

Rainer E. Burkard
Wolfgang Maass
Peter Weibel (Hg.)
Zur Kunst des formalen Denkens

Passagen Verlag

Deutsche Erstausgabe

Gedruckt mit freundlicher Unterstützung des
Bundesministeriums für Wissenschaft und Verkehr in Wien
und des Bundeskanzleramtes Sektion Kunst

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Zur Kunst des formalen Denkens
/ Hrsg. von Rainer Burkard ...
- Dt. Erstausg. - Wien : Passagen-Verl., 2000
(Passagen Kunst) (Perspektiven einer
Wissenschaftskultur in Österreich)
ISBN 3-85165-358-0

Alle Rechte vorbehalten
ISBN 3-85165-358-0
© 2000 by Passagen Verlag Ges. m. b. H., Wien
Graphisches Konzept: Ecke Bonk
Umschlagphoto: Peter Weibel, 1986
Collage: Karin Buol-Wischenau
Fotos (Collage): Landesmuseum Joanneum Graz, Bild und Tonarchiv
Satz: Heike Graf
Druck: Manz, Wien

Inhalt

Zu diesem Band	13
Einleitung	15
Anmerkungen	20
Literatur	20
Peter Weibel	
Kuriosa der Zahlenkunde und die Kunst – kurz gefaßt und leicht faßlich dargestellt	21
Symbole, Persönlichkeiten, Zahlen	22
Primzahlen	24
Die Goldbachsche Vermutung	28
Vollkommene Zahlen und Mersennesche Primzahlen	28
Befreundete Zahlen	32
Die numerische Sensibilität	34
Die Fibonacci-Zahlen	37
Das Pentagramm und der Goldene Schnitt	40
Die harmonikalen Proportionen und der pythagoräische Traum	44
Das Ende des pythagoräischen Traums	49
Der neue pythagoräische Traum: das digitale Prinzip	51
Das gnomonische Wachstum und die logarithmische Spirale	52
Thermodynamisch definierte Formen	55
Fraktales Wachstum: Symmetrie und Selbstsimilarität	56
Formale Grammatiken	57
Anmerkungen	60
Literatur	62
Raimund Seidel	
Die Kunst, die Gunst des Zufalls zu erzwingen	65
Einleitung	65
Den Zufall kontrollieren	67
Zufällige Ordnung und ein geometrisches Optimierungsproblem	68
Zufällige Zeugen und Primzahlbestimmung	73
Mit Hilfe des Zufalls Schätzen und Zählen	78

Eine algebraische Methode	80
Eine Irrfahrtmethode	84
Schlußbemerkungen	86
Danksagung	87
Anmerkungen	87
Literatur	87
Wolfgang M. Schmidt	
Die Kunst im formalen Denken anhand zweier Beispiele aus der Zahlentheorie	91
Anmerkungen	99
Gottfried Tinhofer	
Graphentheorie: Strukturen und Symmetrien	101
Einleitung	101
Graphen	102
Das Postbotenproblem	106
Elektrische Netze	109
Planare Graphen	115
Zusammenhang und Zuverlässigkeit	119
Rundreisen	121
Färbungen	123
Perfekte Graphen	126
Chemische Graphen	130
Schluß	134
Anmerkungen	135
Literatur	135
Farbabbildungen	137
Herbert Edelsbrunner	
Spielereien mit Kreisen und Kugeln. Zum Thema Form und Verformung	153
Zusammenfassung	153
Einleitung	153
Von Kreis und Dreieck zu Form	153
Mathematik und Informatik	154
Übersicht	154

Mehr oder weniger Fläche	155
Problemformulierung	155
Kurze Geschichte	156
Kontinuierliche Kontrahierung	156
Teilkomplexe	158
Vier Kreisscheiben	158
Voronoi und Delaunay Komplexe	159
Duale Komplexe	159
Inklusion-Exklusion Formeln	160
Kurze Geschichte	161
Bausteine des Lebens	161
Atomkugel Modelle	162
Löcher	162
Taschen	164
Anwendungen	164
Form und Formenraum	165
Glatte Oberflächen	166
Topologische Verwandlung	166
Kanonische Verformung	167
Formenraum	168
Literatur	170
Wolfgang Woess	
Irrfahrten	173
Einleitung	173
Irrfahrten in endlichen und unendlichen Straßen	174
Irrfahrten in unendlichen Graphen	179
Irrfahrten, elektrische Netzwerke und Rohrsysteme	181
Isoperimetrische Ungleichungen	187
Starke isoperimetrische Ungleichung für Pflasterungen in der Ebene	189
Schlußbemerkungen	190
Literatur	191
Rainer E. Burkard	
Zuordnungsprobleme - ein Streifzug	
durch die kombinatorische Optimierung	193
Das Heiratsproblem	193
Möglichst große Paarungen	195

Eine Anwendung bei der Kommunikation über Satelliten	196
Lineare Zuordnungsprobleme	200
Rundreiseprobleme	202
Quadratische Zuordnungsprobleme	203
Ausblick	205
Literatur	206
Wolfgang Maass	
Das menschliche Gehirn – nur ein Rechner?	209
Einführung	209
Vorboten eines neuen Typs intelligenter Rechner	212
Vom Logiker zum Käfer: Wechselnde Vorbilder für Maschinelle Intelligenz	213
Maschinen, die aus ihren eigenen Erfahrungen lernen	217
Von der Biologie inspirierte neue Rechnerstrukturen	218
Wie unterscheidet sich ein künstliches Neuronales Netz von seinem Vorbild in der Natur?	223
Konklusion	226
Anmerkungen	229
Literatur	231
Wolfgang Maass und Peter Weibel	
Brain Drain oder Internationalisierung: Überlegungen zur Situation der Mathematik und Theoretischen Informatik in Österreich	235
Appendix	237
Internet Links für weitere Informationen	241
Kurzbiographien der Autoren	243

Zu diesem Band

Rainer E. Burkard, Wolfgang Maass und Peter Weibel

Im deutschen Sprachraum gibt es selten eine Möglichkeit, sich als Nicht-Fachwissenschaftler einen Eindruck von der gegenwärtigen Forschung im Bereich der Mathematik und theoretischen Informatik zu machen. Im englischsprachigen Raum gibt es hingegen nicht nur hervorragende Fachjournalisten, die es vermögen, auch avancierteste Wissenschaftsdisziplinen und -ergebnisse inhaltlich korrekt und dennoch allgemein verständlich darzustellen, sondern auch die Fachwissenschaftler selbst unternehmen immer wieder große Anstrengungen, ihre Leistungen oder die Leistungen ihrer Fachdisziplinen populär und adäquat zugleich darzustellen. Zahlreiche wissenschaftliche Bestseller, von Stephen J. Hawkings über John Casti bis zu John H. Conway, belegen brillant den Erfolg dieser Bemühungen. So entsteht eine Wissenschaftskultur und in der Öffentlichkeit das Bewußtsein von der Relevanz der Wissenschaft für die Gesellschaft. Aus bekannten historischen sozial- und kulturpolitischen Gründen mangelt es im deutschsprachigen Raum an einer solchen Wissenschaftskultur. Daher lebt die deutschsprachige Wissenschaftskultur, in welchem Umfang und Ausmaß sie auch immer existieren mag, in großer Abhängigkeit von ausländischen Importen. Dieses Buch ist daher einer der wenigen Versuche von Fachwissenschaftlern selbst, auch im deutschen Sprachraum die aktuellen Ergebnisse einer avancierten schwierigen Forschungsdisziplin allgemein zugänglich zu machen. Für die beispielhafte Anstrengung und ihre Bereitwilligkeit, trotz des Drucks ihrer zahlreichen fachlichen Verpflichtungen, ihre Kompetenz zur Verfügung zu stellen und damit diese kulturellen Absichten der Herausgeber zu unterstützen, kann den Autoren nicht genug gedankt werden.

Das intellektuelle Vergnügen, das die Vorträge von Herbert Edelsbrunner, Wolfgang Schmidt, Raimund Seidel, Gottfried Tinhofer und Wolfgang Woess bereiteten, reproduziert auch dieser Band, der somit eine seltene Möglichkeit bietet, sich aus erster Hand einen Eindruck von der gegenwärtigen Forschung im Bereich der Mathematik und theoretischen Informatik zu verschaffen. Was diese Vorträge bzw. Artikel über ihre inhaltliche Qualität

und Innovation hinaus auszeichnet, ist ihre stilistische Qualität, das Wissensmanagement. Sie erfordern nämlich kaum spezifische Vorkenntnisse. Das ermöglicht es auch, daß alle Kapitel vollkommen unabhängig in beliebiger Reihenfolge gelesen werden können. Im Laufe der Vorbereitung dieses Bandes ist der Wunsch aufgetaucht, daß auch die Herausgeber selbst einen Einblick in ihr Arbeitsgebiet geben. Die Herausgeber sind dieser Anregung gerne gefolgt und haben daher die Vortragsmanuskripte für diesen Band um eigene Beiträge ergänzt. So ist in der Tat ein Buch entstanden, das nicht nur eine große Tradition aus den 20er Jahren wiederbelebt, sondern vielleicht auch exemplarisch für den Beginn einer eigenen Wissenschaftskultur in Österreich stehen mag.

Daher sei den Autoren für ihre beispielhaften Bemühungen nochmals herzlich gedankt, und auch all jenen Personen, deren Unterstützung, Kompetenz und Arbeit sowohl das Symposium wie auch das Buch ermöglicht haben: Dr. Irolt Killmann (Rektor der TU Graz), Dr. Christa Steinle (Leiterin der Neuen Galerie), Prof. Dr. Robert Tichy (Vorstand des Instituts für Mathematik der TU Graz), Dr. Leonhard Summerer, Heike Graf, Ingrid Preininger, Karin Buol-Wischenau, Margot Goettsberger und Clemens Heuberger. Wir danken der Gesellschaft der Freunde der Neuen Galerie dafür, daß sie durch ihre finanzielle Unterstützung das Symposium und diesen Band ermöglicht hat.

Rainer E. Burkard, Wolfgang Maass, Peter Weibel

Einleitung

Wolfgang Maass und Peter Weibel

Anlässlich der Ausstellung "Jenseits von Kunst" in der Neuen Galerie in Graz vom 7. 2. - 30. 3. 1997 fand am 7. und 8. März 1997 ein Symposium mit dem Titel "Zur Kunst des formalen Denkens" statt, das von Rainer Burkard (Institut für Mathematik an der TU Graz), Wolfgang Maass (Institut für Grundlagen der Informationsverarbeitung an der TU Graz) und Peter Weibel (Neue Galerie am Landesmuseum Graz) konzipiert wurde¹. Ziel dieser Ausstellung und des umfassenden begleitenden Schau- und Lesebuches "Jenseits von Kunst" (1997 ebenfalls im Passagenverlag erschienen), wo die Leistungen und Werke von hunderten österreichischen und ungarischen KünstlerInnen und WissenschaftlerInnen mosaikartig vernetzt und nach einem neuartigen methoden- und problemgeschichtlichen Modell präsentiert wurden, war es, eine innovative und informative Kartographie der Kultur zu erstellen: die überwiegend verdrängte, vertriebene und unbekannte Kunst- und Wissenschaftsgeschichte von Österreich und Ungarn. Beide Länder haben von der Jahrhundertwende bis zur Gegenwart trotz der politischen Destruktionen und Obstruktionen analytische Kunstrichtungen (wie Konstruktivismus, Kinetik, optische Kunst, Körperanalyse, visionäre Architektur) und Denkströmungen (wie Sprachphilosophie, Spieltheorie, Kybernetik, Psychoanalyse, Quantenphysik) begründet oder wesentlich mitgetragen, die eigenständige, spezifische Beiträge zur Weltkultur bilden.

Es handelte sich also um keine traditionelle Kunstaussstellung, sondern um die Darstellung einer kulturellen Produktion und kultureller Zusammenhänge, die von der offiziellen Kulturpolitik bis zur Unsichtbarkeit verdrängt wurden. Im Vergleich der beiden Länder Österreich und Ungarn wurden kulturelle Leistungen sichtbar, die dem anderen wie dem eigenen Land aufgrund der oppressiven politischen Umstände (vom Kollaps der Monarchie bis zum Kalten Krieg) nicht vertraut sind.

Der Schwerpunkt wurde dabei auf analytische und abstrahierende Methoden der Weltauffassung (von den Formalwissenschaften zu den Formalkünsten) gelegt: erstens weil die Formalwissenschaften längst einen

zentralen, wenn auch unsichtbaren Teil unserer Kultur bilden, die in so vielen Bereichen auf der Basis von Technologie konstituiert wird, und zweitens, weil Österreich und Ungarn gerade auf diesem Gebiet der analytischen Methoden überragende Leistungen erbracht haben, ganz im Gegensatz zu dem vom Inland wie Ausland entworfenen repressiven und reaktionären Bild eines barock-expressiven Österreichs und Ungarns. Es mußte daher über die klassischen Disziplinen der Kunst hinausgegangen werden, um ein adäquates interdisziplinäres Bild der Kultur abseits der Klischees zeichnen zu können. Es wurden daher Leistungen in den formalen Künsten, in den Formalwissenschaften Mathematik und Logik sowie in jenen sozial- und naturwissenschaftlichen Disziplinen präsentiert, die mit formalen Methoden operieren.

Viele dieser Leistungen in den Formalwissenschaften und den formalen Künsten wurden und werden zwar von österreichischen und ungarischen BürgerInnen, aber zum Teil nicht in den Ländern Österreich und Ungarn produziert. Unter anderem äußert sich dieses Defizit heute in Österreich in Form einer wenig ausgeprägten Breitenkultur und Innovationsfreudigkeit im Bereich der Computer- und Kommunikationstechnologie, mit entsprechenden nachteiligen wirtschaftlichen Konsequenzen für das Land.

Deswegen haben Wolfgang Maass und Peter Weibel für das Katalogbuch "Jenseits von Kunst" wider den fortlaufenden Exodus der Wissenschaft in Österreich eine konkrete Maßnahme vorgeschlagen, nämlich die Gründung eines Wissenschaftszentrums im Bereich der Formalwissenschaften zur Pflege des wissenschaftlichen Austausches mit ÖsterreicherInnen im Ausland. In leicht veränderter Form ist dieser Vorschlag auch in dieses Buch unter dem Titel "Brain Drain oder Internationalisierung" eingegangen.

Sondierende Gespräche mit der Politik auf Landes- wie auf Bundesebene in bezug auf die Gründung eines solchen Wissenschaftszentrums haben zwar Aufmerksamkeit und Verständnis, aber keine praktischen Folgen gefunden. Daher entstand die Idee, glücklicherweise sofort unterstützt von Prof. Rainer Burkard, im Rahmen unserer eigenen Möglichkeiten dem "brain drain" entgegenzuwirken und eben auf dem Gebiet der Formalwissenschaften im Ausland erfolgreich wirkende Österreicher zu einem Symposium nach Graz einzuladen.

Nicht nur das Symposium selbst war ein entsprechender Erfolg mit großem Publikumsinteresse, sondern es gab auch individuelle Erfolge. Einer der eingeladenen Vortragenden, Wolfgang Woess, zum Zeitpunkt des Symposiums 1997 noch Professor an der Universität von Mailand, ist ab

Herbst 1999 ordentlicher Professor an der Technischen Universität Graz.

