match no. 12 pp. 3-14 1981

NUMBERING OF FINITE RELATIONAL SYSTEMS

Karl Wirth[1] and Martin K.Huber[2]

Organisch-chemisches Institut der
Universität Zürich

(Received: April 1981)

Current addresses:
1) Fehrenstr.23, CH-8032 Zürich, Switzerland
2) Zentrale Forschungslaboratorien, CIBA-GEIGY AG, CH-4002 Basel, Switzerland

A numbering method for finite relational systems is to be described. This method assigns a particular numbering class, the minimum class, to any given relational system. To this end, ordered numeral systems are defined which themselves can be ordered to yield a code for a relational system. In chemistry our numbering method may be used to number uniquely the atoms of a molecule and, thus, to generate a systematic name for it.

1. INTRODUCTION

The work we present in this paper developed[1)] from the well-known problem in chemistry of how to number and, ultimately, name a molecule. The implications of a solution to this problem for a wide variety of computer applications in chemistry have long been recognized. As early as 1965, the significance of numbering and naming became evident when H.L. Morgan published his famous numbering algorithm for molecular constitutions [1] which still is the corner-stone of many a computerized documentation system. Since then, and particularly with the growing acceptance of computers in chemical research, the number of papers concerned with the questions of numbering, naming, and - closely related - symmetry recognition has grown at an ever increasing pace [2]-[13].

We were interested in a solution to these problems which would allow us to treat all structural aspects of a molecule, known or yet to be discovered, in a conceptually uniform, canonical manner. In this context, we wish to report on a numbering method which presupposes that the structure of a molecule can be represented as a finite relational system. Thus we are concerned with a purely mathematical problem of admittedly elementary character.

The method of describing molecules as relational systems will be dealt with in the first [14] of a series of three forthcoming papers. An algorithm for the numbering method based on the present work will be given in the second [15] of these three papers. Once these problems have been treated, the atoms of a molecule can be numbered and the molecule itself named; examples will be presented in the last [16] of the above mentioned papers.

---

1) The basic ideas were conceived in connection with stereo-chemical investigations stimulated by A.S. Dreiding. These investigations were conducted as part of a project under his direction supported by the 'Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung'. Portions of this material were presented at the 172nd National Meeting of the American Chemical Society, San Francisco, Calif., Aug. 29, 1976. We would like to thank A. Häussler and F. Siegerist for helpful discussions.

In mathematical literature one encounters numbering methods mainly in connection with procedures for detecting isomorphism between finite graphs. A comprehensive annotated bibliography relating to the graph isomorphism problem is given in [17][18] .

2. DEFINITIONS

Definition. Let X be a finite set and $R_1, R_2, \ldots, R_m$ relations respectively of $k_1, k_2, \ldots, k_m$ arguments in X, i.e. $R_i \subset X^{k_i}$ for $1 \leq i \leq m$. The ordered (m+1)-tuple

$$S = (X, R_1, R_2, \ldots, R_m)$$

is a finite relational system with the field X. By relational systems, in this paper, we always mean finite ones.

As an example of a relational system let us consider $Q = (X,R)$ with $X = \{a,b,c,d\}$ and the binary relation $R = \{(a,b),(b,a),(b,c),(c,b),(c,d),(d,c),(d,a),(a,d)\}$. The symmetric graph shown in figure 1 visualizes Q.

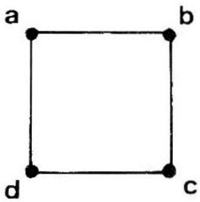

Figure 1

Definition. Two relational systems $S = (X, R_1, R_2, \ldots, R_m)$ and $S' = (Y, T_1, T_2, \ldots, T_{m'})$ are isomorphic if $m = m'$ and if there is a one-to-one function f of X onto Y which preserves the relations, i.e. $(x_1, x_2, \ldots, x_{k_i}) \in R_i \longleftrightarrow (f(x_1), f(x_2), \ldots, f(x_{k_i})) \in T_i$ for $1 \leq i \leq m$. f is an isomorphism of S to S' and we write $f(S) = S'$. An isomorphism of S to S is an automorphism of S.

It is known that the set aut(S) of all automorphisms of S together with the composition of automorphisms forms a permutation group which operates on X; it is the automorphism group of S. In our example:

$$\mathrm{aut}(Q) = \left\{ \begin{pmatrix}abcd\\abcd\end{pmatrix}, \begin{pmatrix}abcd\\bcda\end{pmatrix}, \begin{pmatrix}abcd\\cdab\end{pmatrix}, \begin{pmatrix}abcd\\dabc\end{pmatrix}, \begin{pmatrix}abcd\\adcb\end{pmatrix}, \begin{pmatrix}abcd\\cbad\end{pmatrix}, \begin{pmatrix}abcd\\badc\end{pmatrix}, \begin{pmatrix}abcd\\dcba\end{pmatrix} \right\} .$$

Definition. Let n be the number of elements of the field X of the relational system S. A one-to-one function ν of X into $\{1,2,\ldots,n\}$ is a numbering function, short, a numbering of S.

