the element $a_{13}$ of known edge count 2; (4) the sum of the values of the $a_{43}$ and $a_{13}$ elements is entered into the $a_{14}$ and $a_{41}$ locations, as seen in Figure 1d. The last element, $a_{15}$, in the first horizontal vertex vector has its value determined in the same way; the same value, 2, is also inserted into position $a_{51}$.

In parallel fashion, focusing on the known edge count data in succeeding horizontal vectors, each of the remaining vertex vectors has the value of the missing element inserted, along with the corresponding symmetric element value. All of the

† Present address: P.O. Box 262, Big Horn, WY 82833. Phone: (307) 674-8868.

10.1021/ci0000225 CCC: $19.00 
Published on Web 07/06/2000

operations are simple enough to complete by mere inspection, arriving finally at the completed **D** matrix of 2-methylbutane as seen in Figure 1e.

The accuracy of the inspections can be ascertained by constructing the graph of 2-methylbutane from its **A** matrix data. This is rapidly done by listing all the various numbers from 1 to *N* in an approximate circle (to keep edge intersections of large and intricate graphs to a minimum) and then connecting the numbered vertexes, one edge at a time, by reference to the **A** matrix, as seen in Figure 1f. As the building process of the graph progresses, replicate edges are encountered. The isomorphism of the graphs pictured in Figure 1a,f is obvious.

The above exposition does not lend itself to constructing the **D** matrix of a cycloalkane from its **A** matrix, because of the requirement that the edge count between any two vertexes always be the shortest possible. This method applied to cyclic **A** matrixes resulted in continuing around the cycle in one direction only, passing the midpoint(s) of the cyclic graph with increasingly larger edge counts as one finally approached the initial vertex of the cycle. (However, construction of a cycloalkane graph from its **A** matrix, as described above, presented no problems.) This unique handicap to transformation of the **A** into the **D** matrix of a cycle was surmounted by inserting all the continually increasing edge count element values into the first horizontal vector, and then from the midpoint on subtracting those values from the total vertex count of the cyclic graph. The midpoint is moved one element further on in each succeeding horizontal vertex vector. This vexing cyclic problem does not extend to the alkyl substituents on a cycle; such alkyl substituents can be treated as summarized above for trees.

The sequential numbering of graph vertexes dictated by the IUPAC numbering system facilitated the determination of vertex values from left to right for generating **D** from **A** matrixes as described above. In the unusual and atypical circumstance wherein the graph vertexes are not successively numbered, an **A** matrix unknown edge count location may not be directly to the right of a vertex of known value. In this instance one can sometimes turn either left or right from the 0 point of the first vertical drop step of filling in the edge count values in the vacant spaces of the **A** matrix. Figure 2 replicates the information of Figure 1, using nonsequential vertex numbers for the graph of 2-methylbutane.

The Figure 2b **A**(1) matrix shows $a_{13}$ and $a_{15}$ with an adjacent unity value at $a_{14}$. Hence, $a_{15}$ vertically down to 0, left one column, and back up to known $a_{14}$ unity location gives a value of 2 for position $a_{15}$ (and $a_{51}$). Similarly, $a_{13}$ down to 0, right one column, and again back up to the known $a_{14}$ location yields the value of 2 for positions $a_{13}$ and $a_{31}$, as shown in Figure 2c. Site $a_{12}$ down to 0, right to the fifth column, and then up to the first row value of 2 adds up to values of 3 for $a_{12}$ and $a_{21}$ as pictured in Figure 2d. In row 2 position $a_{24}$ is solved first—down to 0, right one column to 1, and back up to 1 in row 2, giving $a_{24}$ and $a_{42}$ values of 2. Now $a_{23}$ can be determined—vertically down to 0, right one column to 1, and up to the just calculated value of 2 for location $a_{24}$, giving the value of 3 to the $a_{23}$ and $a_{32}$ spots. Position $a_{35}$, of value 2, is likewise defined to complete the **D** matrix of 2-methylbutane shown in Figure 2e. The stem-and-leaf graph of 2-methylbutane derived from the awkwardly numbered Figure 2b **A**(1) matrix is exhibited in Figure 1f.