In den ersten Jahrzehnten dieses Jahrhunderts war Österreich das internationale Zentrum der Mathematischen Logik und der analytischen Philosophie, mit vielfältigen Beziehungen zur damaligen Kunst und Kultur in Österreich. Zu den Folgen der "Vertreibung der Vernunft" [Weibel und Stadler, 1993] gehört bekanntlich nicht nur das Fehlen jener Wissenschaftskultur, welche durch populärwissenschaftliche Darstellungen auch dem Nicht-Fachmann den Zugang zu den neuesten wissenschaftlichen Erkenntnissen ermöglicht, sondern darüber hinaus und vor allem das Fehlen einer Allianz zwischen Kunst und Wissenschaft, zu schweigen von einer naturwissenschaftlichen Fundierung der Kultur. Die Kunst der Neuzeit ist in Wirklichkeit ebenso theorieabhängig wie die Naturwissenschaft der Neuzeit. Wie in dieser herrscht das Primat der Theorie vor der Erfahrung (siehe Maxwells Gleichung von 1873, welche auf die Existenz von elektromagnetischen Wellen hinwies, und ihrem empirischen Nachweis durch Heinrich Hertz 1888). Man könnte sogar sagen, die moderne Kunst ist eine Reaktion auf die maschinenbasierte industrielle Revolution wie die Nachmoderne eine Reaktion auf die informationsbasierte postindustrielle Revolution ist. Man muß das Denken offen halten, um auch in Bereiche vordringen zu können, die nicht direkt unserer Erfahrung zugänglich sind, und die Begriffsbildung verallgemeinern, um neue Erfahrungen, technisch vermittelt, neu einordnen zu können. In solch einem historischen Kontext ist Kultur ohne Wissenschaft nicht zu denken. Ganz im Gegenteil, es gehört zur Epistemologie der Epoche, daß Kunst und Wissenschaft einander bedingen, auch wenn das Wissen darum durch ideologische Umstände verloren ging. Dieses Buch möchte dazu beitragen, dieses zerschnittene Band zwischen Wissenschaft und Kunst wieder zu verknüpfen. Visuelle Kultur als die Gesamtheit der Formen visueller Repräsentationen in allen Medien unserer Lebenswelt kann heute auf eine analytische Denktradition und Fundierung nicht verzichten. Interdisziplinäre und interkulturelle Kompetenz bilden die Basis für zeitgemäße Bilder von Wissenschaft und Kultur. Ja, man könnte sogar sagen, daß sich das visuelle Denken der Kunst (z.B. Klee) und das formale Denken der Wissenschaft in der Moderne gegenseitig unterstützen. Die Komplementarität von logischem Denken und "visuellem Denken" (R. Arnheim, 1969) gehört offensichtlich zu den epistemologischen Bedingungen der Epoche. Der Kunst des visuellen Denkens wird daher eine Kunst des formalen Denkens zur Seite und gegenüber gestellt. Die Mathematik als formalste Wissenschaft hat auf die Künste seit jeher eine große

Faszination ausgeübt und in Form von Proportion, Reihe, Serie, Symmetrie schon immer als Modell für die interne Organisation der visuellen und plastischen Elemente eines Bildwerks oder einer Skulptur gedient. Mathematische Methoden und informationstheoretische Modelle stellen für die Kunst immer wieder eine bedeutende Inspirations- und Orientierungsquelle dar. Auch der Mathematiker denkt in abstrakten Bildern, bevor er das Ergebnis seines visuellen Denkens formalisiert. Die Schnittstelle zwischen visuellem und formalen Denken ist unscharf. In einer zeitgenössischen Theorie der visuellen Kultur kann also zwischen formalem und visuellem Denken nicht mehr scharf unterschieden werden. Die visuelle Kultur von heute begründet sich eindeutig auf Ergebnisse der Kunst des logischen bzw. formalen Denkens. Gerade in den letzten Jahren und Jahrzehnten haben Mathematik und Informatik unsere historischen Vorstellungen umgeworfen, ganz neue Theorien entwickelt und neue Forschungsfelder entdeckt, deren Kenntnis auch für die Kunst von Relevanz sein kann.

Neue Ergebnisse und Forschungsfragen der Mathematik und Theoretischen Informatik werden in diesem Band durch allgemeinverständliche Essays von Spitzenforschern der Öffentlichkeit vorgestellt. Zusätzlich werden Parallelen, Divergenzen und Möglichkeiten zur methodischen und problemzentrierten Zusammenarbeit zwischen wissenschaftlicher Forschung und moderner Kunst erörtert. Zu den überraschenden Parallelen, die sich hierbei ergeben, gehört, daß sich das Gesichtsfeld der Forschung in Mathematik und theoretischer Informatik in den vergangenen beiden Jahrzehnten drastisch erweitert hat und nunmehr – ähnlich wie die Gegenwartskunst – auch nichtdeterminierte offene Systeme und Prozesse umfaßt. Eine andere möglicherweise überraschende Perspektive, die sich aus diesem Band ergibt, ist das innerhalb der exakten Wissenschaften entstehende Wissen um prinzipielle Grenzen der Fähigkeiten von Computern, und das In-Frage-Stellen von herkömmlichen Denkkategorien zur Unterscheidung der Fähigkeiten von Mensch und Maschine. Auch dieser Aspekt eröffnet neue Möglichkeiten der Zusammenarbeit von Künstlern und Wissenschaftlern. Schließlich wird in diesem Band demonstriert, daß traditionelle Bewertungskriterien von Kreativität, wie bedeutend oder unbedeutend, oberflächlich oder tiefinnig, zufällig oder genial, nicht nur in der Gegenwartskunst, sondern auch in den formalen Wissenschaften ihren Sinn verloren haben. Mehrere Aufsätze in diesem Band zeigen, daß zum Beispiel Spieleereien mit einfachen formalen Objekten, wie Primzahlen oder Kreisen, nicht trennbar sind von wissenschaftlichen Schlüsselfragen, die besondere Rele-

vanz für unsere wirtschaftliche und technologische Zukunft haben.

Eine Reihe von Aufsätzen in diesen Band demonstrieren die vielfältigen Beziehungen zwischen formalem Denken und künstlerischer Arbeit. Nicht nur Peter Weibel zeigt in seinem Beitrag, wie bestimmte Ergebnisse und Theorien der Zahlentheorie schon immer die Praktiken und Forschungen in der Musik, Architektur und in der bildenden Kunst beeinflusst und fundiert haben, und daß in der Kunst des visuellen Denkens viele ästhetische interne Ordnungs- und Konstruktionsprinzipien wie Symmetrie, Harmonie, Proportion auf mathematische Vorstellungen basieren. Auch Wolfgang Woess und Raimund Seidel verweisen bei ihren Analysen von Zufalls- und Ordnungsprinzipien in Strukturen und Prozessen auf Parallelen zur Kunst. Ebenso werfen Herbert Edelsbrunners Ergebnisse ein Licht auf die Form und Funktion ästhetischer Objekte und Prozesse. Der bedeutende österreichische Zahlentheoretiker Wolfgang Schmidt spürt dem Ineinanderspiel von Ordnung und Chaos in der Zahlentheorie nach, das überraschende Analogien zur Kunst ergibt, und erläutert die Bedeutung von zunächst rein theoretischen Problemen für Schlüsselfragen der Kryptographie, die unter anderem die Grundlagen für die Abwicklung von Geschäftsvorgängen über das Internet schaffen.

Gottfried Tinhofer stellt einen besonders wichtigen Grundbegriff des formalen Denkens, den des Graphens, mit vielfältigen Beispielen vor, und demonstriert die Vielfalt und Ambivalenz des formalen Denkens zwischen reiner Mathematik räumlicher Vorstellungen, wichtigen Anwendungsproblemen wie dem Entwurf von Computer-Bausteinen und der Zuverlässigkeit des Internets, die von besonderer wirtschaftlicher Bedeutung sind, und formalen Fragen der Ästhetik. Rainer Burkard demonstriert in einem kurzen Streifzug durch die kombinatorische Optimierung, weshalb das wohlbekannte "Heiratsproblem" verwandt ist mit wichtigen Anwendungsproblemen wie zum Beispiel der Nachrichtenübermittlung über Satelliten, und stellt konkrete Beispiele der oft diskutierten Klasse der sogenannten "NP-vollständigen" Probleme vor, bei denen das Auffinden einer optimalen Lösung vermutlich so aufwendig ist, daß selbst die schnellsten Rechner hier an ihre Grenzen stoßen. Raimund Seidel demonstriert wie man zufällige Ergebnisse nutzen kann, um schwierige mathematische Probleme zu lösen (zum Beispiel dem Nachprüfen, ob eine vorgegebene große Zahl eine Primzahl ist), bei denen die üblichen (d.h. deterministischen) Algorithmen versagen. Herbert Edelsbrunner zeigt wie man bei einfachen Spielereien mit Kreisen und Kugeln auf ungelöste Probleme der Mathematik stößt, und gleich-

zeitig auf Grundprinzipien der Bausteine des Lebens, Proteine, wo Form in Funktion umschlägt. Wolfgang Woess untersucht Ordnungen, die sich im scheinbaren Chaos von Irrfahrten ergeben, und erläutert "nebenbei" wichtige Grundbegriffe und Probleme der reinen Mathematik. Im Beitrag von Wolfgang Maass wird das formale Denken gewissermaßen auf sich selbst angewendet. Es werden Ergebnisse über "Rechnen" in biologischen Neuronsystemen - wie zum Beispiel unserem Gehirn - vorgestellt, und es wird der Frage nachgegangen, weshalb es so schwierig ist, einen Computer zum Denken zu bringen.

Der Essay von Maass und Weibel erläutert die Problematik des brain drain am Beispiel der Situation der Mathematik und Theoretischen Informatik in Österreich, und entwickelt einen Vorschlag zur Verkehrung dieser bedenklichen Entwicklung in einen positiven Faktor bei der Internationalisierung dieser Wissenschaften in Österreich.

Im letzten Kapitel "Links" findet der Leser einige Webpages, auf denen weitere Informationen zu in diesem Band angeschnittenen Themen, sowie zu Studiumsmöglichkeiten in den betreffenden wissenschaftlichen Disziplinen erhältlich sind.

Wolfgang Maass und Peter Weibel

Anmerkungen

- ¹ Die Ausstellung "Jenseits von Kunst" in Budapest, Graz und Antwerpen, konzipiert und kuratiert von Peter Weibel, organisiert von Nadja Rottner, Christa Steinle (Graz), Dóra Hegyi, Krisztina Szipöcs (Budapest) und wissenschaftlich koordiniert von Michael Stöltzner (Wien), György Kampis (Budapest), wurde zuerst, unterstützt vom wissenschaftlichen Beirat László Beke und Miklós Peternák, im Ludwig Museum Budapest vom 17. 10. - 24. 11. 1996 gezeigt, wo sie von Elisabeth Mach, Leiterin des Österreichischen Kulturinstituts Budapest und Katalin Néray, Direktorin des Ludwig Museums Budapest, als Beitrag zum Österreichischen Millennium und ungarischen Millecentenarium initiiert worden ist.

Literatur

- [Weibel und Stadler, 1993] Weibel, P., und Stadler, F. (1993). *Vertreibung der Vernunft: The Cultural Exodus from Austria*. Löcker Verlag, Wien. Erweiterte Neuauflage Springer Verlag, Wien New York, 1995.

Kuriosa der Zahlenkunde und die Kunst
– kurz gefaßt und leicht faßlich dargestellt
Peter Weibel

Die Lust an der Multiplikation der Zahl.
Die Trunkenheit ist Zahl. Die Zahl ist im Individuum.

Charles Baudelaire, 1869

Ich blicke euch an, ihr Zahlen,
und ihr erscheint mir verkleidet als Tiere, in euren Fellen,
die Arme gestützt auf ausgerissene Eichen.
Ihr gebt mir Einheit zwischen
der Schlangenbewegung
des Weltalls und dem Tanz
der Waagschalen
ihr erlaubt mir,
die Jahrhunderte zu verstehen als
Zähne eines schnellen Gelächters.
Meine Augen sind aufgerissen, begierig
ALLES zu wissen, was ICH ist,
wann der Teiler zur Eins schrumpft.

Velimir Chlebnikov, vor 1913

Zahlen in Expansion sind ebenso wirklich
wie Tiere, die sich vermehren
Die Zahlen vermehren sich wie die Tiere
Die Tiere vermehren sich wie die Zahlen
In Bewegung sind die mathematischen und
die physikalischen Gesetze wirkliche Tiere
Die Zahlen sind lebende Tiere

Mario Merz, 1970

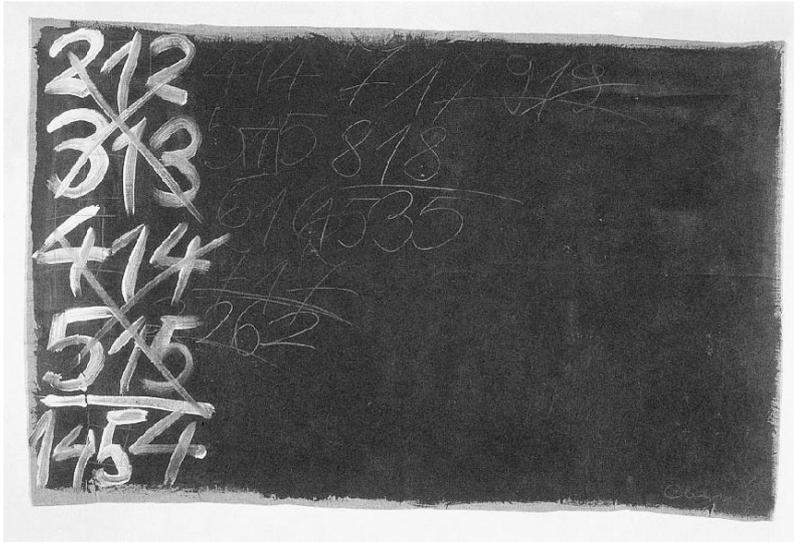


Abbildung 1.1. Oberhuber, Rechnung (Calculation), 1952

Symbole, Persönlichkeiten, Zahlen

Man muß kein fanatischer Numerologe sein, um daran zu glauben, daß einige Zahlen "Persönlichkeit" haben. Wer wird nicht darein übereinstimmen, daß die Zahl 13 einen dubiosen Charakter hat, der Unglück verspricht, schwärzer als der Schatten eines schwarzen Rabens. Die Zahl 7 hat seit alters die Reputation, magisch zu sein: die 7 Schöpfungstage, die 7 Todsünden, die 7 Säulen der Weisheit, der Agent 007, die 7 Weltwunder, die 7 Zwerge, das Buch mit 7 Siegeln, die 7 Planeten, welche die Namen unserer 7 Wochentage lieferten, das verflixte 7. Jahr, die 7 Leben der Katze, die 7 hebräischen Namen Gottes usw. Erst Galileo hat uns erklärt, was schon viel früher bekannt war: die hohe Wahrscheinlichkeit, mit zwei Würfeln eine 7 zu würfeln. Die Zahl 3 gilt nicht erst seit den 3 Grazien oder der Trinität von Vater, Sohn und Heiliger Geist als heilig, sondern schon Pythagoras nannte sie die vollkommene Zahl, da sie den Anfang, die Mitte und das Ende ausdrücke. Selbstverständlich hat man auch einige Beziehungen zwischen Erotik und Mathematik festgestellt, und selbstverständlich in Wien [Hug-Hellmuth, 1916]. So wurden die geraden Zahlen mit weiblichen und die ungeraden Zahlen mit männlichen Eigenschaften versehen, und vice versa, je nach Kultur und Er-