In figure 2 two numberings $\nu = \begin{pmatrix}abcd\\1234\end{pmatrix}$ and $\nu' = \begin{pmatrix}abcd\\4123\end{pmatrix}$ of Q are given. Apparently ν results from ν' by a "90^0-rotation", i.e. by the automorphism $\alpha = \begin{pmatrix}abcd\\bcda\end{pmatrix}$. More precisely: $\nu = \nu'\alpha \leftrightarrow \nu'^{-1}\nu = \alpha$.

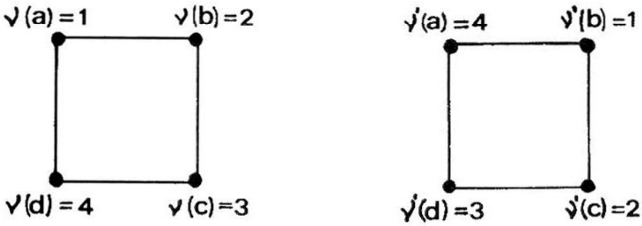

Figure 2

Definition. Two numberings ν and ν' of the relational system S are equivalent if $\nu'^{-1}\nu$ is an automorphism of S.

The set of all n! numberings of S, therefore, is partitioned into classes of equivalent numberings; this follows from the group structure of aut(S). The equivalence classes are called numbering classes of S. If ν is a numbering of S, then $[\nu] = \{\nu\alpha \mid \alpha \in \mathrm{aut}(S)\}$ is the class which contains ν. Since $\nu\alpha \neq \nu\alpha'$ for $\alpha, \alpha' \in \mathrm{aut}(S)$ with $\alpha \neq \alpha'$ each numbering class of S contains exactly g elements where g is the order of aut(S). Hence the number of numbering classes of S equals $\frac{n!}{g}$. In the example of the relational system Q the automorphism group is of the order 8. Of the $\frac{4!}{8} = 3$ numbering classes of Q a representative of each, namely ν, μ and π, is given in figure 3.

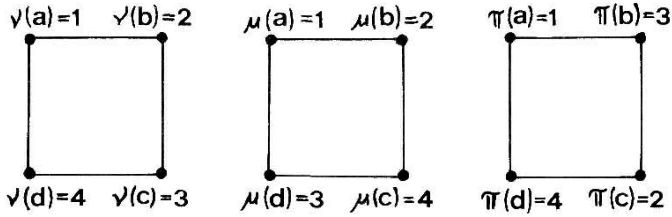

Figure 3

Definition. A numbering method assigns to each relational system S a numbering class stand(S), called standard class of S, where for isomorphic relational systems S and S', with $\nu \in$ stand(S) and $\nu' \in$ stand(S'), $\nu'^{-1}\nu$ is an isomorphism of S to S'.

Note that the various numbering methods mentioned in literature are restricted to graphs [19] [20] [21] [22] [23] [24] or certain extended structures [1] [2] [6] [7] [8] [11] [12] [13] ; both of these might in fact be regarded as special relational systems.

3. MINIMUM METHOD

Using lexicographical order we define a total order on the set of the numbering classes of a relational system S. The smallest class with reference to this order is specified as the standard class of S by what we call the minimum method.

For each numbering ν of S there is a relational system ν(S), isomorphic to S, which we call a numeral system of S. Let us consider, for example, the relational system D = ({a,b,c},{(a,b),(b,c),(c,a)}), which is depicted by the Digraph in figure 4.

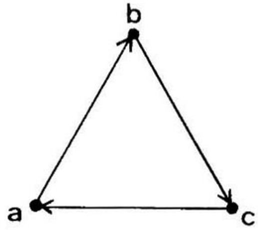

Figure 4

The 3! numeral systems of D are as follows:

$$\binom{abc}{123}(D) = (\{1,2,3\},\{(1,2),(2,3),(3,1)\})$$

$$\binom{abc}{231}(D) = (\{2,3,1\},\{(2,3),(3,1),(1,2)\})$$

$$\binom{abc}{312}(D) = (\{3,1,2\},\{(3,1),(1,2),(2,3)\})$$

$$\binom{abc}{132}(D) = (\{1,3,2\},\{(1,3),(3,2),(2,1)\})$$

$$\binom{abc}{321}(D) = (\{3,2,1\},\{(3,2),(2,1),(1,3)\})$$

$$\binom{abc}{213}(D) = (\{2,1,3\},\{(2,1),(1,3),(3,2)\})$$

The first three numeral systems of D are identical and so are the last three. For a relational system S it is generally true that: $\nu(S) = \nu'(S) \longleftrightarrow \nu'^{-1}\nu(S) = S \longleftrightarrow \nu'^{-1}\nu \in \mathrm{aut}(S)$. Thus two numeral systems $\nu(S)$ and $\nu'(S)$ are equal if and only if ν and ν' are equivalent numberings of S.

Based on the order of the natural numbers, in each numeral system $\nu(S)$, the elements of the field as well as the elements of each individual relation are now lexicographically ordered to tuples. A uniquely determined string of symbols $\overline{\nu(S)}$, which is called an <u>ordered numeral system</u> of S, is thereby obtained. As above, it is still understood that two ordered numeral systems $\overline{\nu(S)}$ and $\overline{\nu'(S)}$ are equal if and only if ν and ν' are equivalent numberings of S.

A one-to-one function Ω is now defined which assigns the ordered numeral system $\overline{\nu(S)}$ to each numbering class $[\nu]$ of S. In the example of the relational system D we can denote Ω as follows:

$$\left[\begin{pmatrix}abc\\123\end{pmatrix}\right] \longmapsto ((1,2,3),((1,2),(2,3),(3,1)))$$
$$\left[\begin{pmatrix}abc\\132\end{pmatrix}\right] \longmapsto ((1,2,3),((1,3),(2,1),(3,2)))$$

The definition of a total order on the set of the ordered numeral systems of S is clear now: It may be found through lexicographical ordering. The smallest element regarding this order is called code of S, written code(S). With Ω^{-1} this order can be transferred in a natural way to the set of the numbering classes of S. The smallest class Ω^{-1}(code(S)) is now the standard class of S by the minimum method, the so-called minimum class of S, termed min(S). For our examples D and Q we find:

$$\text{code}(D) = ((1,2,3),((1,2),(2,3),(3,1)))$$
$$\min(D) = \left[\begin{pmatrix}abc\\123\end{pmatrix}\right] = \left\{\begin{pmatrix}abc\\123\end{pmatrix},\begin{pmatrix}abc\\231\end{pmatrix},\begin{pmatrix}abc\\312\end{pmatrix}\right\}$$

and

$$\text{code}(Q) = ((1,2,3,4),((1,2),(1,3),(2,1),(2,4),(3,1),(3,4),(4,2),(4,3)))$$
$$\min(Q) = \left[\begin{pmatrix}abcd\\1243\end{pmatrix}\right] = \left\{\begin{pmatrix}abcd\\1243\end{pmatrix},\begin{pmatrix}abcd\\2431\end{pmatrix},\begin{pmatrix}abcd\\4312\end{pmatrix},\begin{pmatrix}abcd\\3124\end{pmatrix},\begin{pmatrix}abcd\\1342\end{pmatrix},\begin{pmatrix}abcd\\4213\end{pmatrix},\begin{pmatrix}abcd\\2134\end{pmatrix},\begin{pmatrix}abcd\\3421\end{pmatrix}\right\}$$

The minimum method is a numbering method in the sense defined above. For if S and S' are isomorphic relational systems, we first show that code(S) = code(S'). f being an isomorphism of S to S' and $\nu' \in \min(S')$, $\nu' f$ is a numbering of S. Since code(S') = $\overline{\nu'(S')} = \overline{\nu' f(S)}$, code(S') is lexicographically not smaller than code(S). On the other hand, code(S) is lexicographically not smaller than code(S'), so that code(S) = code(S'). It now follows from $\overline{\nu(S)} = \overline{\nu'(S')} \longleftrightarrow \nu(S) = \nu'(S') \longleftrightarrow \nu'^{-1}\nu(S) = S'$ that, with $\nu \in \min(S)$ and $\nu' \in \min(S')$, $\nu'^{-1}\nu$ is an isomorphism of S to S'.

