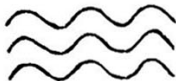


THE CHEMICAL APPLICATIONS OF GRAPH THEORY

D. H. Rouvray - University of the Witwatersrand,
Johannesburg, South Africa

In his search to make sense of his environment, man has from the earliest time utilized symbols to represent the physical and chemical objects about him. Thus, the ancient Egyptians used the symbol 1. to represent water, and the early Greeks depicted alum by means of the symbol 2.

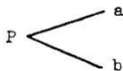


1.



2.

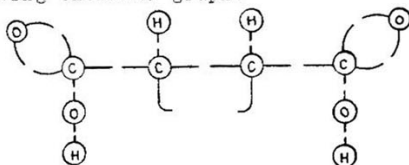
In more recent times, chemical molecules have been represented in terms of points and lines. The points have been used to represent the time-averaged positions of the atoms, and the lines the neighbourhood relations or bonds existing between the atoms. The first person to use such a representation was William Higgins, who in 1789 employed the following representation for the oxides of nitrogen:



Here nitrogen is represented by the symbol P (from its former name phlogisticated air), and the oxygen atoms by the symbols a, b, and c. Although this representation was the first depiction of the constitutional formula, the lines connecting the symbols were not thought of as chemical linkages in the modern sense. The credit for the introduction of the straight line to represent an electron pair covalent bond goes rather to Archibald Couper, who in 1858 first used this notation.

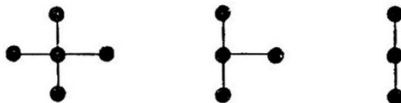
In the 1850's great interest was evident on the notion of chemical structure, and from the studies made then the

modern constitutional formula was gradually developed. Notable contributions to the development were made by Alexander Crum Brown, August Kekulé and Alexander Butlerov. The latter worker was the first to speak of the chemical structure of a species with roughly its modern meaning. In 1864 Crum Brown was the first to publish a list of pairs of isomeric chemical species, each of which he represented in graphical notation. For example, he depicted the molecules of fumaric and maleic acid by the following chemical graph:



Several of Crum Brown's depictions of differing isomers were by means of the same graph. In his depictions Crum Brown was experiencing a difficulty still present today, namely that the chemical graph does not give a unique representation for all chemical species. The problem remains even if we consider polynomials, matrices, or other mathematical representations. In fact, the only truly unique representation of a chemical species is the chemical name of the species. The name, however, provides very little useful information for most calculational purposes. This difficulty is proving especially aggravating in the storage and retrieval of chemical formulae by computer. At present there seems to be no way to overcome this major problem.

The explicit introduction of graph-theoretical concepts into chemistry was pioneered by the English mathematicians Arthur Cayley and James Sylvester. In 1857 Cayley had developed the concept of a tree. With some collaboration from Sylvester, he applied this concept to his study of the isomers of the alkane homologous series in 1874. He was the first person to point out that the constitutional formula is a chemical graph, and that chemical isomers could be enumerated by graph-theoretical techniques. The three types of tree which he enumerated are illustrated below.



These he designated from left to right as carbon trees, boron trees and oxygen trees according as the maximum valence of the vertices is four, three or two. Cayley was also first to use a generating function for the enumeration of trees. Use of such a function enabled him to enumerate the first thirteen members of the alkane series, although his last two determinations were later shown to be in error.

The next major advance in the field of isomer enumeration was made in 1931 when Henry Henze and Charles Blair put forward their iterative method. After devising appropriate recursion formulae they were able to enumerate the constitutional isomers for many of the common homologous series. In general their results did not extend beyond about the twenty fifth member of any series owing to the amount of computational labour involved. The advent of the computer in recent years has now enabled these enumerations to be carried out for the higher members of the series. The method is particularly well suited to the computer.