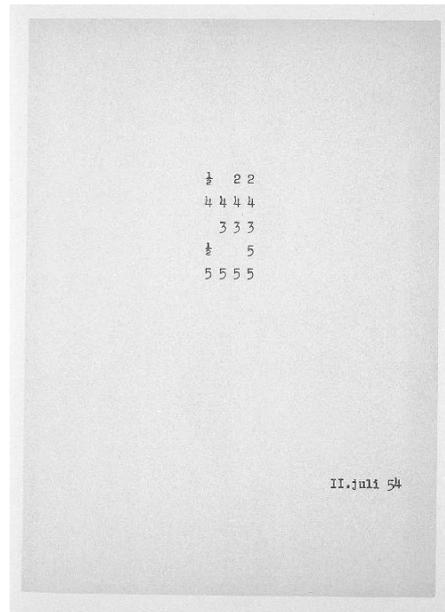


Abbildung 1.2. Gerhard Rühm, Zahlengedicht, 1954

fahrung. Von den Ägyptern bis zu Aristoteles galt die Einheit als Mutter und Ursprung aller Zahlen, ohne selbst eine Zahl zu sein. 1839 veröffentlichte der französische Mathematiker M. Vincent in einer Arbeit über den Ursprung der Zahlen, die auf den Pythagoräern und Boethius fußte, die Vermutung, daß die ersten 3 Ziffern die Geschlechter und deren Vereinigung versinnbildlichen, indem sie als charakteristische Körperteile der Frau und des Mannes und dann die Drei als deren Vereinigung gedeutet worden seien. Die Neun deutete Vincent als Thyphallus, als Zeichen der männlichen Kraft. Denn die Neun ist die Quadratzahl der Drei, die ihrerseits die Vereinigung des männlichen und weiblichen Prinzips darstellt. Das Quadrat oder die zweite Potenz wurde bei den Griechen kurzweg Potenz genannt. Na also! Neben dieser Hypothese, die Entstehung der Namen und Zeichen der Zahlen mit dem sexuellen Leben in Zusammenhang zu bringen, erfreute sich eine andere einer großen Verehrung, nämlich die Zahlen mit Lastern und Tugenden zu verbinden.

So schreibt im 15. Jahrhundert Luca Pacioli in seiner "Divina Proportion": "Die vollkommenen Zahlen endigen abwechselnd mit 6 und 8 und

können eine andere Randziffer nicht haben, denn die Traurigen leben ordnungslos, die Guten und Vollkommenen bewahren immer die vorgeschriebene Ordnung“. Die Triaden 4, 5, 6 und 7, 8, 9 in der Folge von 1, 2, 3 repräsentierten demnach Güte, Gerechtigkeit, Schönheit und Größe, Gesundheit, Kraft.

Die Vier, für die Pythagoräer der “Schlüssel der Natur“, sei als Beispiel der naturwissenschaftlichen, kosmologischen Symbolik der Zahlen erwähnt, die 4 Elemente Erde, Feuer, Wasser, Luft.

Die Griechen wie die Chinesen um das Jahr 1120 v. Chr. haben laut Montucla’s “Histoire des mathématiques“ das Weltall aus den ersten vier geraden und den ersten vier ungeraden Zahlen zusammengesetzt. Die ersten vier ungeraden Zahlen stellen dabei die reinen und himmlischen Elemente dar, die geraden entsprechen denselben Elementen mit irdischer Unreinheit verbunden. Das Weltall, die Verbindung aller himmlischen und irdischen Elemente, wurde also durch die Zahl 36 dargestellt, das heißt die Summe dieser 8 Zahlen [Turing, 1952], [Thompson, 1966], [Thom, 1972], [Thom, 1974], [Cook, 1979], [Mandelbrot, 1983], [Frängsmyr et al., 1984], [Ghyka, 1977], [Coleman und Holmes, 1988]. Die meisten dieser Zahlen haben also “Persönlichkeit“, “Charaktereigenschaften“, die von außen, von Menschen, von der Ideologie, von der Erfahrung an sie herangetragen werden.¹ Doch gibt es auch Zahlen, die von innen her interessante Eigenschaften haben, Zahlen, die auf inhärente Weise interessante “Persönlichkeiten“ sind, wie zum Beispiel die Primzahlen.

Primzahlen

Für mich haben die Primzahlen am meisten “Persönlichkeit“, denn sie sind ein Paradebeispiel für jene Kuriosa der Zahlenkunde, die jenseits ihrer Merk-Würdigkeiten, die der Aberglaube so gerne unter seine Fittiche nimmt, zu fundamentalen Einsichten in die Natur der Zahlen und zu komplexen Theoremen der Zahlentheorie geführt haben² [Stewart, 1995], [Conway und Guy, 1996], [Derlin, 1998], [Hoffman, 1998], [Singh, 1998], [Basieux, 1999]. Eine natürliche Zahl p heißt Primzahl, wenn p ungleich 1 ist und nur die (trivialen, positiven) Teiler 1 und p hat. Primzahlen sind also die positiven ganzen Zahlen, die natürlichen Zahlen, die nur durch sich selbst oder 1 teilbar sind, also unzerlegbare Zahlen. Primzahlen sind $\dots, 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, \dots, 229, \dots, 5\,693, \dots, 199\,999, \dots$ und so weiter. Euklid hat bereits vor 2300 Jahren gezeigt, daß es keine obere Schranke für die Prim-

2	227	509	829	1171	1523	1879	2269	2659	3019	3433	98999	99563	100129	100693	101267	101749	102241	102829	103471	104033	104593
3	229	521	839	1181	1531	1889	2273	2663	3023	3449	99013	99577	100151	100715	101299	101781	102253	102841	103483	104045	104607
5	233	523	853	1187	1543	1901	2281	2671	3037	3467	99027	99591	100169	100733	101317	101799	102271	102859	103501	104063	104625
7	239	541	857	1193	1549	1907	2287	2677	3041	3461	99041	99605	100183	100747	101331	101813	102285	102873	103515	104077	104639
11	241	547	859	1201	1553	1913	2293	2683	3049	3463	99055	99619	100199	100763	101307	101789	102261	102849	103491	104053	104615
13	251	557	863	1213	1559	1931	2297	2687	3061	3467	99079	99633	100219	100783	101327	101809	102281	102869	103511	104073	104635
17	257	563	877	1217	1567	1933	2309	2689	3067	3469	99083	99637	100227	100791	101335	101817	102289	102877	103519	104081	104643
19	263	569	881	1223	1571	1949	2311	2693	3079	3491	99089	99641	100235	100799	101343	101825	102297	102885	103527	104089	104651
23	269	571	883	1229	1579	1951	2333	2699	3083	3499	99103	99645	100243	100807	101351	101833	102303	102891	103533	104093	104659
29	271	577	887	1231	1583	1973	2339	2707	3089	3511	99109	99649	100251	100815	101359	101841	102311	102899	103541	104095	104663
31	277	587	907	1237	1597	1979	2341	2711	3109	3517	99115	99653	100259	100823	101367	101849	102319	102907	103549	104101	104669
37	281	593	911	1249	1601	1987	2347	2713	3119	3527	99121	99657	100267	100831	101375	101857	102327	102915	103553	104107	104675
41	283	599	919	1259	1607	1993	2351	2719	3121	3529	99127	99661	100275	100839	101383	101865	102335	102923	103561	104113	104681
43	293	601	929	1277	1609	1997	2357	2729	3137	3533	99133	99665	100283	100847	101391	101873	102343	102931	103569	104119	104683
47	307	607	937	1279	1613	1999	2371	2731	3163	3539	99139	99669	100291	100855	101399	101881	102351	102939	103577	104125	104689
53	311	613	941	1283	1619	2003	2377	2741	3167	3541	99145	99673	100299	100863	101407	101889	102359	102947	103585	104131	104691
59	313	617	947	1289	1621	2011	2381	2749	3169	3547	99151	99677	100307	100871	101415	101897	102367	102955	103593	104137	104693
61	317	619	953	1291	1627	2017	2383	2753	3181	3557	99157	99681	100315	100879	101423	101905	102375	102963	103601	104143	104695
67	331	631	967	1297	1637	2027	2389	2767	3187	3559	99163	99685	100323	100887	101431	101913	102383	102971	103609	104149	104697
71	337	641	971	1301	1657	2029	2393	2777	3191	3571	99169	99689	100331	100895	101439	101921	102391	102979	103617	104155	104701
73	347	643	977	1303	1663	2039	2399	2789	3203	3581	99175	99693	100339	100903	101447	101929	102401	102987	103625	104161	104703
79	349	647	983	1307	1667	2053	2411	2791	3209	3583	99181	99697	100347	100911	101455	101937	102409	102995	103633	104167	104705
83	353	653	991	1319	1669	2061	2417	2797	3217	3593	99187	99701	100355	100919	101463	101945	102417	103001	103641	104173	104707
89	359	659	997	1321	1693	2069	2423	2801	3221	3607	99193	99705	100363	100927	101471	101953	102425	103007	103649	104179	104713
97	367	661	1009	1327	1697	2081	2437	2803	3229	3613	99199	99709	100371	100935	101479	101961	102433	103015	103657	104185	104719
101	373	673	1013	1361	1699	2083	2441	2819	3251	3617	99205	99713	100379	100943	101487	101969	102441	103021	103665	104191	104721
103	379	677	1019	1367	1709	2087	2447	2823	3253	3623	99211	99717	100387	100951	101495	101977	102449	103029	103673	104197	104723
107	383	683	1021	1373	1721	2089	2459	2837	3257	3631	99217	99721	100395	100959	101503	101989	102457	103037	103681	104203	104729
109	389	691	1031	1381	1723	2099	2467	2843	3259	3637	99223	99725	100403	100967	101511	101997	102465	103045	103689	104209	104731
113	397	701	1033	1399	1733	2111	2473	2851	3271	3643	99229	99729	100411	100975	101519	102005	102473	103053	103697	104215	104735
127	401	709	1039	1409	1741	2113	2477	2857	3279	3659	99235	99733	100419	100983	101527	102013	102481	103061	103705	104221	104741
131	409	719	1049	1423	1747	2129	2503	2861	3301	3671	99241	99737	100427	100991	101535	102021	102489	103069	103713	104227	104743
137	419	727	1051	1427	1753	2131	2521	2879	3307	3673	99247	99741	100435	101003	101543	102029	102497	103077	103721	104233	104745
139	421	733	1061	1429	1759	2137	2531	2887	3313	3677	99253	99745	100443	101011	101551	102037	102505	103085	103729	104239	104747
149	431	739	1063	1433	1777	2141	2539	2897	3319	3691	99259	99749	100451	101019	101559	102041	102513	103093	103737	104245	104751
151	433	743	1069	1439	1783	2143	2543	2903	3323	3697	99265	99753	100459	101027	101567	102049	102521	103101	103745	104251	104753
157	439	751	1087	1447	1787	2153	2549	2909	3329	3701	99271	99757	100467	101035	101575	102053	102529	103109	103753	104257	104755
163	443	757	1091	1451	1789	2161	2551	2917	3331	3709	99277	99761	100475	101043	101583	102061	102537	103117	103761	104263	104757
167	449	761	1093	1453	1801	2179	2557	2927	3343	3719	99283	99765	100483	101051	101591	102069	102541	103125	103769	104269	104759
173	457	769	1097	1459	1811	2203	2579	2939	3347	3727	99289	99769	100491	101059	101603	102077	102549	103133	103773	104275	104761
179	461	773	1103	1471	1823	2207	2591	2953	3359	3733	99295	99773	100499	101067	101611	102081	102557	103141	103781	104281	104763
181	463	787	1109	1481	1831	2213	2593	2957	3361	3739	99301	99777	100507	101075	101619	102089	102561	103149	103789	104287	104765
191	467	797	1117	1483	1847	2221	2609	2963	3371	3761	99307	99781	100515	101083	101627	102097	102569	103157	103797	104293	104767
193	479	809	1123	1487	1861	2237	2617	2969	3373	3767	99313	99785	100523	101091	101635	102105	102577	103165	103801	104301	104771
197	487	811	1129	1489	1867	2239	2621	2971	3389	3769	99319	99789	100531	101103	101643	102113	102585	103173	103809	104307	104773
199	491	821	1151	1493	1871	2243	2623	2999	3391	3779	99325	99793	100539	101111	101651	102121	102595	103181	103817	104313	104775
211	499	823	1153	1499	1873	2251	2647	3001	3407	3793	99331	99797	100547	101119	101659	102129	102605	103189	103821	104319	104777
223	503	827	1163	1511	1877	2267	2657	3011	3413	3797	99337	99801	100555	101127	101667	102137	102615	103197	103829	104325	104779

Abbildung 1.3. Ecke Bonk, Die ersten 10 000 Primzahlen, 1999 (Detail)

zahlen gibt. Es gibt nur die größte bekannte Primzahl.

Es gibt ganze Bücher, die nur aus der Auflistung aller bisher bekannten Primzahlen bestehen. Ein 8-bändiges Werk, das alle Primzahlen, wenn auch fehlerhaft, von 2 bis 100 330 201 aufzählt, das sind 5 761 456 Primzahlen, gibt es in Wien von Kulik, der fast sein ganzes Leben darüber verbracht hat. Vor einiger Zeit haben C. L. Baker und F. J. Gruenberger von der Rand Corporation auf einem Computer die ersten 6 Millionen Primzahlen, von 2 bis 104 395 289, berechnet. Die größte gegenwärtig bekannte Primzahl ist die am 1. Juni 1999 entdeckte Primzahl $2^{6972593} - 1$.

Während es Methoden gibt, zu testen, ob eine gegebene Zahl eine Primzahl ist, gibt es keine Formel, die genau alle Primzahlen erzeugt. Allerdings gibt es Ausdrücke, zum Beispiel durch die Robinson-Formel, mit der man bei weitem nicht alle, aber zumindest eine Gruppe innerhalb der Primzahlen erzeugen kann: $R(k, n) = 2^n k + 1$. Für bestimmte Werte von k und n erzeugt diese Formel Primzahlen. Für $k = 5$ und $n = 1947$ erhalten wir die größte bekannte Robinson-Primzahl, die 586 Ziffern hat. Eine zweite For-

mel, die einige Primzahlen erzeugt, stammt von Fermat: $2^{2^n} + 1$. Fermat glaubte, diese Formel würde für alle Werte von n eine Primzahl erzeugen, doch wurden bis zum Jahr 1980 nur 5 Primzahlen gemäß dieser Formel entdeckt, nämlich 3, 5, 17, 257 und 65 537.