To illustrate the minimum method, we give two more examples. First a relational system with several relations is to be considered, namely W with the field $X = \{a,b,c,d,e\}$ and the relations $R_1 = \{(b,c),(c,c)\}$, $R_2 = \{(b,a),(b,d),(b,e)\}$ and $R_3 = \{(a,c,d),(d,c,e),(e,c,a),(b,c,b),(b,b,b)\}$. The minimum class of W can be found manually:

$$\min(W) = \left\{ \begin{pmatrix} abcde \\ 32145 \end{pmatrix}, \begin{pmatrix} abcde \\ 42153 \end{pmatrix}, \begin{pmatrix} abcde \\ 52134 \end{pmatrix} \right\} .$$

The second example is much harder to handle unless one resorts to a computer. Consider the relational system C with a field of 30 elements and with the binary, symmetric relation as depicted the cat-like figure 5. The minimum class of C contains two numberings which differ only in the numbers assigned to the 'eyes' 22 and 23. As a matter of convenience the points have been labelled directly with the corresponding numbers.

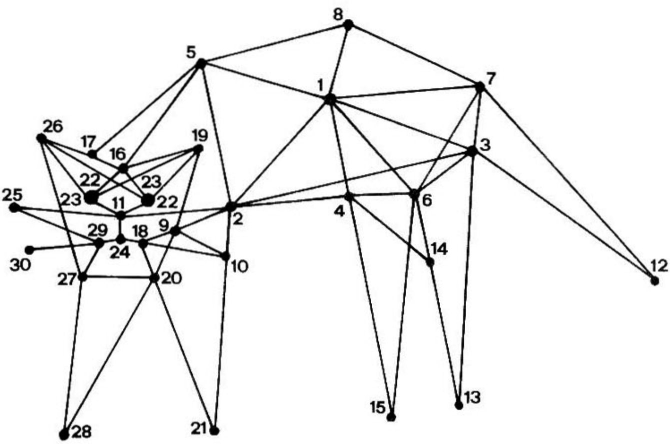

Figure 5

4. REMARKS

The concept of numbering, which originates in chemistry, could be replaced by the concept of total order; we chose numberings for simplicity of presentation.

The terms "numbering of S" and "equivalence of numberings of S" have been explained only with reference to the field X and to aut(S) respectively. However, they are definable for any finite set X and, respectively, any permutation group G which operates on X. One may ask whether a total order on the set of the numbering classes can be defined by using only the structure of G. This can be proved to be the case if and only if $G = N_G$, where N_G is the normalizer of G in the symmetric group of X.

The minimum method assigns to each relational system S the standard class min(S). For certain purposes numbering methods generating different standard classes might be desirable. These may serve to bring into prominence certain structural aspects of special relational systems. Thus a particular numbering method, for instance, leads to remarkable rankings of tournaments (oriented, complete graphs). Some information on the ranking problem can be found in [25] [26].

The set of automorphisms can be determined from a numbering class of S, in particular from min(S). For, if $\nu \in$ min(S), then aut(S) $= \{\nu'^{-1}\nu \mid \nu' \in$ min(S)$\}$ From the two numberings of min(C) (see figure 5), for instance, the identity and the transposition interchanging the 'eyes' are obtained as automorphisms.

Efficient algorithms are needed if a numbering method is to be applied in practice. The minimum class of C, for instance, is, of course, not obtained by checking successively all 30! numberings. It is a matter of creating an algorithm that eliminates a considerable number of candidate numberings with each step. In our case the two numberings of min(C) were found in 0.1 sec on an IBM 370 computer.

The minimum method - or any other numbering method - may be used to detect isomorphism between two relational systems S and S': One checks whether code(S) = code(S'). In this context the question of complexity [17] is of interest: Is there an algorithm capable of detecting isomorphism between two graphs for which the time required does not depend exponentially but polynominally on the length of the input? The answer to this still unsolved question has to be positive if there is a polynominal algorithm for the generation of code(S) for relational systems S which describe a graph.

Code(S) falls within the mathematical concept of coding [17] or canonical representation [24] of a mathematical structure. If a molecule is described by S, then code(S) may be considered to be its structural name [14]. Since this name uniquely represents the structure of the molecule, as far as it is considered, it may be used for purposes of documentation as well as for many other applications of computers in chemistry. It is, however, not only in chemistry but also in other fields that this systematic naming procedure may be applied.

REFERENCES

[1] H.L. Morgan, "The Generation of a Unique Machine Description for Chemical Structures - A Technique Developed at Chemical Abstracts Service", J. Chem. Doc. 5, 107 (1965).

[2] a) W.T. Wipke and T.M. Dyott, "Stereochemically Unique Naming Algorithm", J. Am. Chem. Soc. 96, 4834 (1974).

b) W.T. Wipke and H. Braun, in press.

[3] W.T. Wipke, S.R. Heller, R.J. Feldmann and E. Hyde, Ed., "Computer Representation and Manipulation of Chemical Information", Wiley, New York, 1974.

