What Is Mathematical Chemistry?

Alexandru T. Balaban[1*]
[1] Texas A&M University Galveston, 77551, USA
*e-mail balabana@tamug.edu

Abstract: Examples from the author's personal experience illustrate how mathematical chemistry applies graph theory and discrete mathematics to chemical problems, without involving intermediate physics. In molecular graphs, vertices represent atoms and whose edges represent covalent bonds. In reaction graphs vertices represent molecules or ions and edges symbolize elementary reaction steps.

Keywords: Graph theory; discrete mathematics; molecular graphs; reaction graphs; (3,g)-cages

1. Introduction

Chemistry is considered to be a Central Science that bridges the hard sciences (physics, astronomy) with soft biomedical sciences where experiments are not perfectly reproducible. There is a long-standing controversy whether chemistry is reducible to physics. Paul Dirac wrote: *"The fundamental laws necessary for the mathematical treatment of a large part of physics and the whole of chemistry are thus completely known, and the difficulty lies only in the fact that application of these laws leads to equations that are too complex to be solved."* Chemical physics and physical chemistry both benefit from mathematics, and the Nobel Prize for Chemistry was awarded in 2013 to Martin Karplus, Michael Levitt and Arieh Warshel "for the development of multiscale models for complex chemical systems", using computer programs for solving quantum-chemical equations.

At the beginning of the previous century, when the guano deposits of South America were depleted, Fritz Haber tried to obtain nitrogen fertilizers by combining nitrogen with hydrogen in the presence of iron catalysts. He found out that at 1000°C and under atmospheric pressure, the equilibrium concentration of the resulting ammonia was only 0.01%. However, he was not discouraged and by applying the laws of physical chemistry he increased the yield to 8% by increasing the pressure to 175 atm. and reducing the temperature to 550°C. The industrial production of ammonia was perfected by Carl Bosch and nowadays one reactor can produce a million tons of ammonia per year. Wilhelm Ostwald had shown that ammonia can be oxidized catalytically by oxygen affording nitric acid so that the total balance amounts to:

$$O_2 + 2N_2 + 4H_2O = NH_4^+\,NO_3^-$$

with ammonium nitrate ($NH_4^+\,NO_3^-$) as the most common nitrogen fertilizer, solving (at least for the present) the soil fertility problem. Nobel Prizes for Chemistry honored Ostwald, Haber, and Bosch in 1909, 1918, and 1931, respectively.

The interaction between chemistry with other branches of mathematics (graph theory and discrete mathematics) originates in trying to predict physical or chemical properties or biological activities of substances (numerically measured) from their molecular structures which can be represented by graphs (discrete entities). This is the area in which I became interested 60 years ago, when I was a student, and then working for my Ph. D. degree with Professor Costin D. Nenitzescu at the Faculty of Industrial Chemistry in the Bucharest Polytechnic. I published with him 26 papers in 1958–1964, and two book chapters in a monograph edited in 1964 by George A. Olah who would earn the Nobel Prize for Chemistry in 1994 [1]. The synthesis of pyrylium salts by alkene diacylation (found in my Ph. D. work) is known as the Balaban-Nenitzescu-Praill reaction [2].

2. Chemistry and Graph Theory

For enumerating all possible isomers of heterocyclic aromatic compounds with one ring $X_xY_yZ_z$ where the three types of atoms (X, Y, and Z) provided 2, 1, or 0 π-electrons, respectively, I had to solve a problem analogous to the "necklace problem" containing $x + y + z = m$ beads of three colors; the Hückel Rule for aromatic delocalization of π-electrons requires $2x + y = 4n + 2$, where n is 0, 1, 2, etc. with n = 1 corresponding to an aromatic π-electron sextet as in benzene (formula **1** on next page). With the help of my friend Silviu Teleman, this enumeration was carried out and published in 1958 [3], but as I learned later it could be easily solved by means of the Polya theorem. At that time I could not find Roumanian graph-theorists interested in chemical problems, but Frank Harary, one of the leading American graph-theorists, became interested, visited me in Bucharest, and we started a fruitful collaboration. First we had to learn each other's vocabulary, e. g. with chemical terms in italics: structural formula = molecular graph; atom = vertex; covalent bond = edge; acyclic molecule = tree graph; valency of an atom = vertex degree; number of rings = cyclomatic number; energy level = eigenvalue of adjacency matrix; n-annulene = cycle on n vertices, etc. In 1967 I edited the first book in this field with chapters written by 16 authors (including Harary and myself) from 8 countries [4]; Vladimir Prelog signed its preface in June 1975, and in October he learned that he would be one of the two laureates for the Chemistry Nobel Prize for 1975. A Chinese translation was published in 1983, and since then this research field has developed rapidly in China.