The most powerful method of isomer enumeration developed to date, however, rests upon the now famous Enumeration Theorem of Georg Pólya published in 1936. As in the other procedures mentioned, this method is based upon information obtained from the chemical graphs of the species to be enumerated. In this case, the method examines the symmetry of each species and utilizes this for the derivation of a cycle index for each. After making certain substitutions, the cycle index is used as an appropriately weighted generating function, from the expansion of which the isomers may be enumerated. The method is powerful because it may be used to enumerate many different types of isomer including stereoisomers. Yet there are still some problems to which the method apparently cannot be applied, an example of this being provided by the problem of enumerating the arenes.

Probably the earliest use of graph theory in chemistry, apart from the representation of chemical species, was for the study of the additive properties of chemical substances. Such study was initiated in 1842 by Hermann Kopp who investigated the additive properties of organic substances. He was able to demonstrate for instance an approximately constant difference in the heat of formation of species differing by a methylene (CH_2) group. Some three decades later Vladimir Markovnikov showed for hydrocarbons that branching of a species was always associated with a higher melting point than that for the corresponding unbranched species.

The first formal statement of the Additivity Principle was made in 1917 by Stanley Redgrove, although the principle had been adumbrated in the works of several earlier investigators. It is known now that the principle is strictly invalid because there are no truly additive physico-chemical properties of substances. The closest approach to additivity is displayed by the molecular mass of a molecule which is very nearly equal to the sum of the individual atomic masses. A number of thermodynamic properties are also roughly additive, such as the various heats of reaction. The concept of additivity was first given a mathematical formulation in 1964 when Smolenski used set-theoretical and graph-theoretical notions to derive the formula

$$f(G) = a_0 + \sum_{k=1}^n a_k |X_k|,$$

where $f(G)$ represents a physico-chemical parameter of a molecule designated by the chemical graph G , a_0 and a_k are constants determined from a small representative sample of chemical graphs, $|X_k|$ is a path in G of length k , and n the length of the longest path in G .

Use of this formula by Smolenski to predict the thermodynamic properties of hydrocarbon molecules has generally resulted in good correlations between the predicted and observed values. As the predictions were usually within 5% of the true values, the formula may

be used to obtain estimates of unknown thermodynamic parameters of hydrocarbon species with reasonable accuracy. Such studies have been developed in recent years by many workers, of whom special mention should be made of Tatevski and his coworkers in the Soviet Union. In a number of instances correlations between predicted and measured thermodynamic properties have been excellent.

Within the last decade this type of work has led fairly naturally to the introduction of the topological index. Although not an entirely new concept, as we show below, it has certainly become of increasing importance in the last few years. By a topological index is meant some mathematical formula which is derived from the chemical graph of an individual molecular species. This formula must in each case be reducible to a real number which is regarded as characteristic of the species topology. It will be evident from the foregoing that such an index cannot be uniquely characteristic of species. Yet, by plotting these indices against the known physico-chemical parameters of the bulk species, some excellent correlations have been obtained. In certain cases the correlations are so good that predictions of unknown parameters could be made with reasonable certainty. Further aspects of this relatively new approach are discussed in an article by the present author elsewhere in this volume.

A very important application of graph theory has been in the general area of chemical bonding. Apart from a limited amount of work by chemists on Feynman diagrams, which we shall not explore here, virtually all the applications have centred around Hückel molecular orbital theory. In his attempt to find a simple method for obtaining approximate solutions to the Schrödinger equation for organic species, Hückel introduced in 1931 a number of approximations to the integrals encountered. Using these approximations he was able to set up the so-called Hückel matrix. It was first demonstrated in 1954 by Ruedenberg that the Hückel matrix is equivalent to the topological matrix of a species. The latter matrix is obtained by setting up an $n \times n$ matrix for a species containing n carbon atoms and writing one for the

ij th entry if atoms i and j are bonded and zero otherwise. This matrix must clearly be symmetrical about the principal diagonal. We show below the topological matrix and also the Hückel matrix for benzene.

$$\begin{vmatrix} 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \end{vmatrix}$$

Topological matrix

$$\begin{vmatrix} x & \beta & 0 & 0 & 0 & \beta \\ \beta & x & \beta & 0 & 0 & 0 \\ 0 & \beta & x & \beta & 0 & 0 \\ 0 & 0 & \beta & x & \beta & 0 \\ 0 & 0 & 0 & \beta & x & \beta \\ \beta & 0 & 0 & 0 & \beta & x \end{vmatrix}$$

Hückel matrix

Here x stands for $\alpha - \epsilon$, where ϵ is an energy eigenvalue and α and β represent respectively the Coulomb and resonance integrals, all α and all β terms being taken as identical.