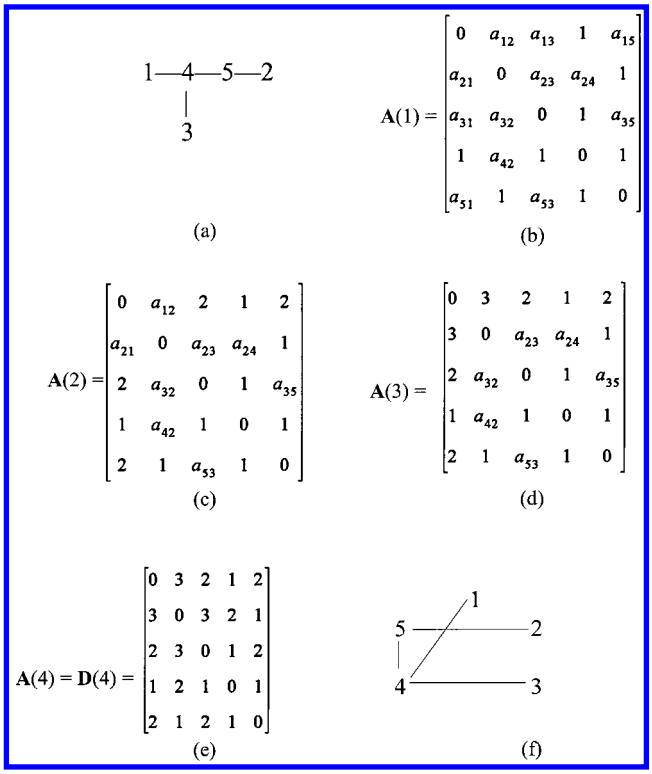


**Figure 2.** (a) 2-Methylbutane graph; (b) **A** matrix; (c) incomplete **D** matrix; (d) incomplete **D** matrix; (e) **D** matrix; (f) 2-methylbutane graph drawn from the **A** matrix.

## CONCLUSIONS

An efficient, relatively rapid procedure is described for transforming **A** into **D** matrixes, trees simply so, cycles somewhat less so. Mere inspection suffices to effect the arithmetic operations. The technique is meant for application in the occasional happenstance circumstance when an isolated need arises to create a **D** matrix from an **A** matrix, not in instances that involve very large or very numerous graphs that require computer usage as is referred to by Mihalić et al.,[1] and is so well illustrated by Müller et al.[5] The **A** matrix of 2-methylbutane required about 1 min for its transformation to the **D** matrix, that of 4,4,6-triethyl-5-isopropyl-2,3,5-trimethyloctane about 10 min.

## ACKNOWLEDGMENT

The reviewers are thanked for their constructive and very penetrating comments.

## REFERENCES AND NOTES

(1) Mihalić, Z.; Trinajstić, N. A Graph Theoretical Approach to Structure—Property Relationships. *J. Chem. Educ*. **1992**, *69*, 701—712.
(2) Mihalić, Z.; Nikolić, S.; Trinajstić, N. Comparative Study of Molecular Descriptors Derived from the Distance Matrix. *J. Chem. Inf. Comput. Sci*. **1992**, *32*, 28—37.
(3) Roberts, F. S. *Discrete Mathematical Model*; Prentice Hall: Englewood Cliffs, NJ, 1976; p 58.
(4) Brown, R. F.; Brown, B. W. *Finite Mathematics*; Ardsley: New York, 1992; pp 534—545.
(5) Müller, W. R.; Szymansky, K.; Knop, J. V.; Trinajstić, N. An Algorithm for Construction of the Molecular Distance Matrix. *J. Comput. Chem.* **1987**, *8*, 170—173.

CI0000225