THEORETICA CHIMICA ACTA

Theoret. Chim. Acta (Berl.) 61, 581–586 (1982)

Extended Connectivity in Chemical Graphs

M. Razinger

Boris Kidrič Institute of Chemistry, POB. 380, YU-61001 Ljubljana, Yugoslavia

The true nature of the extended connectivity, used in Morgan algorithm for the canonical numerotation of points in chemical graphs, is discussed. An alternative method for its calculation based on the number of walks is described and shown to yield results identical to Morgan's method.

Key words: Graph theory – Extended connectivity – Morgan algorithm.

1. Introduction

The notion of extended connectivity (EC) was first introduced by Morgan [1] in his search for a unique canonical name of a given chemical structure. While the connectivity (i.e. the number of nonhydrogen neighbours of points in a graph) itself divides the points of a graph in at most four classes, the EC makes possible a finer differentiation, dividing the points into more than four classes. This statement implies that the graph in question represents the skeleton of a molecular structure, pruned of its hydrogen atoms and containing only atoms whose valence is less or equal to 4.

The Morgan algorithm has been extensively used in the generation of canonical connection tables which form the backbone of the Chemical Abstract Service structural data bank, the largest existing bank of this type. Morgan's method of numerotation of points in graphs was further modified and developed by several authors [2, 3].

In this paper we intend to show that the introduction and use of a new graph invariant, the EC, by Morgan and other cited authors was not necessary. Another topological characteristic, the number of walks in a graph, known in theory before the EC [4], assigns to the points of graphs identical numerical values as does the EC itself.

0040-5744/82/0061/0581/$01.20

2. Calculation of Extended Connectivity

2.1. Morgan's and Related Methods

Starting with the set of connectivities of all points in a graph, Morgan's method [1] consists of a pragmatic iterative summation of connectivities of all neighbours of each point in turn resulting in a new set of connectivities, called EC. The procedure of calculation for an octane isomer is shown in Fig. 1.

The parameter K denotes the number of classes into which the points are divided according to their EC values. The parameter determines where the iterative procedure has to be stopped. We shall not discuss further the principles of numerotation as they are given in detail in the original article [1] and are in fact irrelevant to our findings.

Wipke and Dyott [2] have modified Morgan algorithm to give a stereochemically selective and unique name of a given molecular structure. Their algorithm for the numerotation differs of the Morgan's in the treatment of end-points which conserve their initial connectivity value of 1 throughout the iterative procedure, but nevertheless, the principle of calculation of EC values remains the same as in Morgan's method.

Another interpretation of the Morgan algorithm and its relations to graph potentials were published recently [5].

2.2. Alternative Method

Studying various topological descriptors of chemical structures and the possibilities of their automatic computer generation [6], starting from the adjacency matrix $\boldsymbol{A}$ of corresponding graphs, we focused our attention to the higher powers of $\boldsymbol{A}$, i.e. $\boldsymbol{A}^2, \boldsymbol{A}^3, \ldots, \boldsymbol{A}^N$ (N being the number of points in the graph). Besides many graph-theoretical properties, known to be deducible from these powers $\boldsymbol{A}^i$ [4, 7], we found that the EC values of points, calculated in the ith iteration of Morgan's procedure, can be deduced directly from the ith power of $\boldsymbol{A}$. Powers of the adjacency matrix of the same graph as that in Fig. 1 are shown on the computer printout in Fig. 2.

The EC values of points are calculated as straightforward sums of the corresponding rows (or columns, as $\boldsymbol{A}$ and its powers are symmetrical) of the matrices. These sums are shown in the column on the right in Fig. 2 alongside the matrices from which they are calculated. Although the numerotation of points does not influence the results, an arbitrary numerotation, shown above the computer listing, was chosen for the sake of clarity of presentation.

3. Conclusion

It is known [4, 7] that the value of element $a(i, j)$ of the kth power of adjacency matrix $\boldsymbol{A}$ is equal to the number of all possible walks of length k, starting from point i and ending in point j. The sum of these elements $a(i, j)$ over all values

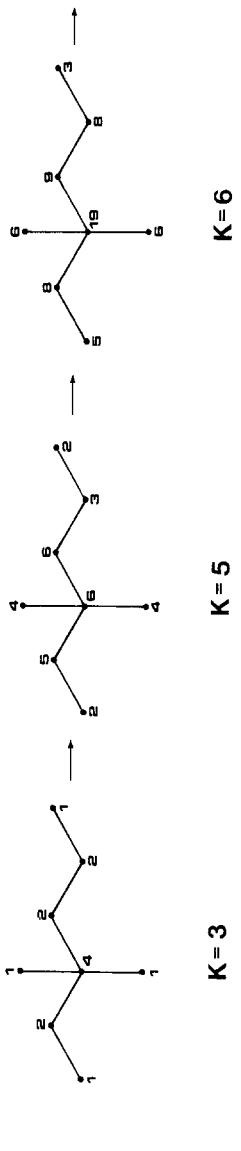
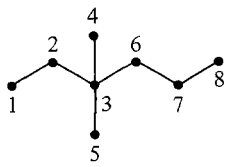