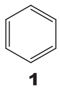
For this type of problems involving discrete mathematics, I founded in 1975 together with André Dreiding, Adalbert Kerber, and Oskar E. Polansky (whom I had known since he was a Professor at the University of Vienna and I was in 1967–1970 a senior research officer at the International Atomic Energy Agency in Vienna) the first peer-reviewed periodical in this field; its main editor was O. E. Polansky at the Max-Planck-Institute for Radiation Chemistry in Mülheim/Ruhr. After Professor Polansky passed away, the journal was edited at the University of Bayreuth by A. Kerber. At present this journal is edited by Ivan Gutman at the University of Kragujevac, Serbia, and it is called MATCH, Communications in mathematical and in computer chemistry. In 1987 Dennis H. Rouvray founded the Journal of Mathematical Chemistry edited at present by Paul G. Mezey (Canada) and Springer Verlag.

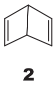
During the following years (1966–1991), I published a series of papers in a sequence with the general title Chemical Graphs (Parts 1 to 54) covering various aspects of this field, to be discussed in the following pages.

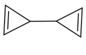
I consider that one of my most important contributions to mathematical chemistry occurred in 1975 when on presenting at Harvard University a lecture on chemical applications of graph theory, I converted to this research field a Croatian post-doc who happened to be in the audience, Milan Randić, who had worked till then on quantum-chemical problems; with his background in mathematics and physics, he made important contributions in this field. After his return to Zagreb, he developed in Yugoslavia a research school that produced many eminent mathematical chemists. One of them, Nenad Trinajstić, published an influential book, Chemical Graph Theory, which appeared in two editions and is highly cited [5]. Another one, Ivan Gutman, edits now MATCH.

3. Valence isomers of annulenes

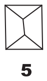
Planar conjugated cyclopolyenes $(CH)_m$, where m is an even number, are called for brevity m-annulenes. According to Hückel's Rule, annulenes with an odd number of conjugated double bonds (i. e. $m = 4n + 2$ as in benzene which has $n = 1$) are stabilized by continuous in-phase overlap of π-electronic orbitals. This stabilization is called "aromaticity" for historic reasons. Among the 6×10^7 known chemicals registered in the Chemical Abstracts database, more than half are aromatic, and their stability is an important aspect for synthetic polymers and for natural macromolecules such as DNA (the four bases that encode genetic information are aromatic heterocycles). Several amino acids that exist in proteins have aromatic rings. In organic chemistry it is customary to draw "hydrogen-depleted formulas" that ignore any hydrogen atoms when they are connected to carbon atoms. Graph-theoretical methods for isomer enumeration allow the prediction of all possible isomers for given molecular formulas. A covalent bond between two atoms, i. e. a shared pair of electrons with opposite spins, is represented by a line. When atoms may be connected by double or triple bonds, structural formulas of such molecules correspond to multigraphs. It can be shown that there are six multigraphs $(CH)_6$ but only the first five (**1**–**5**) can represent molecules. The last one, which cannot be written without crossing lines, is called a non-planar graph.

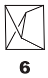

1


2


3


4


5


6

Of course, only benzene (**1**) is a geometrically-planar molecule whereas **2**–**5** (valence isomers of benzene, or cubic graphs with 6 vertices) are tridimensional structures.

During the 1960s, all these valence isomers were synthesized and studied. I enumerated all possible valence isomers of various annulenes in several papers from the "Chemical graphs" series.With two co-authors I published a 3-volume book on valence isomers of annulenes [6], which was used by synthetic chemists in their effort to explore such structures and their hetero-analogues. One of the first valence isomer of $(CH)_{10}$ was first made in Bucharest and is called the Nenitzescu hydrocarbon, as presented in detail in his biography co-authored with M. D. Banciu [7].

