communications in mathematical
match
and in computer chemistry

no. 36, October 1997

ISSN 0340-6253

MATCDY (36) 7-155 (1997)

Konstruktionsalgorithmen bei molekularen Graphen und deren Anwendung*

Thomas Wieland†
Lehrstuhl II für Mathematik, Universität Bayreuth,
D-95440 Bayreuth

Überblick

Die mathematische Disziplin der Graphentheorie hat einen ihrer Ursprünge in der Beschreibung chemischer Substanzen. In der zweiten Hälfte des 19. Jahrhunderts wurde unter Mitwirkung von Liebig, Kekulé, Crum Brown u.a. ein mathematisches Modell für Moleküle entwickelt [12, 17], bei dem die Atome als Ecken und die Bindungen als Kanten eines Bindungsgraphen aufgefaßt werden und das schnell zum Standard etablierte.
Dieses Modell wurde in der Folgezeit von Mathematikern zunächst zur Herausbildung theoretischer Begriffe und zur Ableitung von Abzählformeln verwendet. Bedeutende Beiträge dazu lieferten H. Redfield [106] und G. Pólya [103].
Als später leistungsfähige Rechenmaschinen zur Verfügung standen, wurden auch konstruktive Probleme behandelt. Dabei lag das Zentrum des Interesses zumeist auf der Konstruktion aller Graphen zu gegebenen Bedingungen (s. etwa [30, 36, 52, 56]), also z.B. die Erzeugung aller Bindungsgraphen zu einer gegebenen Bruttoformel und weiteren Nebenbedingungen [53, 55, 96].

Wir wollen in dieser Arbeit weniger untersuchte Konstruktionsaufgaben diskutieren und dabei neue Methoden für das *Verbinden* von Graphen aufzeigen. In vielen praktischen Anwendungen hat sich nämlich gezeigt, daß die ungeheure Vielfalt von Strukturen, die von Graphengeneratoren erzeugt werden können, nur schwer handzuhaben ist und durch Zusatzinformationen eingeschränkt werden muß. Daneben gibt es auch eine Reihe von Erzeugungsproblemen in der mathematischen Chemie und in der Graphentheorie, die sich nicht auf Isomere zurückführen lassen.
Obwohl einige Begriffe im folgenden erst noch klar zu definieren sind, wollen wir kurz die drei Methoden der Verbindung von Graphen umreißen, die in dieser Arbeit beschrieben werden:

1. Verbindung zweier Graphen durch Hinzufügen von Kanten

Wir betrachten numerierte ungerichtete, schleifenfreie zusammenhängende Multigraphen. Unsere Methodik werden wir allgemein formulieren, so daß eine zusätzliche Beschriftung, etwa mit Atomtypen, nicht notwendig, aber auch nicht hinderlich ist.

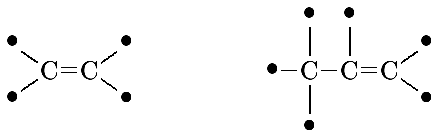
*Diese Arbeit wurde von der Fakultät für Mathematik und Physik der Universität Bayreuth als Dissertation zur Erlangung des Grades eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigt. 1. Gutachter: Prof. Dr. R. Laue, 2. Gutachter: Prof. Dr. A. Kerber, Tag der Einreichung: 18. Oktober 1996, Tag des Kolloquiums: 20. Dezember 1996

†Jetzige Anschrift: Deutsches Fernerkundungsdatenzentrum, Deutsche Forschungsanstalt für Luft- und Raumfahrt e.V. (DLR), Oberpfaffenhofen, D-82234 Weßling, E-Mail: Thomas.Wieland@dlr.de

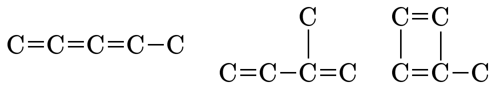
Wir setzen voraus, daß jede Ecke der Graphen neben ihrem aktuellen Grad noch einen maximalen Grad besitzt, den wir in Analogie zur Chemie als Valenz bezeichnen. Dann heißt die Konstruktionsaufgabe, der wir uns in Kapitel 3 widmen:

> *Bestimme alle nichtisomorphen Graphen, die sich durch Zusammenfassen der beiden Graphen und Hinzufügen von Kanten ergeben, so daß der Grad jeder Ecke höchstens so groß wie ihre Valenz ist.*

Die Differenz zwischen dem aktuellen und dem maximalen Grad, die freien Valenzen, liefert uns die Anknüpfpunkte, wie etwa bei folgendem Beispiel:



D.h. es sind alle wesentlich verschiedenen Möglichkeiten zu bestimmen, Einfach- und Mehrfachbindungen zwischen den Strukturen einzufügen. Bei diesem Beispiel gibt es 31, z.B.



(Sämtliche Lösungsgraphen werden in Abb. 3.1, S. 57, gezeigt.) Die Anzahlen an Lösungen wachsen allerdings sehr schnell und können bei einem Graphenpaar mit vier bzw. sieben Punkten bereits über einer Million liegen.

2. Verschmelzung zweier Graphen über alle gemeinsamen Teilgraphen

Auch hierbei betrachten wir Multigraphen mit und ohne Beschriftung. Die Konstruktionsaufgabe lautet:

> *Bestimme alle nichtisomorphen molekularen Graphen, die sich durch Verschmelzen zweier Graphen über alle ihre gemeinsamen Teilgraphen ergeben.*

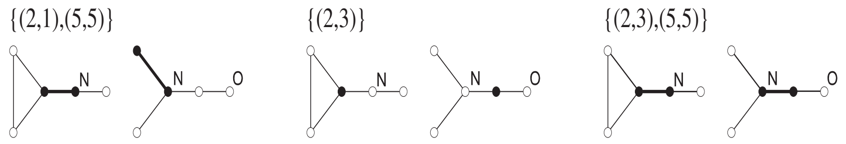
Dazu gehen wir wie folgt vor:

i) Erzeugung aller nichtisomorphen gemeinsamen Teilgraphen (Kapitel 4)

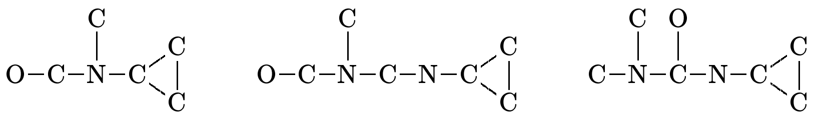
ii) Bestimmung aller nichtisomorphen Abbildungen der gefundenen Teilgraphen aufeinander und Konstruktion der aus diesen Abbildungen resultierenden Graphen (Kapitel 5)

Bereits die Suche nach allen gemeinsamen Teilgraphen ist dabei ein eigenständiges Problem, das sowohl in der Graphentheorie als auch in der Chemie bereits mehrfach untersucht wurde. Wir werden einige dieser Ideen weiterentwickeln, eine durchgängige Formalisierung des Problems vorstellen und neuere mathematische Konstruktionsverfahren zur Suche einsetzen. Mit deren Hilfe sind wir dann in der Lage, beispielsweise die 16 nichtisomorphen gemeinsamen Teilgraphen von

zu bestimmen. Sie werden in Abb. 4.2, S. 69, gezeigt; drei davon sind:



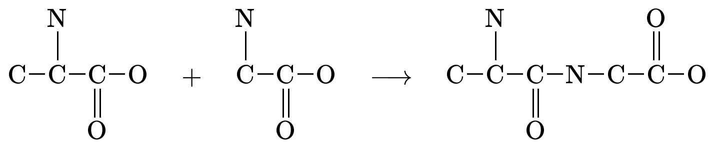
Zu jedem gemeinsamen Teilgraphen wollen wir sodann alle Zuordnungsmöglichkeiten der Ecken und alle Verschmelzungen berechnen, die über diese Ecken unter Einhaltung der maximalen Grade möglich sind. Zu obigem Beispiel gibt es 12 derartige Lösungen; drei davon sind



Diese und die weiteren zeigt auch Abb. 5.2 auf S. 95.

