Construction algorithms at molecular graphs and their application*

Original: Konstruktionsalgorithmen bei molekularen Graphen
und deren Anwendung

Author:
Thomas Wieland†
Lehrstuhl Mathematik II, Universität Bayreuth
D-95440 Bayreuth, Germany

Advisor:
Prof. Dr. Reinhard Laue
Lehrstuhl Mathematik II, Universität Bayreuth

Date of approval: December 1996

1 Introduction

The mathematical discipline of graph theory has of its origins in the description of chemical substances. In the second half of the 19th century Liebig, Kekulé, Crum Brown and others developed a mathematical model for molecules [2,3] where the atoms are regarded as vertices and the bonds as edges of a bond graph. This model became quickly the standard. Later mathematicians began to use it, first for the establishment of theoretical notions and for the deduction of counting formulae. Important contributions came from H. Redfield [26] and G. Pólya [24]. From the time when powerful computing machines were available, constructive problems could be attacked. The center of interest usually lied in the contruction of all graphs corresponding to given conditions (cf., e.g., [6,8,10,13]), like the generation of all structural isomers for a given empirical formula and optionally further restrictions [11,12,22,31].

The work described here discusses construction problems that are studied less frequently and shows up with new methods for the *joining* of graphs. As quite a number of practical applications revealed, the huge variety of structures that can be generated be today's graph generators is extremely hard to handle and must usually be limited by additional pieces of information. Moreover, there are several

*Published as [30]

†Present address: German Remote Sensing Data Centre, German Aerospace Research Establishment (DLR), Oberpfaffenhofen, D-82234 Wessling, Germany, email: `Thomas.Wieland@dlr.de`

construction problems in mathematical chemistry as well as in graph theory which cannot be traced down to isomers.

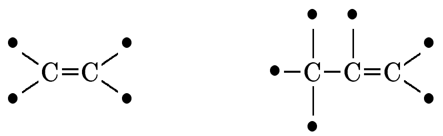
2 Joining graphs by adding edges

We consider labeled, undirected, loop-free connected multi-graphs, i.e. as mappings $\gamma_{\in} m^{\mathbf{p}^{[2]}} := \{f | f : \mathbf{p}^{[2]} \to \{0, \ldots, m-1\}\}$ given as the connection types of their pairs of points $\mathbf{p}^{[2]}$ where $\mathbf{p} := \{1, \ldots, p\}$. An additional labeling with atom names is not necessary but admissible.

We assume that each vertex of a graph has a maximal degree assigned besides its current degree. In analogy to chemistry we will call it valency. So the first construction problem (chapter 3) reads:

> *Determine all non-isomorphic graphs that result from combining two graphs and adding edges in all possible ways such that the degree of each vertex does not exceed its valency.*

The connection sites are given by the difference of the current and the maximal degree, just like in the following example:



I.e., we have to determine all essentially different possibilities of inserting single and multiple edges between the two structures. To formalize this, we define: Let be $\gamma \in m^{\mathbf{p}^{[2]}}$. Set $f^{(\gamma)} := \sum_{i=1}^{p}(val_i - d_i)$. Then the graph $\hat{\gamma} : \{1, \ldots, p + f^{(\gamma)}\}^{[2]} \to \{0, \ldots, m\}$, resulting from connecting each vertex $i \in \mathbf{p}$ with $val_i - d_i$ new bivalent vertices (where d_i is the degree of vertex i), is called *extended graph.* Note that the extension need not comprise all possible connecting sites, but can be restricted appropriately.

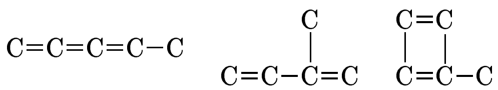
So regarding two graphs $\hat{\gamma}_1$ and $\hat{\gamma}_2$, we are looking for mappings

$$\hat{\epsilon} : \{1, \ldots, f^{(\gamma_1)}\} \times \{1, \ldots, f^{(\gamma_2)}\} \to \{0, 1\}$$

$$\text{with} \quad \sum_{i=1}^{f^{(\gamma_1)}} \hat{\epsilon}(i,j) \leq 1 \quad \forall j \in \{1, \ldots, f^{(\gamma_2)}\}$$

$$\sum_{j=1}^{f^{(\gamma_2)}} \hat{\epsilon}(i,j) \leq 1 \quad \forall i \in \{1, \ldots, f^{(\gamma_1)}\}.$$