4. Chemical documentation

Experimental organic chemists are confronted with the problem of finding out whether a given structure has been synthesized earlier. Organic chemical nomenclature is a complicated, yet powerful, system characterizing uniquely chemical structures. The International Union of Pure and Applied Chemistry up-dates and publishes regularly the IUPAC Rules for organic and inorganic substances. In 1963 I was elected as corresponding member of the Roumanian Academy when I was 32 years old, and the Chemistry Section of the Academy had put me in charge with the Roumanian adaptation of these IUPAC Rules. Although I had been elected as Roumania's delegate at the IUPAC Nomenclature Committee, I was never allowed to travel abroad for debates. However, I supervised the Roumanian adaptation of the two nomenclature volumes for organic and inorganic chemistry by Dr. D. Purdela (1977, 1986), and biochemical nomenclature by Dr. I. F. Dumitru (1984).

Computer search of the huge Chemical Abstracts database has simplified enormously the chemical documentation by the fact that nowadays if one looks for chemical substances it is possible to access all the information be using chemical graphs instead of words. Instead of spending days in libraries, one gets in seconds or minutes all the available information on any of the 60 million substances. This fact makes chemistry the best documented science. Of course, when searching for words (which can be ambiguous or can vary according to authors for one and the same idea, property, phenomenon or theorem), the search needs longer times as in all other sciences that access information by using words. The American Chemical Society (ACS), which sponsors Chemical Abstracts, publishes a journal that changed names (Journal of Chemical Documentation, Journal of Chemical Information and Computer Sciences, and at present Journal of Chemical Information and Modeling). I was on the Editorial Board of this journal for several years, and I published many of my papers in this journal. The Division of Chemical Information of the ACS honoured me with the 1994 Herman Skolnik Award in recognition of outstanding contributions to chemical information. Two years later, a similar award was given to Milan Randić.

With Milan Randić we were the two chemists invited to participate in a symposium entitled Quo Vadis Graph Theory? bringing together scientists from different disciplines [8]. Earlier, I had co-authored a chapter in a book entitled Applications of Graph Theory edited by mathematicians [9]. In both cases and in two other reviews [10,11] I had tried to present the rich chemical applications of non-numerical and symbolical mathematics. More recently, I edited a book linking topological and graph-theoretical aspects of chemical applications with the geometry of molecules [12].

5. Molecular descriptors and QSAR/QSPR

The quantitative measurements of physical, chemical, or biological properties of chemical substances are expressed in numbers that vary on a continuous scale. By contrast, chemical structures of these molecules are represented by discrete graphs. How can one devise correlations between chemical structure and properties, which could be extended to predicting structures with desired properties? The first answers to this question were provided by Harold Wiener in two papers published in 1947. A simpler definition of the Wiener index W and the name “topological index” (TI) for such molecular descriptors (numbers associated with molecular graphs) was provided by Haruo Hosoya in 1971, when he introduced another descriptor, denoted by Z. The most cited and useful descriptor molecular connectivity χ was introduced by Milan Randic in 1975: for each edge ij of the hydrogen-depleted graph, one computes and sums the inverse square root of the product of vertex degrees [13]. These vertex degrees are the sums of entries in the adjacency matrix of the graph over rows or columns; entries are 1 for adjacent vertices and 0 otherwise. Like all topological indices, they do not characterize uniquely the graph, but are “degenerate”. (different graphs have the same TI).