3. Verbindung über Verknüpfungsregeln unter Hinzunahme oder Wegfall von Ecken, in Nachbildung einer chemischen Reaktion

Bei molekularen Graphen gibt es aber auch den Fall, daß nicht alle möglichen Verbindungen eine Rolle spielen, sondern nur einige wenige, genau festgelegte. Solche Graphen ergeben sich z.B. im Verlauf einer chemischen Reaktion. Eine Syntax und einen Algorithmus zur redundanzfreien Konstruktion derartiger Moleküle werden wir in Kapitel 6 entwickeln.
Ein Beispiel ist die Ausbildung einer Peptidbrücke zwischen Aminosäuren:



Insbesondere wollen wir uns der Frage widmen, welche Möglichkeiten für die Anbindung einer Menge von Liganden an eine Gerüststruktur es gibt und wie diese konstruiert werden können.

Anwendungen in der Chemie

Bei den Anwendungen untersuchen wir in Kapitel 7 zunächst den iterativen Einsatz des Verschmelzungsverfahrens bei Problemen aus der Strukturaufklärung, der Suche nach quantitativen Struktur-Aktivitätsbeziehungen (*QSAR*) und der Wirkstoff-Forschung.
Ein weiteres Kapitel (8) ist einer derzeit populären Synthesemethodik der organischen Chemie gewidmet, der *kombinatorischen Chemie*. Grob gesprochen geht es dabei um die Anlagerung einer Menge von Bausteinsubstanzen an ein Gerüstmolekül – ein Problem, dessen konstruktive Aspekte wir ja bereits vorher diskutiert haben. Wir werden sowohl Verfahren zur geschickten Auswahl der Bausteine als auch zur Abzählung der Möglichkeiten der Anlagerungen vorstellen.
Im letzten Kapitel werden schließlich Einsatzmöglichkeiten der Verknüpfungs- und Verschmelzungsalgorithmen zur Verbesserung der Eingabe eines Isomerengenerators vorgestellt.

Mathematische Hintergründe

Unser Ziel sind Algorithmen, die folgenden Anforderungen genügen:

- Sie konstruieren schnell, vollständig und redundanzfrei.
- Sie berücksichtigen Nebenbedingungen des Anwenders (sog. *Restriktionen*) so früh wie möglich.

Nur auf diese Weise ist auch die Relevanz für die Praxis gewährleistet.

Die oben angeführten Problemstellungen führen wir auf Operationen von Permutationsgruppen auf Mengen endlicher Abbildungen zurück. Das zentrale Werkzeug, das bei fast allen Konstruktionsalgorithmen eingesetzt wird, ist die *Ordnungstreue Erzeugung* von Repräsentanten von Gruppenoperationen. Dabei wird die Menge, auf der operiert wird, in invariante disjunkte Teilmengen zerlegt und die Transversale unter Ausnutzung einer Totalordnung sukzessive aufgebaut. Als neues Paradigma führen wir die simultane Ordnungstreue Erzeugung mehrerer Repräsentantensysteme ein.
Um den Aufwand der Verfahren weiter zu reduzieren, machen wir vom sog. *Homomorphieprinzip* Gebrauch. Bildet man nämlich ein Bahnenproblem auf ein leichter zu lösendes ab, etwa auf ein kleineres, so müssen nach diesem Prinzip nur noch Operationen der Urbilder der Stabilisatoren der gefundenen Repräsentanten auf deren Urbildmengen betrachtet werden. Neben der Aufspaltung von Bahnen beschreiben wir mit dem Homomorphieprinzip aber auch die Fusion von Bahnen.
Ein dritter zentraler Begriff sind *Doppelnebenklassen*. Bei der Betrachtung von Paaren von Graphen entsprechen die nichtisomorphen Lösungen der Konstruktionsprobleme oft Doppelnebenklassen nach den Automorphismengruppen der Graphen. Da aber auch diese als Bahnen von Gruppenoperationen beschrieben werden können, fügt sich dieser Begriff nahtlos in unsere übrige Darstellung ein.