The non-isomorphic connections between $\hat{\gamma}_1$ and $\hat{\gamma}_2$ of this kind can be shown to correspond to the representatives of the orbits of the group actions of some subgroups of the automorphism groups $Aut(\hat{\gamma}_1)$ and $Aut(\hat{\gamma}_2)$ on these matrices. In this example, there are 31, e.g.:



The numbers of solutions for this problem grow indeed very fast and can reach a million already for a pair of graphs with four and seven vertices, respectively.

3 Merging two graphs over all common subgraphs

Again we consider multi-graphs with and without atomic labels. The construction problem reads:

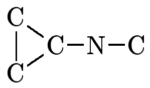
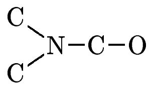
> *Determine all non-isomorphic molecular graphs that result from merging two graphs over all their common subgraphs.*

We proceed as follows:

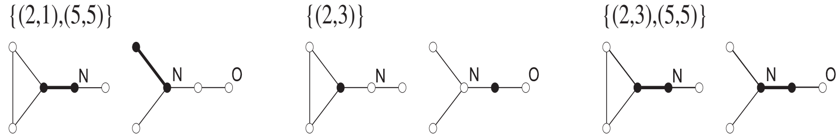
i) Generation of all non-isomorphic common subgraphs (chapter 4)

ii) Determination of all non-isomorphic mappings of the found subgraphs onto one another and construction of the graphs resulting from these mappings (chapter 5)

The search for all common subgraphs alone is a a problem of its own value that has been studies in graph theory as well as in chemistry several times [1,5,7,9,18,20,21,23]. We continue the development of some of those ideas, especially the compatibility graphs, present a strigent formalization of the problem, and utilize newer mathematical construction methods like simultaneous orderly generation for the search. So we are able, e.g., to determine the sixteen non-isomorphic common subgraphs of


and


Three of them can be shown here:



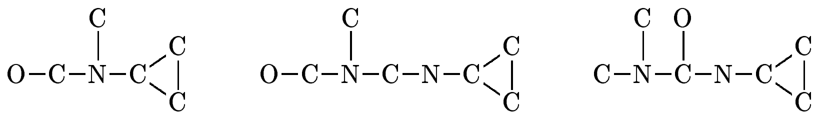
For each common subgraph we then want to calculate all assignment possibilities of the corresponding vertices and all mergings that can be performed over these vertices within the limits given by the maximal degrees.

Let $\gamma_1 \in m^{\mathbf{p}^{[2]}}$ and $\gamma_2 \in m^{\mathbf{q}^{[2]}}$ be the multigraphs; let $\eta_1 \in m^{{T_1}^{[2]}}$ and $\eta_2 \in m^{{T_2}^{[2]}}$ be common subgraphs of γ_1 and γ_2, i.e. $\eta_1 \subseteq \gamma_1$ and $\eta_2 \subseteq \gamma_2$ as well as $|T_1| = |T_2| = k \leq \min(p, q)$, such that there is an edge-preserving isomorphism $\sigma_0 : T_1 \to T_2$. It is obvious that the automorphism group G of η_1 when we see it as a part of γ_1 is usually smaller than the full $Aut(\eta_1)$. The same is the case for the group H of η_2.

From the Gluing Lemma (cf. [19]) we obtain that the set of all essentially different assignments between η_1 and η_2 corresponds to the double cosets

$$\sigma_0(\sigma_0^{-1} G \sigma_0) \backslash Aut(\eta_2)/H.$$

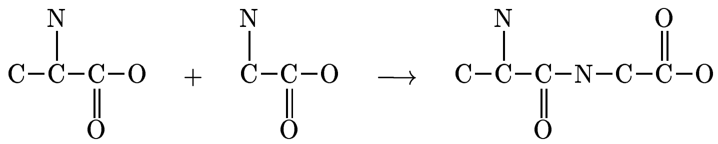
For the example above, there are twelve solutions of this kind; three of them are:



4 Rule-based joining by adding or eliminating of edges, modelling chemical reactions

For molecular graphs, there is also quite often the case that not all possible connections are of interest, but only a few, precisely determined ones. Such graphs result, for example, in the course of a chemical reaction. In chapter six a syntax and an algorithm for the redundancy-free construction of such molecules is developed.