The equivalence between these matrices may be demonstrated in the general case by means of the following reasoning. Let H be the hamiltonian matrix for a π -electron system, S the overlap matrix with all entries equal to s , and ϵ an energy eigenvalue of the system. Then we may write

$$|H - \epsilon S| = 0.$$

The matrices H and S may be resolved as indicated:

$$H = \alpha I + \beta A$$

$$S = I + sA,$$

where I is the unit matrix of the same order as H , S and A , and A is the topological matrix. Substitution of the resolved matrices into the original equation yields the result

$$|H - \epsilon S| = \left| A - \begin{pmatrix} \epsilon - \alpha \\ \beta - \epsilon s \end{pmatrix} I \right| = 0.$$

This reveals that the eigenspectrum of H may be represented by that of the topological matrix A . Furthermore, because

H and A commute, they will be possessed of the same set of eigenvectors.

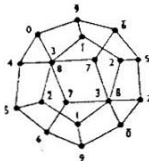
The equivalence of these two matrices leads us to the conclusion that the Hückel energy levels are determined by the molecular topology of the species represented. It is also clear that the topology will determine quantities which may be derived from the Hückel matrix, such as the π -electron charge density on carbon atoms, the bond order between a pair of atoms, and the various parameters used to characterize polarizability. Accordingly, all these parameters must be regarded as topological indices. However, since the bond order of Coulson, for instance, dates back to 1939, the topological index cannot be regarded as a recent discovery. Such indices are in fact of about four decades standing.

Another interesting observation is that the Pairing Theorem of Coulson and Rushbrooke, put forward in 1940, is now seen to be purely a consequence of the topological structure of the species to which it applies. The alternant molecules for which the theorem holds are those possessed of a bipartite graph. It is very well known for such graphs that if ϵ_k is an eigenvalue in the graph spectrum then the eigenvalue $-\epsilon_k$ will also appear in the spectrum. The theorem has in turn led to results such as the so-called Hückel $4n + 2$ rule, which predicts aromatic stabilization in chemical species displaying an electronic orbital filling pattern of the $4n + 2$ type. A large number of other similar results, which we cannot go into here, are now realised to originate in the species topological structure. In passing, it should be mentioned that numerous contributions to this general area have been made in recent years by the Trinajstić group of workers in Yugoslavia.

Thus far only those applications of graph theory have been considered in which the atoms of individual molecular species are represented as vertices. This need not always be the case, and we turn now to take a brief look at some instances where a whole molecule may be represented by a single vertex. The edges of the graphs which may be constructed in this representation symbolize relations between the

molecules themselves. As an example of a relation of this type we cite the case of a transformation occurring between a pair of molecules, which might take the form of an interconversion from one molecular structure or state to another. Because transformations take place in a given direction, the type of graph frequently employed in this field is the directed graph or digraph.

Specific examples of sequences of transformations which have been represented graphically are the interconversions of the penta- and hexa-coördinate complexes of the type ML_n , where M is a metal atom and L a ligand, and n an integer. One possible graphical representation of the interconversions of the ML_5 system due to Muetterties is illustrated below.



This mode of representation has proved especially valuable in the study of isomerization reactions. The principal uses to which the representation has been put are the study of the possible synthesis routes to be followed when starting out from a given set of synthons to produce a desired end product, the determination of the constitution of unknown intermediates or products, and the prediction of conceivable new synthesis products. The type of graphs used in this approach have been designated reaction graphs by Balaban, a worker who has made many notable advances. Other names deserving of mention in this particular field are those of Dubois, Gielen, King and Prelog.