Im Alter von 19 Jahren hat Carl Friedrich Gauß 1798 eine interessante Entdeckung gemacht, um eine Schwierigkeit bei der Konstruktion von regelmäßigen Polygonen von n Seiten zu beheben, wo n eine Primzahl ist, also bei der Konstruktion von Heptagon, 11-gon, 17-gon, usw. Er fand heraus, daß so eine Konstruktion nur möglich ist, wenn die Anzahl der Seiten des regelmäßigen Polygons eine Fermatsche Primzahl ist. Eine Euklidische Konstruktion eines regelmäßigen Polygons mit einer primen Seitenanzahl ist also nur dann möglich, wenn die Anzahl der Seiten 3, 5, 17, 257 oder 65 537 ist. O. Hermes verbrachte 10 Jahre, dieses 65 537-gon zu konstruieren. Sein Manuskript liegt in einer großen Schachtel in der Universität Göttingen.

Andere Formeln zur Erzeugung begrenzter Serien von Primzahlen sind Eulers Polynom $x^2 - x + 41$, das 40 verschiedene Primzahlen für $x = 1, 2, 3, \dots, 40$ ergibt. Legendre's Polynom $2x^2 + 29$ von 1798 erzeugt 29 Primzahlen für $x = 0, 1, 2, \dots, 28$. Untersucht wurden auch Serien von Primzahlen mit einer gleichbleibenden Differenz, zum Beispiel 11, 17, 23, 29, wo die Differenz stets 6 ist. Eine Serie von 10 Primzahlen mit der gemeinsamen Differenz von 210 beginnt mit 199.

Bis heute kennt man jedoch noch kein Verfahren, wie man zu irgendeiner Primzahl ihren unmittelbaren Nachfolger angeben kann. Nur eines wissen wir seit Euklid mit Sicherheit: "Es gibt mehr Primzahlen als jede vorgegebene Anzahl von Primzahlen", die Anzahl der Primzahlen ist also unendlich. Ein Hauch von Unendlichkeit umweht die Primzahlen, doch sie zeigen uns bereits die Unendlichkeit nicht als endlosen Brei, sondern gegliedert. Wie es später Georg Cantor (1845-1918) mit den Begriffen Mächtigkeit und Überabzählbarkeit der unendlichen Zahlenmengen gelungen ist. Der Struktur dieser gegliederten Unendlichkeit verdankt sich auch das noch ungelöste Problem, ob es endlich oder unendlich viele Primzahlzwillinge (wie 17 und 19) gibt.

Das Symbol der Unendlichkeit: die Spirale. Im Hauch der Unendlichkeit der Primzahlen gewinnt bereits das Unendlichkeitssymbol Spirale Kontur. Teilbarkeit und Unteilbarkeit des Lebendigen als Verschränkungen, welche sich in der gleichbleibenden Differenz bei bestimmten mathematischen Serien der Primzahlen spiegeln, kulminieren in der vollkommenen Zahl, wel-

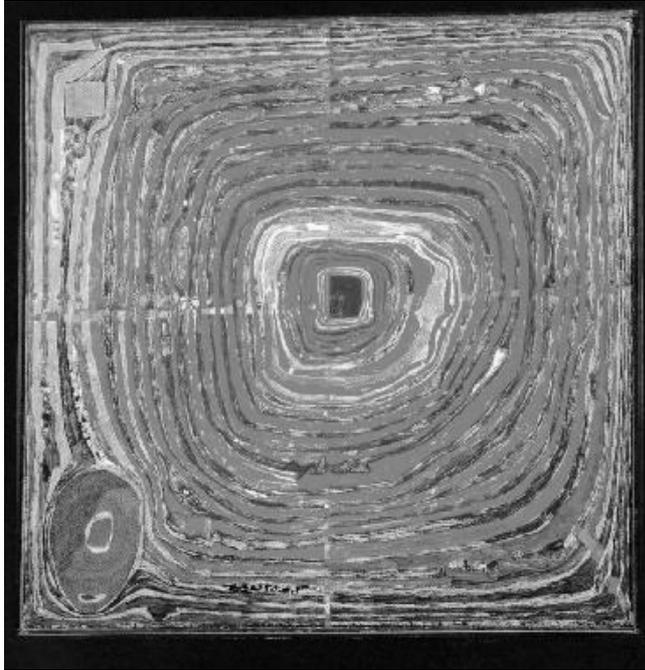


Abbildung 1.4. Friedensreich Hundertwasser, Der große Weg, 1955. ©1998 Gruener Janura AG, Glarus Schweiz

che gleich mit der Summe ihrer Teiler ist. Was ist das für eine wunderbare Identität, die teilbar und geteilt ist, in der Summe ihrer Teiler aber wieder aufersteht? Ist es eine gleichsam mythische Identität ähnlich der des ägyptischen Totengottes Osiris, der als Vegetationsgott zugleich auch für die Auferstehung bürgt? Blaise Pascal, der Mathematiker und religiöse Denker, hat eine Spirale konstruiert, deren Tangentialwinkel konstant ist und deren Mittelpunkt im Unendlichen liegt. Normalerweise hat nur der Kreis einen konstanten Tangentialwinkel und es gehört zum Wesen der Spirale, daß ihr Tangentialwinkel nicht konstant ist. Diese paradoxe, nur mathematisch konstruierbare Spirale hat Pascal als Modell für den Beweis der Unsterblichkeit der Seele genommen: eadem mutata resurgo. Auch nach ihrer Verwandlung (der Tod) wiederaufersteht sie als dieselbe. Siehe Abbildung 1.4 auf Seite 27 und die Farbabbildungen 1.19 und 1.20 (Seite 137, 138).

Die Goldbachsche Vermutung

Da in der Urgeschichte der Mathematik der Gegensatz von geraden und ungeraden Zahlen so eine große Rolle spielte, ja sogar von erheblicher produktiver Kraft war, kommt den Primzahlen schließlich noch eine besonders rätselhafte Funktion zu. Der Königsberger Mathematiker Christian Goldbach (1690-1764) hat nämlich 1742 in einem Brief an Euler die Vermutung ausgesprochen, daß jede gerade Zahl (größer als 2) als Summe zweier Primzahlen, zumeist sogar mehrfach, dargestellt werden kann:

$$\begin{aligned} 88 &= 5 + 83 = 29 + 59 = 41 + 47 \\ 92 &= 3 + 89 = 13 + 79 = 19 + 73 = 31 + 61 \end{aligned}$$

Aber genügen auch in den fernsten Zahlenregionen nur zwei Primzahlen, um durch Addition alle geraden Zahlen zu bilden? Euler hat Goldbach geantwortet: "Daß aber jeder numerus par eine Summe duorum primorum sei, halte ich für ein ganz gewisses Theorema, ungeacht ich dasselbe nicht demonstrieren kann". Der Beweis für Goldbachs Vermutung ist immer noch ausständig. Am nächsten kommt I. M. Winogradows Ergebnis, das besagt, daß es eine ganze Zahl N gibt, so daß jede ganze Zahl größer als N als die Summe von nicht mehr als 3 Primzahlen repräsentiert werden kann, wenn n ungerade ist, und von 4 Primzahlen, wenn n geradzahlig ist. Vier Primzahlen reichen also stets aus, um jede gerade Zahl zu bilden. Nach Chen ist jede hinreichend große gerade Zahl als Summe von $p + q$ darstellbar, wobei p eine Primzahl ist und q Produkt von höchstens zwei Primzahlen.

Vollkommene Zahlen und Mersennesche Primzahlen

Eine ähnlich paradoxe Rolle spielen die Mersenneschen Zahlen, die eine besondere Art von Primzahlen sind und als solche keine ganzzahligen Teiler *ungleich* 1 haben, und die Vollkommenen Zahlen, welche zu den Mersenneschen Zahlen in einer ähnlichen Beziehung stehen wie die geraden Zahlen zu den Primzahlen.

Rekapitulieren wir: Primzahlen p sind nur durch sich selbst oder durch 1 teilbar, haben also keine (ganzzahligen) Teiler zwischen 1 und p bzw. sind unteilbar. Vollkommene Zahlen sind hingegen gerade diejenigen Zahlen m , für die die Summe aller kleineren positiven Teiler von m die Zahl m selbst wieder ergeben. Es zeigt sich nun, daß jede Vollkommene Zahl genau eine Mersennesche Primzahl als Teiler besitzt und umgekehrt, daß jede Mersennesche Primzahl genau eine Vollkommene Zahl festlegt. Obwohl in ihren

Teilbarkeitseigenschaften so gegensätzlich, besteht also eine eindeutige Zuordnung zwischen den beiden Zahlenmengen. Jeder Mersenneschen Primzahl entspricht eine Vollkommene Zahl und es gibt keine bekannte Vollkommene Zahl, die nicht mit einer Mersenneschen Primzahl korrespondiert. Wie ist das möglich?

Es gibt unter den ersten 30 000 000 Zahlen nur vier Vollkommene Zahlen:

$$\begin{aligned}
 6 &= 1 + 2 + 3 && \text{wobei 6 durch 1, 2 und 3 teilbar ist.} \\
 28 &= 1 + 2 + 4 + 7 + 14 \\
 496 &= 1 + 2 + 4 + 8 + 16 + 31 + 62 + 124 + 248 \\
 8128 &= 1 + 2 + 4 + 8 + 16 + 32 + 64 + 127 + 254 + 508 + 1016 + \\
 &\quad 2032 + 4064
 \end{aligned}$$

Die nächste Vollkommene Zahl ist 33 550 336, welche erst 1460 festgestellt wurde, während man die ersten vier bereits seit 2000 Jahren kannte und mit Euklid die Vollkommenen Zahlen nannte. Bis heute kennen wir 38 Vollkommene Zahlen und jede davon ist geradzahlig. Man kennt keine ungeraden Vollkommenen Zahlen. Vollkommene Zahlen haben interessante Eigenschaften, zum Beispiel daß sie, außer 6, als letzte Ziffernsumme 1 haben, wenn man in einem iterierten Prozeß immer wieder die Summe ihrer Ziffern bildet.

$$496 : 4 + 9 + 6 = 19, \quad 1 + 9 = 10, \quad 1 + 0 = 1 .$$

Jede vollkommene Zahl, außer 6, kann auch als Summe der Kuben aller ungeraden Zahlen bis zu einer gewissen Stelle geschrieben werden, zum Beispiel

$$\begin{aligned}
 28 &= 1^3 + 3^3 \\
 496 &= 1^3 + 3^3 + 5^3 + 7^3 .
 \end{aligned}$$

Sie haben vielleicht auch bemerkt, daß die Teiler der vier vollkommenen Zahlen mit 1 beginnen und sich zunächst verdoppeln, bis an einer Stelle das Doppelte des vorhergehenden Teilers um 1 vermindert erscheint und von da an sich die Teiler wieder verdoppeln. Der kritische Übergang erfolgte

	für	bei den Teilern
6	2	3
28	4	7
496	16	31
8128	64	127

Wenn man diese beiden Teiler miteinander multipliziert, erhält man wiederum die jeweilige Vollkommene Zahl. Bei diesen Teilern kann man auch ersehen, daß der erste offensichtlich eine Potenz von 2, der zweite die um 1 verminderte nächsthöhere Potenz von 2 ist. V läßt sich also definieren als die Multiplikation von zwei Potenzen von zwei, wobei von letzterer 1 abgezogen wird.

$$\begin{aligned} V_1 &= 6 = 2^1 \cdot (2^2 - 1) \\ V_2 &= 28 = 2^2 \cdot (2^3 - 1) \\ V_3 &= 496 = 2^4 \cdot (2^5 - 1) \\ V_4 &= 8128 = 2^6 \cdot (2^7 - 1) \end{aligned}$$

Daraus ergibt sich die allgemeine Struktur der geraden Vollkommenen Zahlen $V_i = 2^{n-1} \cdot (2^n - 1)$, die Euklid um 300 v. Chr. und Euler 1750 gefunden haben, die aber nur dann Vollkommene Zahlen erzeugt, wenn $2^n - 1$ nicht teilbar ist, also eine Primzahl. Zum Beispiel würde auch 120 dieser Struktur genügen, aber dennoch ist sie keine Vollkommene Zahl: $120 = 2^3(2^4 - 1) = 8 \cdot 15$, aber 15 ist keine Primzahl. Nur wenn $2^n - 1$ nicht teilbar, also eine Primzahl ist, liefert die Formel $2^{n-1} \cdot (2^n - 1)$ eine Vollkommene Zahl, diese perfekt teilbare Zahl. Der Satz von Euklid-Euler lautet also:

Eine gerade Zahl V ist genau dann vollkommen, wenn V von folgender Struktur ist:

$$V = 2^{n-1}(2^n - 1)$$

wobei $2^n - 1$ eine Primzahl ist.

In dieser paradoxen Definition der perfekten Teilbarkeit mit Hilfe einer teilweisen Nichtteilbarkeit ist auch die Dialektik von Gerade und Ungerade mit eingeschlossen. Denn es gibt nur gerade Vollkommene Zahlen dieser Bauart, während die Primzahlen ($2^n - 1$), welche für deren Definition benötigt werden, alle ungerade sind.

Primzahlen von der Form $2^n - 1$ sind ganz besondere Primzahlen, nämlich die sogenannten Mersenneschen Primzahlen M_i , zur Erinnerung an Pater Marin Mersenne (1588-1648) so genannt, der 1644 ankündigte, einige neue vollkommene Zahlen entdeckt zu haben, allerdings weder richtig noch vollständig, wobei er aber feststellte, daß $2^n - 1$ Primzahlen ergibt. Bis heute wissen wir dies für 38 Werte von n , angefangen mit $n = 2, 3, 5, 7, 13, 17, 19, 31, 61, 89, 107, 127, 521, 607, 1279, 2039, 2281,$

3 217, 4 253, 4 423, 9 689, 9 941, 11 213, 19 937. Die 10 größten der 38 gegenwärtig bekannten Werte von n für die $2^n - 1$ eine Primzahl ergibt sind $n = 110503, 132049, 216091, 756839, 859433, 1257787, 1398269, 2976221, 3021377, 6972593$. Für den letztgenannten Wert von n ergibt $2^n - 1$ die größte gegenwärtig bekannte Primzahl. Dies ist eine Zahl, die in der üblichen Dezimalschreibweise über 2 Millionen Stellen hat (siehe <http://www.utm.edu/research/primelargest.html>).

Es gibt keine allgemein gültige Methoden, um Mersenne Primzahlen zu erzeugen. Notwendig ist auf jeden Fall, daß n selbst eine Primzahl ist. Naheliegender wäre für die Mersennesche Primzahlen die Behauptung: $2^n - 1$ ist stets eine Primzahl, wenn n eine Primzahl ist. Doch wir haben gesehen, daß bestimmte Primzahlen (wie zum Beispiel 23) nicht in der Liste von denjenigen Zahlen aufscheinen, welche als Werte für n bei der Formel $2^n - 1$ eine Mersennesche Primzahl ergeben. Daher muß man für jede Primzahl n separat nachprüfen, ob $2^n - 1$ wieder eine Primzahl ergibt. Die Art, wie dieser Nachweis jeweils geschehen ist, spiegelt interessante Wandlungen in der Arbeitsweise der Zahlentheorie wider. Während bis vor wenigen Jahrzehnten Mathematiker diesen Nachweis mühsam mit Papier und Bleistift führen mußten, ermöglichte die Erfindung elektronischer Rechenmaschinen, zu viel größeren Werten von n vorzustoßen. Anfänglich hat man dazu überwiegend Supercomputer in Rechenzentren verwendet. Dagegen wurden die letzten vier Mersenneschen Primzahlen (für $n = 1398269, 2976221, 3021377, 6972593$) durch eine mittels Internet koordinierte Zusammenarbeit von mehreren tausend Hobbyforschern ermöglicht, die die nicht benötigte Rechenzeit auf ihrem PC hierfür zur Verfügung gestellt haben. Auf der Webpage <http://www.mersenne.org/prime.htm> kann sich jedermann über den gegenwärtigen Stand dieser Suche nach immer größeren Primzahlen informieren und kann dort auch erfahren, wie er sich selbst an der Aktion GIMPS (=The Great Internet Mersenne Prime Search) beteiligen kann.