One example is the establishment of a peptid bridge between amino acids:



The following problems are discussed in particular: What possibilities exist for the assignment of a set of ligands to a skeleton structure and how can these be constructed?

5 Applications to chemistry

First we study the iterative usage of the merging technique for problems from structure elucidation, the search for quantitative structure activity relationships ($QSAR$, cf. [4,17,28]) and drug research.

Another chapter is devoted to a currently popular synthesis technique: *combinatorial chemistry* [27,29]. Roughly speaking, this is about the assignment of a set of building block compounds to a skeleton molecule. While the construction aspects of this problem have been discussed in a previous chapter, we present here a method for the intelligent selection of the building blocks as well as formulae for the enumeration of the size of the resulting combinatorial libraries.

In the last, but not least chapter the perspectives of the usage of the joining and merging algorithms for the enhancing the input to isomer generators are shown.

6 Mathematical backgrounds

The primary objective of this work are algorithms that meet the following requirements:

- The construction is fast, complete, and redundancy-free.
- Restrictions of the user are taken into account as early as possible.

This is the only way to guarantee relevance for the practical use.

The problems mentioned above are formulated in the context of actions of permutation groups on sets of finite mappings [15,19]. The central tool that is used in almost all our generation algorithms is *orderly generation* [8,14,25] of representatives of orbits of group actions. There the set on which the action is defined is decomposed into invariant disjoint subsets and the transversal is built successively by employing a total order. As a new paradigm, “simultaneous orderly generation” is introduced.

In order the reduce the expense we make use of the so-called *homomorphism principle*. Because after mapping an orbit problem to an easier one (like a smaller one), we just have to consider – according to this principle – the action of the pre-images of the stabilizers of the found representatives on the pre-image sets. Beside the split of orbits we also employ the homomorphism principle to describe fusions of orbits.

The third decisive notion are *double cosets* [16,19]. When examining pairs of graphs, the non-isomorphic solutions of the construction problems often correspond to double cosets after the automorphism groups of the graphs. By interpreting these cosets as orbits of certain group actions, we can integrate this notion seamlessly into our overall context.

7 Final remarks

All algorithms presented in this work have been implemented by the author in the C++ programming language in an object-oriented way and combined to the program system NEXTOR.

A lot of the mathematical methods are very general and can hence be applied to many kinds of discrete structures. The concentration on graphs and molecular graphs should therefore be understood as an example, motivated by application, but not required by the methodology.

References

[1] D.M. Bayada, R.W. Simpson, A.P. Johnson, and C. Laurenço, *An algorithm for the multiple common subgraph problem*, J. Chem. Inf. Comput. Sci., Vol. 32 (1992), 680–682.

[2] C. Benecke, R. Grund, R. Hohberger, A. Kerber, R. Laue, and T. Wieland, *Die Veranschaulichung von Molekülen im Chemieunterricht mit Hilfe des Programmsystems MOLGEN*$^+$, MNU - Der mathm.-naturw. Unterr., Vol. 48 (1995), 156–163.

[3] N.L. Biggs, K.E. Lloyd, and R.J. Wilson, *Graph theory 1736-1936*, Clarendon Press, 1977.

[4] D. Bonchev, *Information Theoretic Indices for Characterization of Chemical Structures*, Research Studies Press, Chichester, 1983.

[5] L. Chen and W. Robien, *MCSS: A new algorithm for perception of maximal common substructures and its application to NMR spectral studies. 1. The algorithm*, J. Chem. Inf. Comput. Sci., Vol. 32 (1992), 501–506.

[6] C.J. Colbourn and R.C. Read, *Orderly algorithms for generating restricted classes of graphs*, J. Graph Theo., Vol. 3 (1979), 187–195.

[7] M.M. Cone, R. Venkataraghavan, and F.W. McLafferty, *Molecular structure comparison program for the identification of common substructures*, J. Am. Chem. Soc., Vol. 99 (1977), 7668–7671.