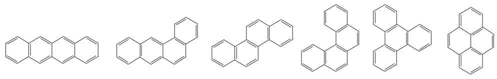
$$\chi = \sum_{ij} (v_i v_j)^{-1/2}$$

The Wiener index is derived from another matrix called the distance matrix, which is also symmetrical; its entries ij are the topological distances between vertices i and j, along the shortest path between them. The Wiener index is the half-sum of all entries in the distance matrix and has a higher degeneracy than χ. Sums over rows or columns in the distance matrix are called distasums. In 1982 I defined an even more selective TI, denoted by J, which is now called the Balaban index [14]. It is similar to χ, but instead of vertex degrees it has distasums; in order to avoid an automatic increase both with branching and graph size, the sum over all edges is multiplied by a factor $e/(e - v + 2)$, where e is the number of edges and v is the number of vertices. Among the interesting properties of index J is the fact that for an infinitely-long linear alkane it has the asymptotic value of the number π. It also allows an easy adaptation for the presence of multiple bonds and heteroatoms. Together with James Devillers (from Lyon, France) I edited a book on molecular descriptors [15]. Several chapters in that book are written as collaborations with my former Ph. D. students Ovidiu and Teodora Ivanciuc, who followed me to Galveston, Texas when I retired from the Bucharest Polytechnic on reaching the retirement age. Other TIs that my collaborators and I developed were information-theory-based indices with my son [16] and distance-based indices with Mircea Diudea [17].

The most important use of TIs was for finding quantitative structure-property or structure-activity relationships (QSPRs and QSARs). A nice illustration is provided by finding an immune-suppressing decapeptide by Lahana and co-workers on using index J together with a few other descriptors; the computational search started with a quarter-million possible candidates, and ended with five that were synthesized and tested, yielding a structure that was 100 times more active than the “lead compound” [18]. Nowadays, more refined computer programs exist for drug design, helping the pharmaceutical industry to find continuously newer medicinal drugs, because germs develop resistance to antibiotics. I reported about TIs and their uses in several reviews [19,20]. The handbook by Todeschini and Consonni [21] contains the most comprehensive overview about TIs.

6. Benzenoids and their dualists

Polycyclic aromatic hydrocarbons (benzenoids for brevity) with condensed benzenoid rings such as naphthalene are valuable starting materials for plastics, dyestuffs, pesticides, detergents, and pharmaceuticals. Enumerating all possible isomers of benzenoids with h six-membered rings can be carried out by hand [22] but for large h values this is now done by computer programs [23]. With Frank Harary, I called attention to the fact that the graph (“dualist”) obtained by inscribing a vertex at the center of each benzenoid ring, and connecting by edges vertices of adjacent rings characterizes uniquely all benzenoids. Moreover, it helps classify benzenoids into catafusenes, perifusenes, or coronafusenes according to whether the dualist has no rings, 3-membered rings, or larger rings, respectively [24]. In the following figure for benzenoids with four rings, one sees four non-branched catafusenes with their notation based on the angles in the dualist (digit 0 for 180° and 1 or 2 for 120° or 240°), followed by a branched catafusene, and a perifusene (no notation for it).

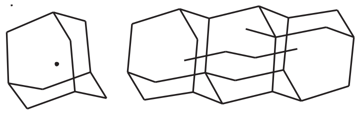


00 01 12 11 1(.)

According to Eric Clar [25], the aromatic sextet of electrons is not shared equally between adjacent benzenoid rings but depends on the topology of the benzenoid. In collaborations with Milan Randić [26] I. Gutman [27] and Matevz Pompe (from Slovenia) [28], we analyzed the partition of all π-electrons in benzenoid molecules based on various assumptions. Many properties of benzenoids can be correlated bi-parametrically with the numbers h of rings and the numbers of linearly-condensed rings. Some topological properties of benzenoids were discussed in collaborations with scientists in USA [29,30] and Roumanian mathematicians: Ioan Tomescu [31] and Cristian Artemi [32] from the Faculty of Mathematics of the Bucharest University; with Solomon Marcus [33], we applied picture grammars, context-free grammars and push-down automata to chemical problems.

7. Diamondoids and their dualists

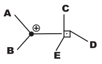
Hydrocarbons with a carbon scaffold similar to that found in diamond and called diamondoids. Here again the dualist concept can serve for classification into perimantanes and catamantanes according to whether the dualist has or does not have cycles, respectively. The first formula on the next page is for adamantane $C_{10}H_{16}$, and its center is shown by a dot. The second formula is for [121]tetramantane and the zigzag line is its dualist .

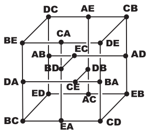


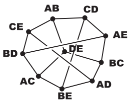
Diamondoids are the most stable among all their polycyclic isomers. With Schleyer (who had discovered a new synthesis of adamantane $C_{10}H_{16}$) we developed a naming and coding system for catamantanes based on the geometry of the dualist [34] which is now generally used. One can isolate diamondoids from petroleum, and they may be used as starting materials for new pharmaceuticals.