Die Vollkommenen Zahlen V_6 bis V_8 wurden im 16. Jahrhundert entdeckt, die V_9 erst am Ende des 19. Jahrhunderts. Ob es eine größte Vollkommene Zahl gibt, wissen wir bis heute genauso wenig, wie ob die Mersenneschen Primzahlen und damit die Vollkommenen Zahlen unendlich sind. Als Zwischenlösung müßten wir aber definitiv wissen, ob es wirklich keine ungeraden Vollkommenen Zahlen gibt. Wenn sie existieren sollten, müßten sie die Form $12m + 1$ oder $36m + 9$ haben, wobei m eine Primzahl ist. Bisher konnte allerdings nur nachgewiesen werden, daß es keine ungeraden Vollkomme-

Madri der, hat angeleitet, man solle die Zahlen 220 und 284 aufschreiben, die kleinere dem Objekt der Begierde zum Essen geben und selbst die größere essen. Er selbst habe die erotische Wirkung davon in eigener Person erprobt, genau wie Ibn Chaldun von den wunderbaren Kräften dieser Zahlen als Talisman Gebrauch gemacht habe.

Befreundete Zahlen heißen zwei natürliche Zahlen m und n dann, wenn die Summe aller positiven Teiler d von m außer m selbst n ergibt, und die Summe aller positiven Teiler d von n außer n selbst m ergibt.

Pythagoras hat die zwei Befreundeten Zahlen 220 und 284 schon gekannt. Zählen wir alle positiven Teiler von 220 auf außer 220 selbst.

$$220 : 1 + 2 + 4 + 5 + 10 + 11 + 20 + 22 + 44 + 55 + 110 .$$

Summieren wir diese Teiler, so ergibt sich die Zahl 284.

Machen wir dasselbe mit 284, also

$$284 : 1 + 2 + 4 + 71 + 142 ,$$

so ergibt sich die Zahl 220.

Die Summe der Teiler einer Befreundeten Zahl ergibt also jeweils die andere Zahl. Verwenden wir für die Summe aller positiven echt kleineren Teiler einer Zahl n das Symbol $S(n)$, so können wir sagen $S(220) = 284$ und $S(284) = 220$, abstrahiert $S(a) = b$, $S(b) = a$, für Befreundete Zahlen a und b . Daraus folgt, daß $S(S(n)) = S^2(n) = n$ für jede Befreundete Zahl n gilt.

Wegen ihrer Teilbarkeitsvorschrift stehen die Befreundeten Zahlen natürlich in einem gewissen Zusammenhang mit den Vollkommenen Zahlen, sie sind sozusagen eine Art Abspaltung. Die Befreundeten Zahlen stehen mit den Vollkommenen Zahlen in folgendem Zusammenhang: Natürlich ist jede Vollkommene Zahl n mit sich selbst befreundet. Denn setzt man n für a und b in S ein, so erhält man $S(n) = n$ und $S(n) = n$, was obige Bedingung für $S(a) = b$ und $S(b) = a$ erfüllt. Ebenso gilt $S(n) = n \rightarrow S(S(n)) = n$. Weniger als 1200 solche Befreundete Zahlen sind bis 1980 bekannt. Euler hat 1750 davon 59 entdeckt. Einige Paare von Befreundeten Zahlen seien aufgeschrieben:

$$220 \quad 1184 \quad 2620 \quad 5020 \quad 6232 \quad 10744 \quad 17296 \quad 9363584$$

$$284 \quad 1210 \quad 2924 \quad 5564 \quad 6362 \quad 10856 \quad 18416 \quad 9437056$$

Doch gibt es nicht nur Paare von Befreundeten Zahlen, sondern auch Ketten, wie zum Beispiel diese Fünfer-Kette:

$$12496, 14288, 15472, 14536, 14264.$$

Hier ergibt die Teilersumme der ersten Zahl die zweite Zahl, deren Teilersumme die dritte usw. und die Teilersumme des letzten Gliedes ergibt wiederum die erste Zahl.

$$n = 12496 \text{ (als Startglied)}$$

$$S(n) = 14288$$

$$S(S(n)) = S^2(n) = 15472$$

...

$$S^5(n) = n = 12496$$

Eine berühmte Kette von 28 befreundeten Zahlen hat $n = 14316$ als Startzahl und es gilt $S^{28}(14316) = 14316$, d.h. nach 28 Gliedern endet die Kette wieder bei der Startzahl. Als eine Misch-Struktur von Teilbarkeit und Ungerad- bzw. Geradzahligkeit erscheint uns nun die schon besprochene Goldbachsche Vermutung, daß jede gerade Zahl, größer als 2, als Summe zweier Primzahlen darstellbar ist.

Wenn eine Zahl teilbar ist und die Summe aller ihrer möglichen Teiler wiederum die Zahl ergibt, so ist das schon eine recht ansehnliche Sache. Wegen dieser perfekten Teilbarkeit nennt man diese Zahlen auch vollkommene Zahlen. Das Perplexe an diesen perfekten Zahlen ist aber, daß ihre vollkommene Teilbarkeit auf vertrackte Weise auf perfekt unteilbaren Zahlen, den Primzahlen, aufgebaut ist. Wie schon früher angemerkt wurde, ist die allgemeine Struktur der geraden vollkommenen Zahlen V : $V = 2^{n-1} \cdot (2^n - 1)$. Aber nur wenn $2^n - 1$ eine Primzahl ist, also nicht teilbar, liefert die Struktur $2^{n-1} \cdot (2^n - 1)$ eine vollkommene Zahl. Das ist wirklich der Gipfel einer Art Vollkommenheit - die gegenseitige Abhängigkeit von Teilbarkeit und Nichtteilbarkeit. Dieses Prinzip des gegenseitigen Bedingens gilt auch für die geraden und ungeraden Zahlen. Denn die vollkommenen Zahlen sind gerade und die Primzahlen, welche für deren Definition benötigt werden, sind ungerade Zahlen.

Die numerische Sensibilität

Eine numerische Sensibilität nimmt die Zahlenverhältnisse und deren Eigenschaften als Modelle für andere Verhältnisse in der Welt (vom Bilderrahmen bis zur Architektur). Die numerische Sensibilität sucht den numerischen Code hinter allen Dingen: Die Zahl als Maß aller Dinge, wie es Pythagoras formulierte. Die Fragmente des Philolaos: B1 "Die Natur aber ward in der Weltordnung aus grenzenlosen und grenzenbildenden Stücken

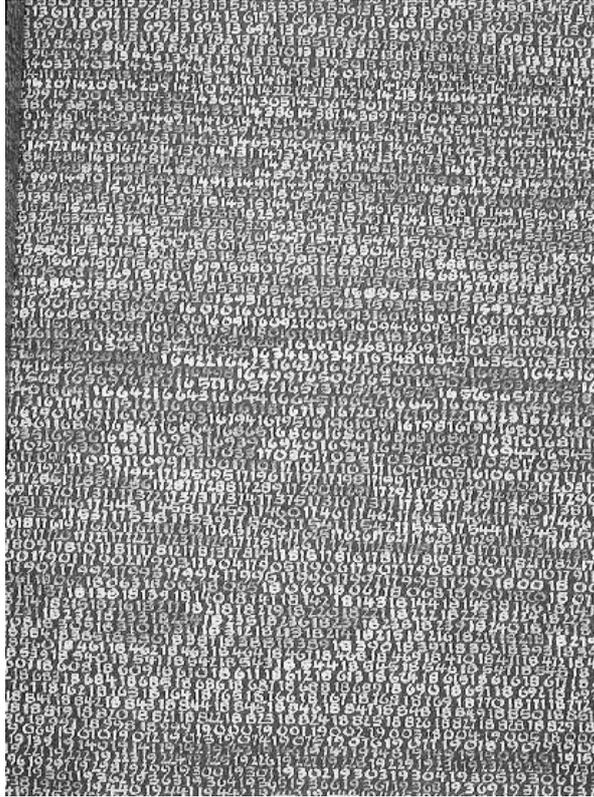


Abbildung 1.7. Roman Opalka, 1-∞ (Detail), 1965

zusammengefügt, sowohl die Weltordnung als Ganzes wie alle in ihr vorhandenen Dinge. " B4 "Und in der Tat hat alles, was man erkennen kann, Zahl." B11 "Denn nichts von den Dingen wäre irgendeinem klar, weder in ihrem Verhältnis zu sich, noch zueinander, wenn die Zahl nicht wäre und ihr Wesen." Die Pythagoräer behaupteten, "das Wesen aller Dinge sei Zahl" (Aristoteles). Die Einheit war für die Pythagoräer sowohl gerade wie ungerade, sie leitete sich aus dem Begrenzten und Unbegrenzten her. Laut Aristoteles nahmen die Pythagoräer an, das Unendliche sei identisch mit dem Geraden. Denn dieses gewähre für sich abgeteilt und von dem Ungeraden begrenzt den Dingen die Unendlichkeit. Die Pythagoräer haben auch die "figurierte Zahl" eingeführt. Sie stellten Zahlen durch die Anordnung von Steinchen her. Sie legten Figuren mit Steinchen (Dreiecke, Quadrate),

welche Zahlen darstellten. Die Pythagoräer kannten auch schon die Proportionslehre. Archytas (B2): "Es gibt aber drei mittlere Proportionalen in der Musik: erstens die arithmetische, zweitens die geometrische, drittens die harmonische". Auch die Ägypter wußten schon über Proportionen Bescheid.

Das Wesen von All und Nichts wird also durch den numerischen Code bestimmt. Auch der genetische Code unterliegt dem numerischen.

Vermehrung und Vererbung gehorchen dem numerischen Code ("Die Zahlen vermehren sich wie die Tiere"; Merz). Fibonacci ist also ein Abkömmling von Pythagoras, dessen Schule ja gerade in Süditalien und Sizilien eine große Anhängerschaft gefunden hatte.

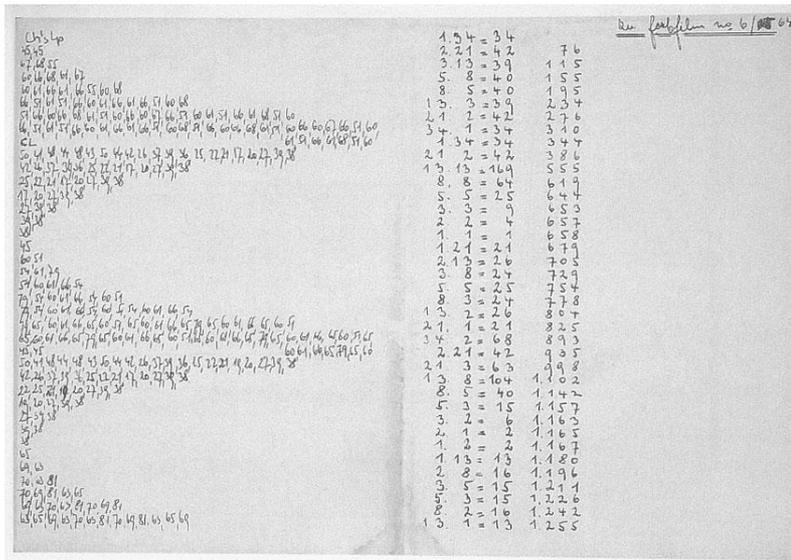


Abbildung 1.8. Kurt Kren, Mama und Papa, Kaderplan, 6/64

Der numerischen Sensibilität dienen also Zahlenverhältnisse als Modelle für andere Verhältnisse der Welt. Zahlenverhältnisse als

1. Hilfsmittel für die Konstruktion eines Werkes, Erklärung.
2. Verhältnis von Zahl und Natur, Naturphänomene als Zahlenphänomene. Die Zahl als Maßsystem der Natur. Zahlgesetze als Weltgesetze, Bewegung der Planeten etc.

3. Parallelen zwischen Zahlen und anderen Wesenheiten wie Farben, Worte.
4. Manifestationen physikalischer, biologischer, sozialer und physiologischer Grundsätze, die für Harmonie, Ordnung etc. sorgen.
5. Offenbarungen geheimer Zusammenhänge (Cryptanalysis, Wortzahlenmystik im Neuen Testament, Numerologie).

Die Fibonacci-Zahlen

Lenardo von Pisa, Sohn (= filius) des Bonacci, deshalb auch Fibonacci genannt, wollte in seinem Werk "Liber Abacci" von 1202 das arithmetische und algebraische Wissen seiner Zeit zusammenfassen, wodurch übrigens die arabischen Zahlen in Europa bekannt wurden. Auf einer dieser Seiten steht die kuriose Aufgabe: Wieviel Kaninchenpaare werden im Laufe eines Jahres, von einem einzigen Paar ausgehend, gezeugt? Unter der Voraussetzung, daß jedes Paar monatlich ein neues Paar wirft und daß die Kaninchen vom zweiten Monat an gebärfähig sind, gelangte er für die einzelnen Monate zu folgenden Zahlen: 1, 2, 3, 5, 8, 13, 21, 34, Diese Zahlenreihe, wo jede Zahl (nach der zweiten) als die Summe ihrer zwei Vorgänger definiert wird, nennt man Fibonacci-Reihe.

$$f_1 = f_2 = 1 \quad f_n = f_{n-1} + f_{n-2} \quad (n \text{ größer als } 2)$$

$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, \dots$$

Die Fibonaccizahlen stehen also in einem arithmetischen Verhältnis, da die Differenz zweier aufeinanderfolgender Zahlen (ähnlich wie bei der arithmetischen Reihe $a - b = b - c$) die Struktur der Fibonacci-Reihe bestimmt, wenngleich sie nicht wie bei der arithmetischen Reihe konstant ist. Um Mißverständnisse zu vermeiden, möchte ich hier anmerken, daß ich in diesem Aufsatz die Begriffe Folge und Reihe synonym verwende. In der formalen Mathematik bezeichnet man in der Regel nur eine Folge von Teilsummen als Reihe.

Die Fibonacci-Reihe hat u.a. folgende Eigenschaften: Jedes Glied der Reihe (ab dem dritten) ist das arithmetische Mittel aus seinem Nachfolger und dem Vorläufer seines Vorläufers. Die Differenz zweier Glieder dieser Folge, die ein Mittelglied einschließen, ist gleich diesem Mittelglied. Eine der schönsten der elementaren Eigenschaften dieser Folge ist, daß sie ihre eigene

Differenzfolge ist: Bildet man die Differenzen $b_n = a_{n+2} - a_{n+1}$ aufeinanderfolgender Glieder der Folge (a_n) , so entsteht wieder die Folge (a_n) selbst, wenn auch leicht verschoben.