Fig. 1. Morgan's calculation of the EC values



Matrix of power 1

i/j	1	2	3	4	5	6	7	8	Sum a(i, j) (j = 1, . . 8)
1	0	1	0	0	0	0	0	0	1
2	1	0	1	0	0	0	0	0	2
3	0	1	0	1	1	1	0	0	4
4	0	0	1	0	0	0	0	0	1
5	0	0	1	0	0	0	0	0	1
6	0	0	1	0	0	0	1	0	2
7	0	0	0	0	0	1	0	1	2
8	0	0	0	0	0	0	1	0	1

Matrix of power 2

i/j	1	2	3	4	5	6	7	8	Sum a(i, j) (j = 1, . . 8)
1	1	0	1	0	0	0	0	0	2
2	0	2	0	1	1	1	0	0	5
3	1	0	4	0	0	0	1	0	6
4	0	1	0	1	1	1	0	0	4
5	0	1	0	1	1	1	0	0	4
6	0	1	0	1	1	2	0	1	6
7	0	0	1	0	0	0	2	0	3
8	0	0	0	0	0	1	0	1	2

Matrix of power 3

i/j	1	2	3	4	5	6	7	8	Sum a(i, j) (j = 1, . . 8)
1	0	2	0	1	1	1	0	0	5
2	2	0	5	0	0	0	1	0	8
3	0	5	0	4	4	5	0	1	19
4	1	0	4	0	0	0	1	0	6
5	1	0	4	0	0	0	1	0	6
6	1	0	5	0	0	0	3	0	9
7	0	1	0	1	1	3	0	2	8
8	0	0	1	0	0	0	2	0	3

Fig. 2. Listing of the powers of adjacency matrix with the calculated number of walks

Matrix of power 4

i/j	1	2	3	4	5	6	7	8	Sum a(i, j) (j = 1, . . 8)
1	2	0	5	0	0	0	1	0	8
2	0	7	0	5	5	6	0	1	24
3	5	0	18	0	0	0	6	0	29
4	0	5	0	4	4	5	0	1	19
5	0	5	0	4	4	5	0	1	19
6	0	6	0	5	5	8	0	3	27
7	1	0	6	0	0	0	5	0	12
8	0	1	0	1	1	3	0	2	8

Matrix of power 5

i/j	1	2	3	4	5	6	7	8	Sum a(i, j) (j = 1, . . 8)
1	0	7	0	5	5	6	0	1	24
2	7	0	23	0	0	0	7	0	37
3	0	23	0	18	18	24	0	6	89
4	5	0	18	0	0	0	6	0	29
5	5	0	18	0	0	0	6	0	29
6	6	0	24	0	0	0	11	0	41
7	0	7	0	6	6	11	0	5	35
8	1	0	6	0	0	0	5	0	12

Classification of points

	1 K = 3	2 K = 5	3 K = 6	4 K = 6	5 K = 7
1	1	2	5	8	24
2	2	5	8	24	37
3	4	6	19	29	89
4	1	4	6	19	29
5	1	4	6	19	29
6	2	6	9	27	41
7	2	3	8	12	35
8	1	2	3	8	12

Fig. 2 (*cont.*)

of j (i.e. the sum of the ith row) gives the number of all possible walks of length k from point i to all other points in a (connected) graph. This is true for cyclic graphs as well as for the trees one of which was used for the demonstration of our alternative of Morgan algorithm. As the EC values of points, calculated by the now classical algorithm, are exactly the same as those obtained by the enumeration of walks of corresponding length, the conclusion is evident: the introduction of a new graph invariant, i.e. the EC, was in fact superfluous since

a previously defined graph invariant can be used for the identical canonical numerotation of points in graphs.

Acknowledgement. Part of this work was done on the Institut de Topologie et de Dynamique des Systèmes, Université Paris VII. I thank Professors J. E. Dubois and J. Chrétien of ITODYS for stimulating discussions and also Professors D. Hadži and J. Zupan of the Boris Kidrič Institute of Chemistry, Ljubljana, for their interest and support.

References

1. Morgan, H. L.: J. Chem. Doc. **5**, 107 (1965)
2. Wipke, W. T., Dyott, T. M.: J. Am. Chem. Soc. **96**, 4834 (1974)
3. Moreau, G.: Nouv. J. Chim. **4**, 17 (1980)
4. Harary, F.: Graph Theory, Reading, Mass: Addison-Wesley, 1969
5. Golender, V. E., Drboglav, V. V., Rosenblit, A. B.: J. Chem. Inf. Comput. Sci. **21**, 196 (1981)
6. Razinger, M., Chrétien, J., Dubois, J. E.: J. Chim. Phys. (submitted)
7. Deo, N.: Graph theory with applications to engineering and computer science, Englewood Cliffs, N.J.: Prentice-Hall, 1974

Received March 24, 1982