8. Reaction graphs and trivalent cages

In addition to molecular graphs, one can consider other associations between graph theory and chemistry. In one of these, graph vertices symbolize reaction intermediates, and edges symbolize elementary reaction steps. The first among such "reaction graphs" dealt with the rearrangement of ethyl-carbenium cations with substituents A, B, C, D, and E, where one the two carbons (black dot and white square in the figure) was isotopically labelled. Any of the three substituents can migrate to the positively-charged carbon that has only two substituents, so that in the reaction graph one vertex is adjacent to three other vertices. The result is the Levi graph of the Desargues configuration with 20 vertices [35] denoted by the two substituents (in alphabetical order if attached to the black dot, or contrary to that order if attached to the white square). The second figure below shows this graph.

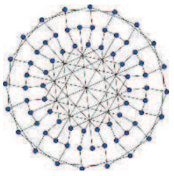






If there is no isotopic label, the graph reduces to 10 vertices, again denoted by two letters, and the result is the famous Petersen graph, which is also the trivalent 5-cage (a regular graph where all vertices are incident to three edges such that the smallest circuit has girth g = 5 vertices; the third of the preceding figures represents this graph). In 1994 I published a review on reaction graphs [36].

The trivalent g-cage problem had a gap for 9-, 10-, and 11-cages. When I was in Vienna I tried to solve this open problem and I conjectured possible solutions [37]. Mathematicians (Exoo, McKay, Myrwold) proved with computer programs and years-long calculations that two of my conjectures were correct. The 10-cage shown in the next figure, the first from the three possible 10-cages [38], has 70 vertices, a nice symmetry is known as "Balaban's 10-cage". "Balaban's 11-cage" is unique and has 112 vertices. No trivalent g-cages are yet known for g > 12.



9. Reply to the title question

Both August Comte and Immanuel Kant had written that mathematical methods were incompatible with chemistry. Several authors attempted to provide replies to the title question of this article [39,40] and I prefer an adaptation of Ivar Ugi et al.'s goal of mathematical chemistry [41]: it is the mathematization of chemistry without the intermediacy of physics, and the direct solution of chemical problems by qualitative (often discrete) mathematical methods.

10. International Academy of Mathematical Chemistry

In 2005 the International Academy of Mathematical Chemistry (IAMC) was founded in Dubrovnik, Croatia, with me as president and Milan Randić as vice-president. Since 2008 the president has been Roberto Todeschini, and at the annual meeting in June 2014 new elections will be held in Split, Croatia; it is likely that Milan Randić will be the next president. The structure, member names with their countries, and objectives of IAMC are stated in the web page: www.iamc-online.org. So far, from Roumania there are 5 out of all 98 members: A. T. Balaban, M. Diudea, O. Ivanciuc, G. Paun, and I. Tomescu.

11. References

1) Balaban, A. T. and Nenitzescu C. D., in Friedel-Crafts and Related Reactions, editor Olah, G. A. Wiley-Interscience, New York, 1964, vol. 2, pp. 979-1047; Nenitzescu C. D. and Balaban, A. T. vol. 3, pp.1033-1152.

2) Balaban, A. T., In: Name Reactions in Heterocyclic Chemistry II (J. J. Li, editor) Wiley, Hoboken, NJ, 2011 pp. 402-419.

3) Balaban, A. T. Studii si cercetari chim., Acad. R. P. Roumania, 7 (1959) 257 (in Roumanian).

4) Balaban, A. T. (editor): Chemical Applications of Graph Theory, Academic Press, London, 1976.

5) Trinajstić, N. Chemical Graph Theory, CRC Press, Boca Raton, Florida, 1999.

6) Balaban, A. T.; Banciu, M. and V.Ciorba, V. Annulenes, Benzo-, Hetero-, Homo-Derivatives and Their Valence Isomers, CRC Press, Boca Raton, Florida, 1986, 3 volumes.

7) Balaban, A. T. Annals Discrete Math., 55 (1993) 109; reprinted in Quo Vadis, Graph Theory? (edirtors Gimbel, J.; Kennedy, J. V. and Quintas, L. V.), North Holland, Amsterdam, 1993.