Wir haben die Fibonacci-Reihe einfach durch die Werte ihrer Vorläufer definiert. Wenn wir aber einen allgemeingültigen Ausdruck für diese Glieder haben wollen, kommen wir zu dieser komplizierten Formel:

$$a_n = \frac{1}{\sqrt{5}} \left[\left(\frac{1}{2}[1 + \sqrt{5}] \right)^{n+1} - \left(\frac{1}{2}[1 - \sqrt{5}] \right)^{n+1} \right]$$

Das Bestürzende an dieser Formel ist die Tatsache, daß die natürliche Zahl a_n durch einen Ausdruck definiert wird, in dem die Irrationalzahl $\sqrt{5}$ eine wesentliche Rolle spielt. Dies wird noch besonders deutlich, wenn wir setzen

$$c_n = \frac{a_n}{a_{n+1}}$$

Dann können wir mit Hilfe voriger Formel feststellen, daß sich c_n mit wachsendem n dem Wert $g = \frac{1}{2}(\sqrt{5} - 1)$ annähert.

Die Zahl g charakterisiert aber den Goldenen Schnitt. Dieses Verhältnis zeigt sich auch beim Pentagramm oder Drudenfuß. Ein "goldenes" Rechteck ist eines, dessen Seiten sich wie $1 : g$ verhalten.

Besonders interessant für die numerische Sensibilität macht die Fibonaccizahlen ihr exponentielles Wachstum. Sie sind gleichsam "figurierte Zahlen" des Wachstums. Es gehört zur innersten mathematischen Eigenschaft der Fibonaccizahlen, daß sie exponentiell, d.h. rasch wachsen. Es wäre interessant, mathematisch mit Julia Robinson zu zeigen, daß alle rekursive Folgen diophantisch definierbar sind, wenn es eine solche Folge mit exponentiellem Wachstum gibt, das ist eben die Fibonacci-Reihe. Deswegen hat ja Fibonacci angenommen, mit dieser Zahlenreihe hätte er das Muster des natürlichen Wachstums entdeckt, vom Wachsen der Pflanzen bis zum Wachsen der Hasen. Die Fibonacci-Reihe als numerisches Modell der Evolution.

Bei der Blattanordnung mancher Pflanzen, der Phyllotaxis, spielen die Fibonaccizahlen eine Rolle. Beim Übergang von einem zum nächsten Blatt tritt oft eine Schraubung auf, die zum Beispiel eine halbe Drehung enthält. Das Arrangement der Blätter kann also als Bruch definiert werden:

$$\frac{\text{Zahl der vollendeten Drehungen}}{\text{Zahl der Blätter pro Zyklus}}$$

Man spricht dann von $\frac{1}{2}$ Phyllotaxis (Ulme, Linde). Es treten aber auch die Werte $\frac{1}{3}$ (Buche, Haselstrauch), $\frac{2}{5}$ (Eiche, Aprikose), $\frac{3}{8}$ (Pappel, Birne),

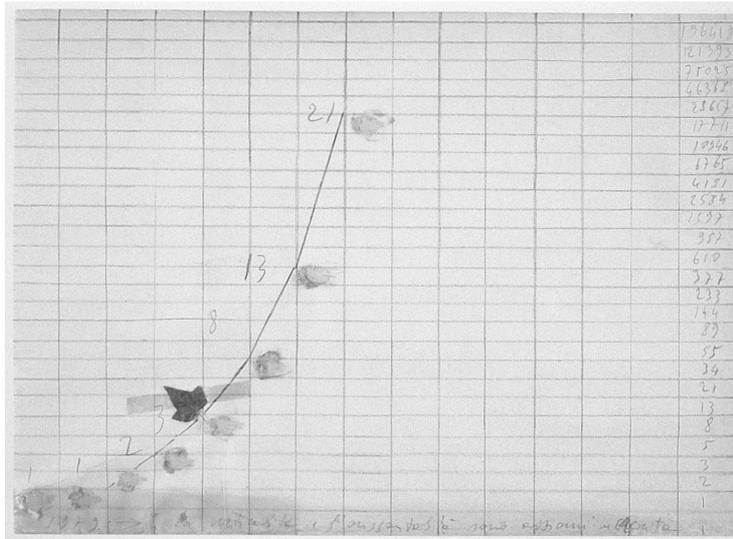


Abbildung 1.9. Mario Merz, la verticalità sono assiomi in crescita, 1979

$\frac{1}{13}$ (Weide, Mandel), auf. Diese Phyllotaxis-Brüche bestehen stets aus Zahlen der Fibonacci-Reihe. Gerade diese Phyllotaxis ist eines der häufigsten ikonographischen Motive bei bildenden Künstlern neben der Spirale, oder eine propellerartige Mischung von Phyllotaxis und Spirale. Die Fibonacci-Zahlen eignen sich eben besonders zur Darstellung des Wachstums und des Lebens; sie sind eine Frühform von formalen Produktionssystemen. Nicht nur über die Phyllotaxis ist das erfaßbar, sondern auch über die Spirale. Denn bei der Ananas zum Beispiel finden wir Arrangements aufsteigender Spiralen als Ergebnis der Phyllotaxis. Spirale und Phyllotaxis sind also schon in der Natur verbunden.

Von den vielen Eigenschaften und Paradoxien der Fibonacci-Folgen seien jedoch besonders jene untersucht, die sie mit dem Goldenen Schnitt in Beziehung bringen.³

Die Überzeugung der Pythagoräer war es, daß jedes Ding und jeder Begriff in der Welt durch eine Zahl gekennzeichnet werden kann (Philolaos von Kroton: "Und wirklich hat alles, was erkannt wird, Zahl. Denn es ist unmöglich, daß ohne diese irgend etwas im Denken erfaßt oder erkannt wird".) Die gegenseitigen Beziehungen der Dinge sind durch das Verhältnis der ihnen zugeschriebenen ganzen Zahlen ausdrückbar. Das griechische

Wort logos für Verhältnis heißt im lateinischen ratio. Daher nennen wir die Verhältnisse ganzer Zahlen rationale Zahlen, sie umfassen die ganzen Zahlen ($5 : 1 = 5$) und die gewöhnlichen Brüche ($1 : 2 = 1/2$). Diese schreiben wir mit einer ganzen Zahl als Zähler über und einer ganzen Zahl als Nenner unter den Bruchstrich: zum Beispiel $\frac{5}{1}$, $\frac{1}{2}$ usw. Da man die ganzen Zahlen als Verhältnis schreiben kann, gehören auch sie zu den rationalen Zahlen wie die Brüche. Der Pythagoräismus war der erste Ausdruck der numerischen Sensibilität.

Das Pentagramm und der Goldene Schnitt

Doch ausgerechnet das Geheimabzeichen der Pythagoräer, der regelmäßige Fünfstern aus fünf Linien, das Pentagramm, das Geheimabzeichen der Pythagoräer, brachte das pythagoräische Weltbild, daß man die Beziehungen sämtlicher Dinge durch das Verhältnis (den logos) ganzer Zahlen (arithmoi) beschreiben könne, zum Einsturz. Denn wenn wir das Verhältnis zwischen

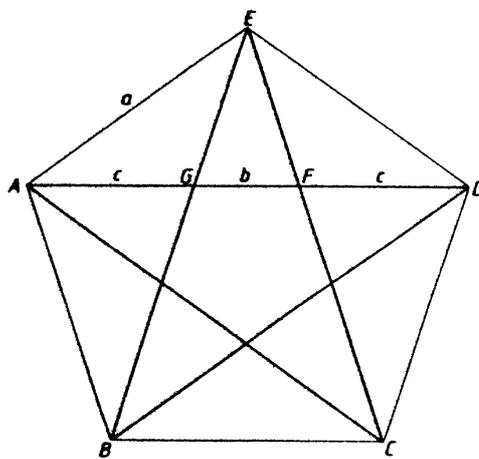


Abbildung 1.10. Pentagramm

einer Diagonale d des regelmäßigen Fünfecks (zum Beispiel die Linie von B nach E in Abb. 1.10) und seiner Seite s (zum Beispiel die Linie von A nach B in Abb. 1.10), das geometrisch einfach zu bestimmen war, versuchen numerisch darzustellen, dann verhält sich zwar $s : d$ ungefähr wie $3 : 5$ oder wie $5 : 8$ oder wie $13 : 21$, aber nicht genau. Die genannten Verhältnisse sind

bloß Näherungswerte für das tatsächliche geometrische Verhältnis. Beim Pentagramm lautet die Beziehungsgleichung zwischen Diagonale und Seite $d^2 = s^2 + d \cdot s$ oder $\frac{d}{s} = \frac{s}{d-s}$.

Die Diagonale d verhält sich also zur Seite s wie die Seite s zur Differenz von Diagonale und Seite ($d - s$). Diese Verhältnisse müßten nach pythagoräischer Lehre rational sein und d und s folglich ganze Zahlen. Wenn wir uns nun fragen, welche Zahlenfolgen kommen am ehesten für diese Beziehungsgleichung in Frage, so kommen wir wieder auf die Fibonaccizahlen. Man kann nämlich die Fibonaccizahlen paarweise in Brüche aus aufeinanderfolgenden Fibonaccizahlen verwandeln: $1/2, 2/3, 3/5, 5/8, 8/13$, usw. Wenn wir nun diese Brüche in Dezimalzahlen verwandeln, kommen wir zu folgenden Werten (bei drei Dezimalstellen):

1 : 2	0,500
2 : 3	0,667
3 : 5	0,600
5 : 8	0,625
8 : 13	0,615
13 : 21	0,571
21 : 34	0,618
34 : 55	0,618

Wir sehen, die Brüche der Fibonaccizahlen streben einem Grenzwert zu: 0,618... Wir können aber auch andere Brüche aus Paaren von Fibonaccizahlen herstellen, indem wir den Kehrwert von den ersteren bilden und diesen dann ebenfalls in eine 3-stellige Dezimalzahl verwandeln:

$\frac{1}{1} = 1.000$	$\frac{13}{8} = 1.625$
	$\frac{21}{13} = 1.615$
	$\frac{34}{21} = 1.619$
$\frac{2}{1} = 2.000$	$\frac{55}{34} = 1.618$
$\frac{3}{2} = 1.500$	$\frac{89}{55} = 1.618$
$\frac{5}{3} = 1.667$	$\frac{144}{89} = 1.618$
$\frac{8}{5} = 1.600$	$\frac{233}{144} = 1.618$

Hierbei sind also zunächst die Zähler und dann die Nenner fortschreitende Fibonaccizahlen. Auch diese rationalen Brüche oder ganzzahligen

Verhältnisse streben einem Grenzwert zu: $1.618\dots$, dem sie immer näher kommen. 5-stellig lautet er $1.61803\dots$ und ist mit einem anderen berühmten Bruch identisch, den wir erhalten, wenn wir eine beliebige Strecke der Länge $x + y$ so in zwei Teile teilen, daß das Verhältnis der ganzen Strecke $(x + y)$ zur größeren Strecke x das gleiche ist wie das Verhältnis der größeren Strecke x zur kleineren Strecke y .

Dieses Verhältnis nennen wir den Goldenen Schnitt:

$$\frac{x + y}{x} = \frac{x}{y} .$$

Ausmultipliziert ergibt das $x^2 = xy + y^2$ oder $x^2 - xy - y^2 = 0$, oder $(x - y/2)^2 = \frac{5y^2}{4}$, welches das Verhältnis

$$\frac{x}{y} = \frac{(1 + \sqrt{5})}{2}$$

liefert. Es ist $\frac{(1 + \sqrt{5})}{2} = 1.61803\dots$, also die gleiche Zahl wie der Grenzwert von Brüchen von Zahlen aus der Fibonacci-Reihe.

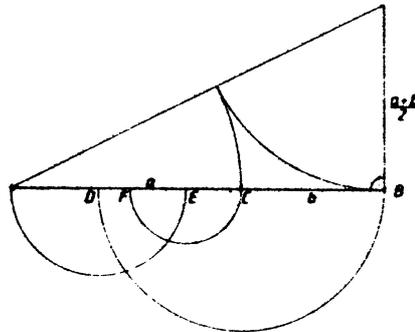


Abbildung 1.11. Geometrische Konstruktion des Goldenen Schnitts

Die "heilige Verhältniszahl" des Goldenen Schnitts ist übrigens die einzige Zahl, die ganz einfach in ihre reziproke Zahl verwandelt werden kann, indem man 1 abzieht: $x - 1 = \frac{1}{x}$. Daraus erhält man $x^2 - x - 1 = 0$, also obige Gleichung mit $y = 1$ eingesetzt, was $x = \frac{1 + \sqrt{5}}{2}$ als Lösung hat. Es ist also

$$\frac{(1 + \sqrt{5})}{2} - 1 = \frac{2}{(1 + \sqrt{5})} .$$

Die irrationale Zahl 1.61803... hat für einige Verwirrung in der Literatur gesorgt und zu Verwechslungen von Fibonacci-Reihe, Goldener Schnitt und Harmonikalität gesorgt, die zwar in Beziehung zueinander stehen, aber nicht dasselbe sind. Sie alle haben mehr oder minder mit der pythagoräischen Tradition zu tun, deswegen wollen wir sie etwas genauer untersuchen und differenzieren.

Die Verhältniszahl des Goldenen Schnitts als Harmoniegesetz hat in der klassischen Kunst, von Dürer über Raffael zu Tizian, eine bekannt große Rolle gespielt. Dem Goldenen Schnitt GS als Maß in der Kunst entspricht auch ein GS als Maßverhältnis in der Natur.

„Sehr oft läßt sich an Blättern und Blüten das Maßverhältnis des Goldenen Schnitts nachweisen, so beim Blatt des Goldregens, der Alpengänsedistel, der Maiblume usw. Das Schneeglöckchen ordnet seine Blütenblätter im gleichseitigen Dreieck an, während uns die Blüten der verschiedenen Lilienarten das zum Sechsstern verdoppelte gleichseitige Dreieck zeigen. Außerdem dürfte bekannt sein, daß die Bienenwabe aus vielen nebeneinander gefügten reinen Sechsecken gebildet wird.

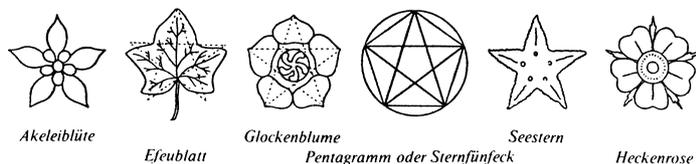


Abbildung 1.12. Proportionsverhältnis bei Pflanzen und Meerestieren

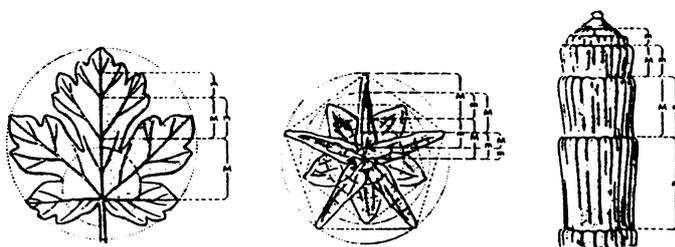


Abbildung 1.13. Goldene-Schnitt-Proportionen bei Pflanzen, von links Hahnenfuß, Seidenpflanze, Schachtelhalm

Auch das Quadrat läßt sich als Grundform vielfach in der Natur nachweisen: so treffen wir es in der Verdoppelung als Achtstern bei den Kreuz-

blütlern, bei Männertreu und Wiesenschaumkraut, und als Achteck in seltener Regelmäßigkeit bei der Einbeere, einem staudigen Liliengewächs. Dagegen bildet die ungefüllte Blüte der Dahlie einen achtstrahligen Stern. Noch häufiger begegnet uns aber das Fünfeck und das Sternfünfeck oder Pentagramm, eine Figur, der Jahrtausende hindurch geheimnisvolle Bedeutung beigemessen wurde. In der Pflanzenwelt treffen wir diese Form am klarsten in der Akeleiblüte, in der Tierwelt beim Seestern an. Zahlreiche andere Blüten, wie die Glockenblume, die Nelke, die Heckenrose, die Lindenblüte, der Phlox und andere zeigen diese Grundform, ebenso wie verschiedene Blätter, etwa das Himbeerblatt. Ziehen wir beispielsweise die Pentagrammform über ein Efeublatt, so stellen wir bei aller Unregelmäßigkeit, die dieses Blatt sonst aufweist, fest, daß die Grundverwandtschaft mit der Fünfecksform besteht. Es scheint, als bemühe sich die Natur, der idealen mathematischen Grundform so nahe wie möglich zu kommen. Die Abbildungen verdeutlichen dies.