[8] I.A. Faradzhev, *Generation of non-isomorphic graphs with given partition of vertex degrees*, Algorithmic Studies in Combinatorics (I.A. Faradzhev, ed.), Nauka, Moskau, 1978, (In Russian), pp. 11–19.

[9] V.E. Golender and A.B. Rozenblit, *Logical and combinatorial algorithms for drug design*, Research Studies Press, Letchworth, Herts., 1983.

[10] R. Grund, *Konstruktion schlichter Graphen mit gegebener Gradpartition*, Bayreuther Mathem. Schr., Vol. 44 (1993), 73–104.

[11] R. Grund, *Konstruktion molekularer Graphen mit gegebenen Hybridisierungen und überlappungsfreien Fragmenten*, Bayreuther Mathem. Schr., Vol. 49 (1995), 1–113.

[12] R. Grund, A. Kerber, and R. Laue, *Construction of discrete structures, especially isomers*, Discr. Appl. Math., Vol. 67 (1996), 115–126.

[13] T. Grüner, *Ein neuer Ansatz zur rekursiven Erzeugung von schlichten Graphen*, Master's thesis, Universität Bayreuth, 1995.

[14] T. Grüner, R. Laue, and M. Meringer, *Applications for group actions applied to graph generation*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science (Providence, RI) (L. Finkelstein and C. Kantor, eds.), 1995.

[15] A. Kerber, *Algebraic Combinatorics Via Finite Group Actions*, BI-Wissenschaftsverlag, Mannheim, Wien, Zürich, 1991.

[16] A. Kerber and R. Laue, *Group actions, double cosets, and homomorphisms: unifying concepts for the constructive theory of discrete structures*, 1994.

[17] L.B. Kier and L.H. Hall, *Molecular Connectivity in Structure-Activity Analysis*, Research Studies Press, Chichester, 1986.

[18] V. Kvasnička and J. Pospichal, *Maximal common subgraphs of molecular graphs*, Rep. Mol. Theo., Vol. 1 (1990), 99–106.

[19] R. Laue, *Construction of combinatorial objects – a tutorial*, Bayreuther Mathem. Schr., Vol. 43 (1993), 53–96.

[20] G. Levi, *A note on the derivation of the maximal common subgraphs of two directed or undirected graphs*, Calcolo, Vol. 9 (1972), 341–352.

[21] J.J. McGregor and P. Willet, *Use of a maximal common subgraph algorithm in the automatic identification of the ostensible bond changes occuring in chemical reactions*, J. Chem. Inf. Comput. Sci., Vol. 21 (1981), 137–140.

[22] S.G. Molodtsov, *Computer-aided generation of molecular graphs*, MATCH, Vol. 30 (1994), 213–224.

[23] V. Nicholson, C.C. Tsai, M. Johnson, and Naim. M., *A subgraph isomorphism theorem for molecular graphs*, Graph Theory and Topology in Chemistry (R.B. King and D.H. Rouvray, eds.), Elsevier Science Pub., Amsterdam, 1987, pp. 226–230.

[24] G. Pólya, *Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen*, Acta mathematica, Vol. 68 (1937), 145–253.

[25] R.C. Read, *Every-one a winner*, Ann. Discr. Math., Vol. 2 (1978), 107–120.

[26] H.J. Redfield, *The theory of group-reduced distributions*, Am. J. Math., Vol. 49 (1927), 433–455.

[27] T. Wieland, *Mathematical simulations in combinatorial chemistry*, MATCH, Vol. 34 (1996), 179–206.

[28] T. Wieland, *The use of structure generators in predictive pharmacology and toxicology*, Arzneim.-Forsch./Drug Res., Vol. 46 (I) (1996), 223–227.

[29] T. Wieland, *Combinatorics of combinatorial chemistry*, J. Math. Chem., Vol. 21 (1997), 141–157.

[30] T. Wieland, *Konstruktionsalgorithmen bei molekularen Graphen und deren Anwendung*, MATCH, Vol. 36 (1997), 7–156.

[31] T. Wieland, A. Kerber, and R. Laue, *Principles of the generation of constitutional and configurational isomers*, J. Chem. Inf. Comput. Sci., Vol. 36 (1996), 413–419.