Alle in dieser Arbeit vorgestellten Algorithmen wurden auch vom Autor in der Programmiersprache *C++* objektorientiert implementiert und zum Programmsystem NEXTOR zusammengefaßt.
Die meisten der hier vorgestellten Verfahren sind neu und in dieser Form nicht in der Literatur zu finden. (Bei der Diskussion des Algorithmus zur Suche nach gemeinsamen Teilgraphen wird gesondert auf die in dieser Arbeit

entwickelten Neuerungen hingewiesen.) Es soll an dieser Stelle betont werden, daß viele der vorgestellten mathematischen Verfahren ganz allgemein und auf viele Arten diskreter Strukturen anwendbar sind. Die Konzentration auf Graphen und molekulare Graphen ist somit exemplarisch zu verstehen und ist durch die Anwendung motiviert, aber nicht durch die Methodik bedingt.

Inhaltsverzeichnis

Vorwort

Angeregt durch die Vorarbeiten, die von D. Moser [70, 98] und insbesondere Dr. R. Grund am Lehrstuhl II für Mathematik der Universität Bayreuth geleistet wurden [53, 54] und die zum Strukturgenerator **MOLGEN** [14, 131] führten, ergab sich der Wunsch nach einer Ausweitung der bestehenden Verfahren. Die Anforderungen entsprangen dabei zwei relativ unterschiedlichen Richtungen: Zum einen sollte ein theoretischer Beitrag im Rahmen der am Lehrstuhl untersuchten *konstruktiven Theorie diskreter Strukturen* [56, 69, 84, 85] geleistet werden, zum anderen brachte der durch den Vertrieb von **MOLGEN** in Gang gekommene intensive Kontakt mit einer Reihe von Chemikern verschiedener Forschungsrichtungen viele neue Fragestellungen aus der mathematischen Chemie mit sich. Somit wurde bei der Ausgestaltung dieser Arbeit der Versuch unternommen, beiden Anforderungen gleichmäßig Rechnung zu tragen, was zu einer Zweiteilung führte in einen theoretischen Teil, der verschiedende Algorithmen und ihre Hintergründe vorstellt, und einen anwendungsbezogenen Teil, der sich jedoch nicht damit begnügt, exemplarische Fälle der zuvor diskutierten Verfahren aufzuzeigen, sondern eigenständige Resultate vorzustellen versucht.

Denn auch die Mathematik steht heute, am Ende des zwanzigsten Jahrhunderts, vor neuen Herausforderungen: Die Menschen bringen mehr Erkenntnisse denn je zuvor hervor; ihr Wissen verdoppelt sich alle zehn Jahre. Daher können auch die einzelnen Fachdisziplinen ihre Grenzen nicht mehr in derselben Weise ziehen wie im 19. Jahrhundert, als sie sich etablierten. Kaum ein Wissenschaftler kann es sich heute noch leisten, sein Interesse nur auf seine Fachrichtung einzuschränken und andere völlig zu ignorieren. In den immer stärker fließenden Übergangszonen bilden sich neue Wissensgebiete heraus. So dürfen auch Abhandlungen, die die Grenzen zu überbrücken versuchen, ja die bemüht sind, den nicht immer einfachen Dialog zwischen den Fachsprachen zu verbessern, nicht nur unter dem Blickwinkel eines Fachgebietes gesehen, sondern müssen in ihrer Gesamtheit beurteilt werden.

Mein Dank gilt zunächst den Profs. R. Laue und A. Kerber, die mich an den Themenkreis heranführten, stets ein offenes Ohr für Fragen aller Art hatten und mir viel Freiraum für eigene Ideen ließen. Daneben danke ich allen Mitarbeitern an **MOLGEN** und am Lehrstuhl II für Mathematik, insbesondere Herrn Dr. C. Benecke, der neben vielen technische Hilfestellungen auch einige nützliche Programmroutinen beisteuerte, und Herrn Dipl.-Math. T. Grüner, der mir bei fast allen theoretischen Schwierigkeiten weiterhelfen konnte und dessen Implementierungen verschiedener Verfahren von großem Wert waren. Nicht zuletzt danke ich meiner Frau Andrea, deren Geduld und Verständnis diese Arbeit erst möglich gemacht haben.