Gerade der Fünfstern, also das bereits erwähnte Pentagramm, hat für unsere weiteren Betrachtungen besondere Bedeutung, teilen sich doch die Pentagrammseiten 'stetig' im Goldenen Schnitt, dem wir in der Natur auch beim Wachstum der Pflanze, bei einem edel gebauten menschlichen Körper oder den Abmessungen eines Pferdekörpers begegnen. So führt das stetig fortschreitende Wachstum der Pflanze, wie es der kleine Pappelzweig zeigt, häufig zu einer stetigen Teilung, wenn auch das ungeschulte Auge oft keine Regel erkennen kann. Die Strecken zwischen den einzelnen Knotenpunkten stehen sehr schön im 'Goldenen Verhältnis', das auch bei weiterem Längenwachstum beibehalten wird." [Hangenmaier, 1977].

Die harmonikalen Proportionen und der pythagoräische Traum

Die Harmonik geht auf Pythagoras zurück. Ihre Lehre ist, daß nicht nur unser Ohr ganzzahlige Intervallproportionen bevorzugt - die Pythagoräer glaubten ja daran, daß die ganze Welt durch das Verhältnis (logos) ganzer Zahlen beschrieben werden kann -, sondern daß diese Intervalle der Musik auch Naturgesetze sind, siehe Johannes Keplers "Weltharmonik". Die 12 musikalischen Hauptintervalle entstehen durch die Teilungen einer Saite nach ganzzahligen Verhältnissen. Schwingt eine Saite (einer beliebigen Länge), erhalten wir den Grundton, die Tonika. Vibriert nur mehr die Hälfte, besteht also das Verhältnis 1:2, steigt der Ton und wir erhalten die Oktave. Das Verhältnis der Saitenlängen, die schwingen, zu denen, die nicht

schwingen, kann auch als Verhältnis von Wellenlängen und von Frequenzen aufgefaßt werden. Wo immer aber die gleiche Proportion zwischen schwingendem und ruhigem Saitenabschnitt vorhanden ist, erklingt das gleiche Intervall. Die weiteren Proportionen der 12 Hauptintervalle sind

2:3	Quinte (der Ton steigt um ein Fünftel)	5:9	kleine Septime
3:4	Quarte	8:9	große Sekunde
3:5	große Sexte	8:15	große Septime
4:5	große Terz	15:16	kleine Sekunde
5:6	kleine Terz	32:45	Tritonus
5:8	kleine Sexte		

Für die traditionelle Musiktheorie gelten die Intervalle bis zur kleinen Sexte als Konsonanzen, der Rest als Dissonanzen.

In harmonischer Proportion befinden sich die 3 Zahlen a, b und c in folgender Proportion:

$$\frac{1}{a} - \frac{1}{b} = \frac{1}{b} - \frac{1}{c} \text{ oder } \frac{c}{a} = \frac{c-b}{b-a}$$

Man nehme zum Beispiel für a, b und c : 2, 3 und 6. Gerade die Folge der Konsonanzen hat Glieder: 1 : 2, 2 : 3, 3 : 5, 5 : 8; (Oktave, Quinte, große Sexte, kleine Sexte), die mit den ersten Gliedern der Fibonacci-Reihe, als Brüche definiert, übereinstimmen: $1/2, 2/3, 3/5, 5/8 \dots$ bzw. $2/1, 3/2, 5/3, 8/5 \dots$. Und gerade diese rationalen Brüche oder ganzzahligen Verhältnisse bzw. Proportionen kommen dem tatsächlichen geometrischen Verhältnis zwischen Diagonale d und Seite s im Pentagramm immer näher, und zwar je dichter das Proportionsintervall, desto genauer. Die Beziehungsgleichung zwischen Diagonale d und Seite s lautet ja beim Pentagramm: $d^2 = s^2 + d \cdot s$. Demgemäß gilt $d : s = s : (d - s)$. d soll sich zu s wie s zur Differenz von $d - s$ verhalten.

Dieses Verhältnis müßte nach pythagoräischer Lehre rational sein, daher d und s ganzzahlig. Versuchen wir nun, diese Forderung zu erfüllen, und dieses (geometrische) Verhältnis von d und s mit Zahlen zu besetzen, so sehen wir, daß dies mit Paaren aus der Fibonacci-Reihe oder Harmonik am besten gelingt.

d	1	2	3	5	8	13	21
s	1	1	2	3	5	8	13
$d - s$	0	1	1	2	3	5	8

$\frac{d}{s} = \frac{s}{d-s}$ kann man ja transformieren in $d(d-s) = s \cdot s$.

Wenn wir nun die obigen Zahlen in diese Gleichung einsetzen, müßten bei $d(d - s)$ und $s \cdot s$ die gleichen Summen herauskommen:

$$\begin{array}{r} d(d - s) \dots \quad 0 \ 2 \ 3 \ 10 \ 24 \ 65 \ 168 \\ s \cdot s = s^2 \dots \quad 1 \ 1 \ 4 \quad 9 \ 25 \ 64 \ 169 \end{array}$$

Es klappt nur fast, denn wenn wir die untereinanderstehenden Zahlen dieser Zeilen vergleichen, sehen wir, daß der Unterschied immer geringer wird, je größer die Zahlen werden. Dennoch wird man nie zwei Zahlen finden, seien sie noch so groß, für welche $d(d - s)$ exakt gleich $s \cdot s (= s^2)$ ist.

Was war nun diese Beziehung von Diagonale d und Seite s im Pentagramm eigentlich? Wenn wir d als ganze Strecke x plus y nehmen, und s als die größere Strecke, dann ergibt die Differenz $d - s = (x + y) - x$ also die kleinere Strecke y . Wenn sich nun $d : s$ wie $(x + y) : x$ verhalten soll, dann verhält sich auch $s : (d - s)$ wie $x : y$. Das bedeutet aber: $d : s = s : (d - s)$ beschreibt die gleiche Proportion wie $(x + y) : x = x : y$. Das Verhältnis $d : s = s : (d - s)$ des Pentagramms hieß bis zum Mittelalter "proportio divina" und seit der Renaissance Goldener Schnitt.

Diese Proportion war zwar durch die Fibonaccizahlen am annäherndsten mit Hilfe ganzer Zahlen zu errechnen - wie es der Pythagoräische Traum vorschrieb -, aber wie wir gesehen haben, nie exakt. Es blieb immer ein Rest. Dieser bedeutet, daß Diagonale und Seite eines Pentagramms in ihrem Verhältnis nicht durch zwei ganze Zahlen, als rationaler Bruch, darstellbar sind, daß also d und s kein gemeinsames Maß haben, also inkommensurabel sind. Das Verhältnis der größeren Strecke x (Major) zur kleineren Strecke y (Minor) des Goldenen Schnitts tendiert zwar wie die Brüche aus aufeinanderfolgenden Fibonaccizahlen einem Grenzwert zu, nämlich $\frac{x}{y} = \frac{(1+\sqrt{5})}{2} = 1.61803\dots$, doch ist dies wegen $\sqrt{5}$ ein irrationaler Bruch.

In der Praxis wäre diese Inkommensurabilität, Unmeßbarkeit, durch ein Fortsetzen der Fibonacci-Reihe zu immer größeren Zahlen vernachlässigbar und das Verhältnis von $d : s$ bzw. $x : y$ mit gewünschter Genauigkeit erreichbar gewesen, doch theoretisch nicht mehr. So zerbrach der Pythagoräische Traum und der Begriff der Inkommensurabilität tauchte auf. Für den Hauptwert des Goldenen Schnitts (0,618), den Major, kannte man bereits seit der Antike als Annäherungslösung die harmonikale Proportion 5:8 (=0,625), die nur um 0,007 von ihm abweicht. Wahrscheinlich hat man aus praktischen Gründen immer mit solchen Annäherungswerten gearbeitet, die man aus der Harmonik oder der Fibonacci-Reihe nahm.

Man kann mit Quadraten von den Seitenlängen 1, 2, 3, 5, 8, 13 . . . (also Fibonaccizahlen) Rechtecke aufbauen, deren Seitenverhältnisse ständig besser dem des Goldenen Schnitts entsprechen. Die Fibonacci'schen Zahlenverhältnisse nähern sich immer besser dem irrationalen Verhältnis $d : s$ an, so wie man aus jedem Rechteck, dessen Seitenlängen gleich Diagonale d und Seite s eines regelmäßigen Fünfecks sind, durch Abspaltung eines Quadrates ein dem vorigen ähnliches Rechteck bekommt - man aber an kein Ende kommt.

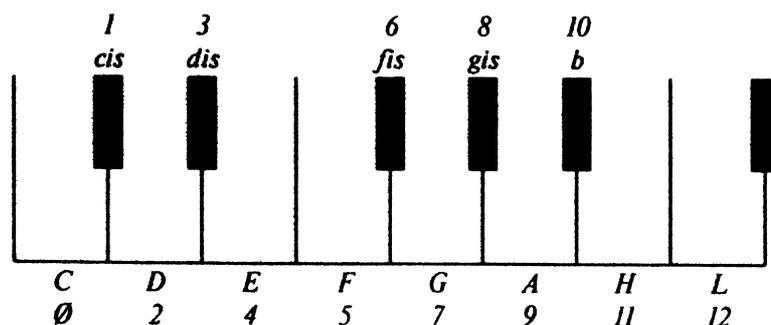


Abbildung 1.14. J. S. Bach, wohltemperiertes Klavier

Eine andere Annäherung an die harmonikalischen Proportionen hat J. S. Bach mit seinem wohltemperierten Klavier durchgeführt. Während die griechische Harmonielehre bzw. die Fibonacci-Reihe auf ganzen Zahlen und deren Verhältnis aufgebaut (also arithmetischer Natur) waren, konstruierte Bach eine geometrische Reihe (ähnlich der geometrischen Natur des Goldenen Schnitts) als Basis für seine Harmonik.⁴ Wie bereits erwähnt, ist das Schwingungsverhältnis von Tonika zu Oktav 1:2.

Die 11 zwischen Tonika und Oktav liegenden Töne sollen sich nicht durch (verschiedene) ganzzahlige Verhältnisse ergeben (wie dies ein sehr guter Geigenvirtuose auszuführen vermag, indem er mit dem Bogen die Saiten in entsprechenden Proportionen zum Schwingen bringt), sondern alle in gleichem Verhältnis zueinander stehen. Das bedeutet für aufeinanderfolgende Töne x, y, z die Beziehung $x : y = y : z$ oder $y = \sqrt{x \cdot z}$ (geometrische Reihe). Damit das Verhältnis 1:2 zwischen Tonika und Oktav auch erhalten bleibt, wenn alle benachbarten Töne ein konstantes Schwingungsverhältnis haben, ist es notwendig, daß dieses gleich $1 : \sqrt[12]{2}$ ist. Von der Tonika zur Oktav kann man daher die einzelnen 12 Töne durch die folgende geometrische

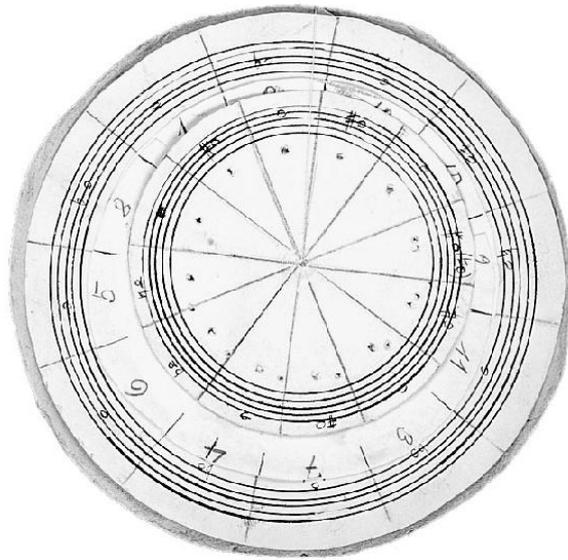


Abbildung 1.15. Arnold Schönberg, Zwölftondrehscheibe für "Bläserquintett op. 26", ©Arnold Schönberg Center, Wien

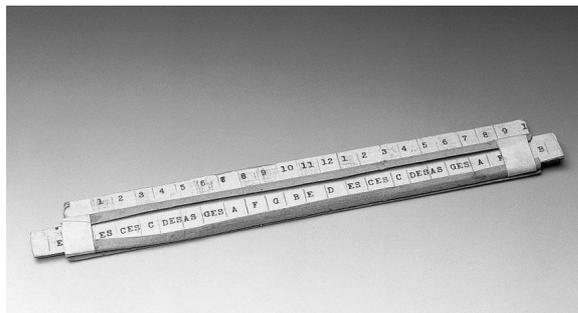


Abbildung 1.16. Arnold Schönberg, Zwölftonreihenschieber für "Serenade op. 24 (box 24)", ©Arnold Schönberg Center, Wien

Proportion genau festlegen: $1, \sqrt[12]{2}, (\sqrt[12]{2})^2, (\sqrt[12]{2})^3, (\sqrt[12]{2})^4, \dots, (\sqrt[12]{2})^{11}, 2$.

Aus dieser Gleichberechtigung aller Töne durch Glättung der arithmetischen Verhältnisse entwickelte sich die 12-Tonmusik Hauer's und Schönberg's mit ihren eigenen Gesetzmäßigkeiten.

Das Ende des pythagoräischen Traums

Wegen der Ungenauigkeit der Messungen kann das Vorkommen des Goldenen Schnitts in Wirklichkeit ein Beweis für harmonikalisches Bauen sein. Die Fibonacci-Reihe schlägt eine Brücke zwischen zwei unversöhnlichen Gegnern, dem Goldenen Schnitt, der Serie irrationaler Brüche, und der Harmonik, der Serie rationaler Brüche.