8) Rouvray D. H. and Balaban, A. T., in Applications of Graph Theory, (editors Wilson, R. J. and Beineke, L. V.), Academic Press, London, 1979, pp. 177-221.

9) Banciu, M. and Balaban, A. T. C.D. Nenitzescu: His Life and His Work (in Roumanian), Editura Academiei Romane, Bucharest, 1995.

10) Balaban, A. T. Sigsam Bull., 8 (1984) 29.

11) Balaban, A. T. in EUROCAL-85, Lecture Notes in Computer Science, No. 203 (editor Buchberger, B.), Springer, Berlin, 1985, pp. 68-79.

12) Balaban, A. T. (editor), From Chemical Topology to Three-Dimensional Geometry, Plenum Publishing Corporation, New York, 1997.

13) Randić, M. J. Am. Chem. Soc. 97 (1975) 6609.

14) Balaban, A. T. Chem. Phys. Lett. 80(1982)399.

15) Devillers, J. and Balaban, A. T. (editors), Topological Indices and Related Descriptors in QSAR and QSPR, Gordon and Breach, Amsterdam.

16) Filip, P. A.; Balaban, T. S. and Balaban, A. T. J. Math. Chem. 1 (1987) 61.

17) Diudea, M. V. and Balaban, A. T. J. Chem. Inf. Comput. Sci. 33 (1993) 421.

18) Grassi, G.; Calas, B.; Yasri, A.; Lahana, R. et al. Nature Biotechnol. 16 (1998) 748.

19) Balaban, A. T.; Motoc, I.; Bonchev, D. and Mekenyan, O. Topics Curr. Chem., 114 (1983) 21.

20) Balaban, A. T. Pure Appl. Chem. 55(1983) 199.

21) Todeschini, R. and Consonni, V. Molecular Descriptors for Chemoinformatics, Wiley-VCH, Weinheim, 2009.

22) Balaban,A.T. et al. Z. Naturforsch., 42c (1987) 863.

23) Brinkmann, G., Caporossi, G., Hansen, P. J. Chem. Inform. Comput. Sci. **43** (2003) 842.

24) Balaban, A. T. and Harary, F. Tetrahedron, 24 (1968) 2505.

25) Clar, E. The Aromatic Sextet, Wiley, London, 1972.

26) Balaban, A. T. and Randić, M. New J. Chem. 32 (2008) 1071.

27) I. Gutman and A. T. Balaban, J. Serb. Chem. Soc. 76 (2011) 1505.

28) Balaban, A. T.; Pompe, M. and Randić, M. J. Phys. Chem. A, 112 (2008) 4148.

29) Balaban, A. T. and Klein, D. J. J. Phys. Chem. C, 113 (2009) 19123.

30) Balaban, A. T. and Schmalz, T. G. J. Chem. Inf. Model. 46 (2006) 1563.

31) Balaban, A. T. and Tomescu, I. Discrete Appl. Math. 19 (1988) 5.

32) Balaban, A. T. and Artemi, C. Polycyclic Aromatic Compounds, 1 (1970) 171.

33) Balaban, A. T.; Barasch, M. and Marcus, S. MATCH, Comm. Math. Comp. Chem. 8 (1980) 193.

34) Balaban, A. T. and Schleyer, P. von R. Tetrahedron, 34 (1978) 3599.

35) Balaban, A. T. in Graph Theoretical Approaches to Chemical Reactivity, (Bonchev, D. and Mekenyan, O. editors), Kluwer Academic Publishers, The Netherlands, 1994, pp.137-180.

36) Balaban, A. T.; Rentea, C. C. and Ciupitu, E. Rev. Roum. Chim., 13 (1968) 231.

37) Balaban, A. T. J. Comb. Theory B, 12 (1972) 1.

38) Balaban, A. T. Rev. Roum. Math. Pures Appl., 18 (1973) 1033.

39) Trinajstić, N. and Gutman, I. Croat. Chem. Acta 75 (2002) 329.

40) Balaban, A. T. Foundations Chem. 7 (2005) 289.

41) Ugi. I. et al. Angew. Chem. Int. Ed. Engl. 32 (1993) 201.