[4] M.F. Lynch, J.M. Harrison, W.G. Town, and J.E. Ash, Ed., "Computer Handling of Chemical Structure Information", Macdonald, London, 1971.

[5] J.E. Rush, "Status of Notation and Topological Systems and Potential Future Trends", J. Chem. Inf. Comp. Sci. 16, 202 (1976).

[6] L.M. Masinter, N.S. Srindharan, R.E. Carhart and D.H. Smith, "Labeling of Objects Having Symmetry", J. Amer. Chem. Soc. 96, 7714 (1974).

[7] a) C. Jochum and J. Gasteiger, "Canonical Numbering and Constitutional Symmetry", J. Chem. Inf. Comp. Sci. 17, 113 (1977).
b) R.E. Carhart, "Erroneous Claims Concerning the Perception of Topological Symmetry", ibid. 18, 108 (1978).
c) C. Jochum and J. Gasteiger, "On the Misinterpretation of our Algorithm for the Perception of Constitutional Symmetry", ibid. 19, 49 (1979).
d) R.E. Carhart, "Perception of Topological Symmetry", ibid. 19, 56 (1979).

[8] W. Schubert and I. Ugi, "Constitutional Symmetry and Unique Descriptors of Molecules", J. Amer. Chem. Soc. 100, 37 (1978).

[9] J. Kudo and S. Sasaki, "The Connectivity Stack, a New Format for Representation of Organic Chemical Structures", J. Chem. Doc. 14, 200 (1974).

[10] H.W. Davis, "Computer Representation of the Stereo-chemistry of Organic Molecules", Birkhäuser, Basel, 1976.

[11] a) M. Randic, "On the Recognition of Identical Graphs Representing Molecular Topology", J. Chem. Phys. 62, 3920 (1974).
b) M. Randic, "On Canonical Numbering of Atoms in a Molecule and Graph Isomorphism", J. Chem. Inf. Comp. Sci. 17, 171 (1977).

[12] a) C.A. Shelly and M.E. Munk, "Computer Perception of Topological Symmetry", J. Chem. Inf. Comp. Sci. 17, 110 (1977).
b) C.A. Shelly and M.E. Munk, "An Approach to the Assignment of Canonical Connection Tables and Topological Symmetric Perception", ibid. 19, 247 (1979).

[13] a) M. Uchino, "Algorithms for Unique and Unambiguous Coding and Symmetry Perception of Molecular Structure Diagram. I. Vector Functions for Automorphism Partitioning", J. Chem. Inf. Comp. Sci. 20, 116 (1980).
b) M. Uchino, "II. Basic Algorithm for Unique Coding and Computation of Symmetry Group", ibid. 20, 121 (1980).
c) M. Uchino, "III. Method of Subregion Analysis for Unique Coding and Symmetry Perception", ibid. 20, 124 (1980).

[14] K. Wirth, M.K. Huber, D. Pazis and A.S. Dreiding, "Structure and Names, I. Describing Method", manuscript in preparation.

[15] K. Wirth, M.K. Huber and F. Siegerist, "Structure and Names, II. Coding Method", manuscript in preparation.

[16] K. Wirth, M.K. Huber, D. Pazis, Ph. Floersheim and A.S. Dreiding, "Structure and Names, III. Naming Method in Chemistry", manuscript in preparation.

[17] R.C. Read and D.G. Corneil, "The Graph Isomorphism Desease", J. Graph Theory 1, 339 (1977).

[18] G. Gati, "Further Annotated Bibliography on the Isomorphism Disease", J. Graph Theory 3, 95 (1979).

[19] F. Stockton, "Linearization and Standardization of Graphs", Shell Development Co., Technical Progress Report No. 2-68, project No. 34430 (1968).

[20] J.F. Nagle, "On Ordering and Identifying Undirected Linear Graphs", J. Math. Phys. 7, 1588 (1968).

[21] Y.J. Shah, G.T. Davida and M.K. McCarthy, "Optimum Features and Graph Isomorphism", IEEE Trans. Systems, Man Cybernet 4, 313 (1974).

[22] A. Proskurowski, "Search for a Unique Incidence Matrix of a Graph", BIT 14, 209 (1974).

[23] B. Weisfeiler, "On Construction and Identification of Graphs", Lecture Notes in Mathematics Vol. 558, Springer, Berlin, 1976.

[24] B. Mahr, "Aspects of Graphrepresentation", Appl. Comp. Sci. 13, 307 (1979).

[25] J.W. Moon, "Topics on Tournaments", Holt, Rinehart and Winston, New York, 1968.

[26] W. Maurer, "Ueber Turniere", Didaktik der Mathematik 3, 49 (1975).