To sum up we may say that the main uses of graph theory in chemistry are twofold. In the first instance graphs have been employed for the representation, systematization, enumeration, and construction of topological indices for individual chemical species. Secondly, graphs have been employed for the representation of molecular species which are undergoing transformations or interconversions of some kind such as that of isomerization. As a result of these

studies, there has been a steadily growing interest in the chemical applications of graph theory and a virtual explosion in the number of publications appearing in this field. The topology of a species is now seen to influence, if not actually determine, a wide variety of the physical and chemical properties of the bulk substance. Yet, if our studies be confined only to those with a bearing on the species topology, only limited progress can be expected for much information is lost in the purely topological representation of species. For this reason it may well be necessary to supplement the information provided by the chemical graph with other information. This could involve for example the use of group-theoretical information when studies are being made on the chiral nature of species, or the use of spaces of dimension higher than three when the interactions of species are being considered.

In conclusion, I propose the establishment of a new journal which would specialize in the chemical applications of set theory, graph theory, topology, and combinatorial theory. Although somewhat hesitant in suggesting yet another new journal, I feel that that this one could fulfill a real need for those of us working in the above indicated fields. Many of us have experienced considerable difficulties in the past in getting our work accepted in the leading journals. Furthermore, contributions in this general area are extremely widely scattered at the moment. I call upon all those interested to meet together after this lecture to discuss my proposal. (As a result of this call it was decided to start a new journal to be called Mathematical Chemistry(MATCH) covering the areas indicated. The present paper appears in the first issue of this new journal.)

General Bibliography

A. T. Balaban, Chemical Graphs. III. Reactions with Cyclic Six-membered Transition States, Rev. Roumaine Chim., 12, 875 (1967).

A. T. Balaban, Aplicatii Ale Teoriei Grafurilor in Chimie,

Editura Academiei Republicii Socialiste Romania, Bucharest, 203 (1972).

A.T. Balaban (ed.), Chemical Applications of Graph Theory, a monograph published by Academic Press, London, 1976.

I. Gutman and N. Trinajstić, Graph Theory and Molecular Orbitals, Topics Curr. Chem., 42, 49 (1973).

D.H. Rouvray, Graph Theory in Chemistry, Royal Inst. Chem. Revs., 4, 173 (1971).

D.H. Rouvray, The Additivity Principle from a Topological Viewpoint, Chem. Technol., 3, 379 (1973).

D.H. Rouvray, The Search for Useful Topological Indices, Amer. Sci., 61, 729 (1973).

D.H. Rouvray, Uses of Graph Theory, Chem. Britain, 10, 11 (1974).

D.H. Rouvray, Isomer Enumeration Methods, Chem. Soc. Revs. London, 3, 355 (1974).

Editor's Note: We have received only one comment on a lecture delivered at the Microsymposium. It is the following comment of Prof. Dr. H.-H. Schmidtke, University of Dusseldorf, concerning the lecture delivered by Dr. D. H. Rouvray on Tuesday morning.

SOME COMMENTS ON THE LECTURE BY DR. D. H. ROUVRAY

by

Prof. Dr. H.-H. Schmidtke

Listening to this lecture and to those from yesterday about the chemical application of graph theory one may come to the conclusion that this theory only can describe information which is on the standard of chemical and physical knowledge of the last or possible of the beginning of this century. The results which were obtained from quantum mechanics can only be covered, if at all, by a more sophisticated theory of graphs in which physical insight has been lost to a high extent. The claim of graph theory that it introduces a simple and straightforward concept for classifying chemical and physical experience will not be justified and the respective situation seems to be similar to that in quantum chemistry in its advanced stage.

For instance, in recent years plenty of novel type compounds are synthesized which cannot be described by simple "Strichformeln". These compounds may be described by three center, multi center or even isotropic bonds in the constitutional formulae. Graph theory consequently must introduce in the mapping X to Y for the set X new sets $P^{\{3\}}$, $P^{\{4\}}$, . . . , $P^{\{\infty\}}$ (in the denotation of Dr. Kerber's lecture) which unduely complicates the theory losing more and more of its physical insight.