Morgan Revisited

John Figueras

781 Baker's Pond Road, Orleans, Massachusetts 02653

Received April 19, 1993*

Graph theoretic analysis demonstrates that the Morgan algorithm is equivalent to computing a power of the adjacency matrix.

INTRODUCTION

Ever since its description in 1965, the Morgan algorithm[1] has been important in computer applications that handle chemical structure. It was originally proposed as a key step in the unique coding of chemical structures and was subsequently extended in slightly modified form by Wipke and Dyott[2] as part of their chemical synthesis program, SECS. Corneil and Read[3] mention it as a possible but incomplete method for partitioning vertices for isomorphism determinations. Shelley and Munk[4] and Figueras[5] used it in conjunction with other graph invariants for topological symmetry perception, and the algorithm played an important role in Shelley's structure display program.[6]

In place of paraphrase, let me quote Morgan's original description of his algorithm:[1]

"The 'connectivity values' are computed by first assigning to each node an initial 'connectivity value' equal to the number of nonhydrogen atoms attached to that node.... The computer then calculates the number (k) of different 'connectivity values' which had been assigned. An iterative process is then established which calculates a new 'connectivity value' for each node. This new value is the sum of the assigned values for the nodes connected to the one under consideration. Having computed a value for each node based on the previous values, the computer calculates the number (k') of different values in the set of new values. If $k' > k$, the new values are assigned to the corresponding nodes, k is set equal to k', and the summation process is repeated. If, however, $k' \leq k$, the process is terminated and the last set of values assigned to the nodes is used to induce a partial ordering among the nodes...." It is apparent that, for each atom in the structure, each cycle of this computation brings in the next shell of atoms for computing the new connectivity values.

A relationship between Morgan's connectivity values and matrix powers of the adjacency matrix was noted by Razinger,[7] who gave an empirical example but offered no proof. More recently, Rücker and Rücker[8] derived the relationship by proving that total walk counts, which can be computed as row sums of a power of the adjacency matrix, are identical to Morgan's connectivity values. In this paper, the relationship is derived by direct representation of the Morgan algorithm as a recursive matrix multiplication.

THEORETICAL ANALYSIS

For a structure composed of n atoms, let $\mathbf{A}$ be the $n \times n$ adjacency matrix representing the connectivity in the structure such that $a_{ij} = 1$, if atoms i and j are connected, else $a_{ij} = 0$. Diagonal elements are taken as zero. Obviously, $a_{ij} = a_{ji}$. For the kth iteration of the algorithm, let $\mathbf{V}_k$ be a column vector that contains the current connectivity values $\{v_{1k}v_{2k}...v_{nk}\}$, where v_{ik} is the connectivity value assigned to the ith atom in the kth iteration. According to the algorithm stated above, the next connectivity value for atom i is computed by summing all of the current values associated with the atoms attached to i; this is the scalar product of row i of the adjacency matrix and the column vector $\mathbf{V}_k$:

$$v_{i,k+1} = [a_{i1}a_{i2}a_{i3}...a_{in}]\cdot\{v_{1k}v_{2k}...v_{nk}\} = \sum a_{ij}v_{jk} \quad (1)$$

Generalizing to the whole adjacency matrix, we can write directly from (1) the matrix product

$$\mathbf{V}_{k+1} = \mathbf{A}\mathbf{V}_k \quad (2)$$

From the recursive equation (2) it is a direct step to the final equation (3), where $\mathbf{V}_0$ is the vector of initial connectivity values assigned to the n atoms and $\mathbf{A}^r$ is the rth power of the adjacency matrix.

$$\mathbf{V}_r = \mathbf{A}^r\mathbf{V}_0 \quad (3)$$

There are several interesting conclusions that can be drawn from eq 3:

(a) The Morgan algorithm is separable into two distinct factors, one that depends only upon the adjacency matrix and the other upon the initial connectivity values assigned to the atoms. Thus, modifications such as the introduction of atom types into computation of initial connectivity values improve partitioning on their own, and do not alter the basic behavior of the algorithm.

(b) The i,jth element of the rth power of the adjacency matrix is, from graph theory,[10] the number of walks of length r that exist between atoms i and j. The connectivity value V_{ir} associated with atom i after the rth iteration is the sum of all walks of length r starting from atom i and ending at atoms $j = 1...n$, each weighted by the initial connectivity value V_{j0}. The contribution of all atoms k at a distance r bonds away from atom i to the computation of the connectivity value at i will not be felt until r iterations have been carried out.

(c) The initial vector $\mathbf{V}_0$ generally used in the Morgan algorithm is the vector of degrees (the number of non-hydrogen attachments) at each atom. However, a vector made up of unit components serves as well, since the first multiplication of $\mathbf{V}$ by $\mathbf{A}$ automatically installs node degrees in the vector. The initial vector may of course be modified, as mentioned earlier, to take into account other node properties such as atomic weight or ring membership, that will help partitioning.

The terminating condition for the Morgan algorithm requires repetition of the summation process until k, the number of different connectivity values, no longer increases. The upper limit on the number of iterations is n, since, as has

* Abstract published in *Advance ACS Abstracts,* August 15, 1993.

0095-2338/93/1633-0717$04.00/0

been pointed out,[8] the Cayley–Hamilton theorem implies that powers of an adjacency matrix higher than n generate linear combinations of lower matrix powers. In some cases, fewer than n iterations may be required to reach the terminating condition because some of the row sums become degenerate. Also, in certain cases, oscillatory behavior is observed; that is, the number k of different connectivity values does not converge to a single value, but alternately decreases and increases as the process is repeated, and in these cases, fewer than n iterations will be required to reach the terminating condition. Strictly speaking, however, the Morgan algorithm *always* terminates, because, according to the original recipe, as soon as k no longer increases, the process is stopped and the partition induced by the previous round is accepted.

Rücker and Rücker[10] have developed an algorithm that successfully uses powers of the adjacency matrix itself to separate atoms into equivalence classes for many cases in which the Morgan algorithm breaks down. In their work, they find that by classifying atoms using the actual matrix elements generated for each power of the adjacency matrix, they can obtain partitioning that is not uncovered by the Morgan algorithm. The Morgan algorithm hides differences among terms of the adjacency matrix power by effectively summing its row elements, resulting in degeneracy and throwing away information that the Rücker algorithm successfully uses for partitioning.

ACKNOWLEDGMENT

I wish to thank the reviewer, Ch. Rücker, who called to my attention the previous work by Razinger[7] and who kindly made available to me his own work prior to publication.[8]

REFERENCES AND NOTES

(1) Morgan, H. L. The Generation of a Unique Machine Description for Chemical Structures—A Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* **1965**, *5*, 107–113.
(2) Wipke, W. T.; Dyott, T. M. Stereochemically Unique Naming Algorithm. *J. Am. Chem. Soc.* **1974**, *96*, 4834–4842.
(3) Read, R. C.; Corneil, D. G. The Graph Isomorphism Disease. *J. Graph Theory* **1977**, *1*, 339–363.
(4) Shelley, C. A.; Munk, M. E. An Approach to the Assignment of Canonical Connection Tables and Topological Symmetry Perception. *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 247–250.
(5) Figueras, J. Automorphism and Equivalence Classes. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 153–157.
(6) Shelley, C. A. Heuristic Approach for Displaying Chemical Structure. *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 61–65.
(7) Razinger, M. Extended Connectivity in Chemical Graphs. *Theor. Chim. Acta* **1982**, *61*, 581–586.
(8) Rücker, G.; Rücker, C. Counts of All Walks as Atomic and Molecular Descriptors. *J. Chem. Inf. Comput. Sci.*, see paper on pp 683–695 of this issue.
(9) Gould, R. *Graph Theory*; The Benjamin/Cummings Publishing Co., Inc.: 1988.
(10) Rücker, G.; Rücker, C. On Using the Adjacency Matrix Power Method for Perception of Symmetry and for Isomorphism Testing of Highly Intricate Graphs. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 123–126.