Desweiteren wurde die Apotheose des Goldenen Schnitts durch Luca Pacioli (1445-1514) in seinem Buch "Divina proportione" (mit Abbildungen von Leonardo da Vinci) unter dem Eindruck des religiösen Dogmas von der Trinität geführt, sodaß es durchaus wahrscheinlich ist, daß in Wahrheit bis dahin nach harmonikalischen Intervallproportionen gebaut worden war, was einfacher, meßbarer und rationaler (im doppelten Sinne des Wortes) war als mit den irrationalen, inkommensurablen, komplizierten Goldenen Schnitt-Zahlenverhältnissen. Oder man baute nach Fibonacci-Zahlen, deren Brüche ja auf einen Grenzwert zu tendieren, welcher der Major des Goldenen Schnitts (0,618...) ist. Die Fibonacci-Reihe schlägt also eine Brücke zwischen dem Goldenen Schnitt und der Harmonik: Der Major liegt nämlich den musikalischen harmonikalischen Konsonanzen in Gestalt der beiden Sexten (3:5, 5:8) am nächsten.

Harmonik, Goldener Schnitt und Fibonacci-Reihe berühren sich in vielen Punkten, stehen aber im Grunde jeder für sich, wenn nicht in Gegensatz zueinander. Sie bilden die bzw. kommen aus der ersten Grundlagenkrise der Mathematik. Wir haben gesehen, die Harmonik besteht aus ganzzahligen Intervallproportionen, welche wie die Fibonaccizahlen ein ganzzahliges Verhältnis zueinander haben, ein arithmetisches (arithmos bedeutet ganze Zahl) Verhältnis. Der Goldener Schnitt ist aus Strecken gebildet, die nicht exakt durch das Verhältnis zweier ganzer Zahlen ausgedrückt werden können, sie sind inkommensurabel, aber man kann sie zeichnen, sie haben ein geometrisches Verhältnis. Die Fibonacci-Reihe ist die Zahlenfolge, mit der man am besten die stetige geometrische Teilung des Goldenen Schnitts ganzzahlig darstellen kann (in ganzzahligen rationalen Brüchen).

Die Fibonaccizahlen kamen dem pythagoräischen Traum am nächsten. Denn das Verhältnis von Diagonale d und Seite s des Pentagramms, das in der Form $d : s = s : (d - s)$ als "göttliche Proportion" bzw. "Goldener Schnitt" bekannt ist, ist durch Fibonaccizahlen am annäherndsten zu errechnen, wenn auch nicht ganz. Denn wie die Brüche aus aufeinanderfolgenden Fibonaccizahlen strebt auch das Verhältnis der größeren Strecke zur kleine-

ren Strecke des Goldenen Schnitts einem Grenzwert zu, nämlich

$$\frac{x}{y} = \frac{(1 + \sqrt{5})}{2} = 1.61803\dots,$$

doch ist dies wegen $\sqrt{5}$ ein unendlicher irrationaler Bruch, was soviel heißt wie, daß eben der Goldener Schnitt nicht restlos durch rationale Brüche ganzer Zahlen beschreibbar ist, sondern ein unauflöslicher, wenn auch bei immer größeren Fibonaccizahlen immer kleiner werdender Rest bleibt – das Unendliche (welches die Punkte . . . bedeuten).

Der Goldene Schnitt und die Harmonik sind gewissermaßen Gegner. Ein irrationaler Bruch ist ein Quotient von Zahlen, der (im allgemeinen im Zähler) mindestens eine Wurzel (aus einer positiven Zahl) enthält, sodaß er einen unendlichen nichtperiodischen Dezimalbruch ergibt. Ein rationaler Bruch ist ein Quotient zweier natürlicher (oder ganzer) Zahlen, was einen endlichen oder periodischen (in seiner Bauart also durch endlich viele Ziffern beschreibbaren) Dezimalbruch ergibt.

Der harmonikalische pythagoräische Traum geht davon aus, daß die Welt durch das Verhältnis ganzer Zahlen beschrieben werden kann. Da das griechische Wort für Verhältnis *logos* im Lateinischen *ratio* heißt, nennen wir die Verhältnisse ganzer Zahlen (*arithmoi*) rationale Zahlen. Sie bilden auch den Mythos für einen rationalen, logischen Aufbau der Welt. Doch aus der Zahl selbst, aus den Eigenschaften der *arithmoi*, den Verhältnissen ganzer Zahlen, wurde die Irrationalität geboren. Aus dem Schoße der *Ratio*, dem Zahlenverhältnis, entsprang die Irrationalität. Das Symbol des Maßes, die Zahl, gebar auch die Idee der Inkommensurabilität, des Nichtmeßbaren. Denn ausgerechnet das Geheimabzeichen der Pythagoräer, das Pentagramm, ließ sich nicht exakt ganzzahlig, rational darstellen. Wenn schon nicht einmal das Pentagramm als Verhältnis ganzer Zahlen beschrieben werden konnte, dann natürlich auch nicht die Welt. Die Pythagoräer trugen diesen Traum dennoch weiter, obwohl ihnen schon seit dem 5. Jahrhundert vor Christi der Begriff der Inkommensurabilität bekannt war: Der Beweis für die Unmöglichkeit, das Verhältnis der Quadratwurzel aus 2 zur Einheit in ganzen Zahlen auszudrücken, also der Unmöglichkeitsbeweis für die Beziehung $d^2 : s^2 = 2s^2 : d^2$. Die Seite s und die Diagonale d des Quadrates waren schon inkommensurabel. Die ungeheure Tragweite dieser Entdeckung der Inkommensurabilität und des Irrationalen aus dem Schoße des Maßes und der *Ratio* selbst hat nicht nur die griechische Mathematik erschüttert.⁵

Der neue pythagoräische Traum: das digitale Prinzip

Der neue pythagoräische Traum lautet: kann ein Computer die Welt perfekt simulieren? Nicht nur in der Antike, auch in der Neuzeit träumt man vom pythagoräischen Prinzip. Die Beschreibbarkeit der Welt durch ganze Zahlen steigerte sich mit dem Church-Turing Prinzip zum Ideal der effektiven Berechenbarkeit der Welt. David Deutsch definierte 1985 dieses Prinzip nach den Namen zweier Logiker: "Every finitely realizable physical system can be perfectly simulated by a universal model computing machine operating by finite means"[Deutsch, 1985].

George Boole (1815-1864) hat im 19. Jahrhundert in seinem Buch "Laws of Thought", der Begründung des Logikkalküls, diesem Traum die Grundlage geliefert, indem er die Grundlage für eine Verbindung von Zahl und Elektrizität lieferte. In diesem Buch steht der Satz "Die Bedeutung der Symbole 0 und 1 im System der Logik sind das Nichts und das All". Er folgt dabei der Spur des binären Zahlensystems von Leibniz, das darin besteht, alle Zahlen durch bloß zwei Ziffern (0, 1) darzustellen. Die binäre Darstellung der Zahlen gehört zur Grundlage der digitalen Welt. Denn erst mit ihrer Hilfe kann man durch Elektrizität Zahlen darstellen, kann der Computer rechnen. Hat nämlich der Logiker Boole die Symbole 0 und 1 mit Nichts und All gleichgesetzt und darauf einen Logikkalkül, eine Netzalgebra etc. aufgebaut, so konnte der Ingenieur daraus Schaltungssysteme ableiten, indem er für das Symbol 1 Elektrizität und für das Symbol 0 Nicht-Elektrizität setzte. Floß Strom durch die Leitung hieß das 1, floß kein Strom durch die Leitung hieß das 0. Mit dieser sequentiellen Abfolge von Strom und Nichtstrom, von Elektrizität und Nichtelektrizität, von 1 und 0 konnte der Computer (eine elektrische Maschine) Zahlen darstellen und Rechenvorgänge ausführen. Die Darstellung digitaler Zahlen (das sind Zahlen, die nur durch zwei Ziffern ausgedrückt werden, nämlich 0 und 1) durch Elektrizität und Nicht-Elektrizität bildet die Basis für die elektronische Welt des Computers, für unsere elektronische, digitale Welt. Ein Dualismus, der sich aber zu etwas neuem verbindet (nämlich zwei Ziffern, die jeweils durch eine andere Anordnung eine andere Zahl darstellen), zeichnet das digitale Denken aus. Dieser binäre Code ist die neue Form der numerischen Sensibilität, der neue pythagoräische Traum: Elektronisierung der Welt. Zahl: Elektrizität: Digitalisierung der Welt. Die Zahl steckt also nicht nur hinter der Proliferation des Hasen, sondern steht auch in Zusammenhang mit der Elektrizität. Das digitale Prinzip ist also der neue pythagoräische Traum.

Der quantifizierende Geist, l'esprit de la géometric, hat bereits im 18. Jahrhundert einen ersten Höhepunkt erreicht. Nach 1900 erschienen mehrere Werke, die den "geometrischen Geist" bzw. das digitale Prinzip spezifisch auf das Wachstum von Formen lebender Organismen anwendeten. Theodore Andrea Cook publizierte 1903 "Spirals in Nature and Art" und 1914 "The Curves of Life", ein Werk, das ganz der Spirale (the most beautiful of curves" - A.R. Wallace) und ihrem geschichtsmächtigen Wirken in der Biologie, Botanik, Kosmologie, Architektur und visuellen Kunst gewidmet war [Cook, 1979]. Die endlose Kurve der Spirale vermittelte den Eindruck kontinuierlicher Bewegung.

Die Spirale oder die Helix ist ein Phänomen, das so oft in allen Formen des Wachstums, bei Pflanzen, bei Tieren, beim menschlichen Körper, vom Andromeda-Nebel bis zur DNA (Doppel Helix) beobachtet wurde, daß die Vermutung nahe liegt, sie läge allem Wachstum und damit allen Formen des Lebens zugrunde. Insbesondere die logarithmische Spirale ("Spira mirabilis" nannte sie Bernoulli 1691) wird als Manifestation jener Energie betrachtet, die beim Wachstum organischer Körper arbeitet. Cook ist sich allerdings klar, daß der Geometrisierung des Lebens ("to define a natural object in mathematical terms") Grenzen gesetzt sind und ebenso dem organischen Aspekt der logarithmischen Spirale ("a logarithmic spiral is as near as we can get in mathematics to an accurate definition of the living thing") [Cook, 1979].

In der Geschichte der Kunst sind allerdings in der Tat Spiralförmigkeiten, besonders wegen ihres Aspektes der Symmetriebrechung und der Chiralität (Links- oder Rechts-Händigkeit), gerne als "Kurven des Lebens" zur Darstellung des Lebens benützt worden.

Das gnomonische Wachstum und die logarithmische Spirale

Die logarithmische Spirale und ihre Eigenschaft des gnomonischen Wachstums hat auch D'Arcy Thompson bei seinem Studium des Wachstums lebender Organismen fasziniert. In seinem Buch "On Growth and Form" (1914) hat er der "gleichwinkligen Spirale", wie Descartes und Roger Cotes die logarithmische Spirale nannten, ein eigenes Kapitel gewidmet. Was ihn bei der Spirale als Wachstumsform interessierte, war, daß das spiralförmige Gehäuse, wie auch das darin enthaltene Lebewesen, an Größe zunimmt, aber seine Form nicht verändert: "Die Existenz dieses konstanten Wachstumsverhältnisses oder dieser konstanten Formähnlichkeit ist das Wesentli-



Abbildung 1.17. Robert Smithson, Spiral Jetty, Great Salt Lake, Utah, 1970

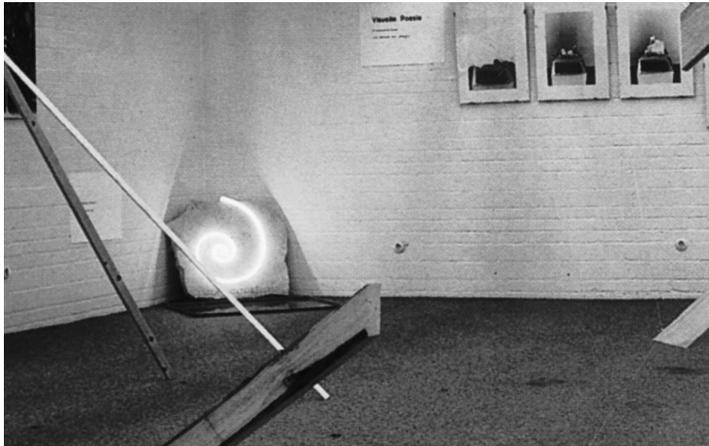


Abbildung 1.18. Peter Weibel, Eadem Mutata Resurgo, logische Spirale, 1975.

che bei der gleichwinkligen Spirale.”[Thompson, 1966].

Darauf hatte schon Christian Wiener in den “Grundzügen der Weltordnung” (1863) hingewiesen. Jeder Zuwachs gleicht seinem Vorgänger. Das Wachstum geschieht durch symmetrische Ausdehnung und so bleibt seine Form unverändert bewahrt. Diese Eigenschaft kontinuierlicher Ähnlichkeit gibt es unter allen Kurven nur bei der gleichwinkligen bzw. logarithmischen Spirale. Daraus entsteht die gnomonische Eigenschaft der Spirale, nämlich beim Wachstum außer der Größe keine Veränderung zu erleiden: Formin-

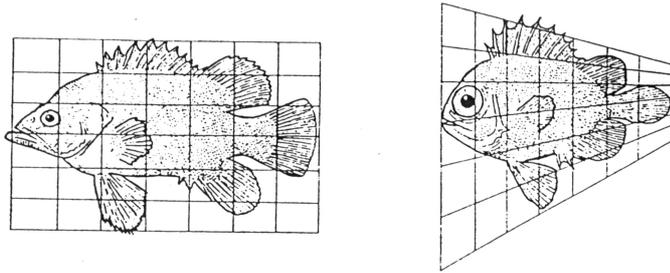


Abbildung 1.19. Links: Polyprion, rechts: Pseudopriacanthus altus

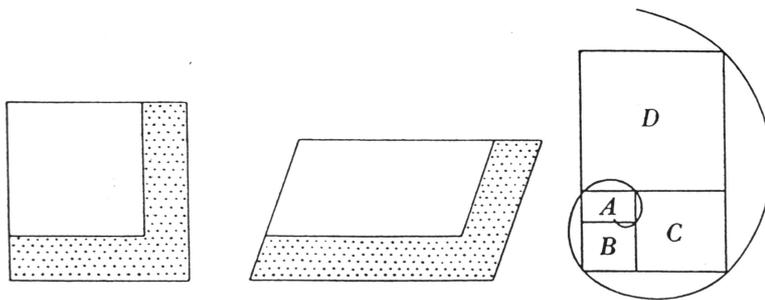


Abbildung 1.20. Gnomonische Figuren

varianz bei Skalierungsvarianz, vergleichbar den Fraktalen. Ein Gnomon ist eine Figur, die, wenn sie zu irgendeiner anderen Figur hinzugefügt wird, eine dem Original ähnliche Figur entstehen läßt. Symmetrisches Wachstum ist also gnomonisch, das heißt, "daß jeder folgende Zuwachs seinem Vorgänger ähnlich ist, und zwar ähnlich sowohl in Bezug auf die Vergrößerung wie auf die Anordnung." [Thompson, 1966].

Doch Thompson warnte vor der "mystischen Vorstellung", die Spirale selbst als Manifestation des Lebens zu sehen, da sie ja aus lebloser Substanz geformt wird. Thompson beharrte darauf, daß die Formen und Formveränderungen von Organismen, die durch Bewegung und Wachstum in Erscheinung treten, mit Hilfe physikalischer Methoden und Überlegungen, die mit mathematischen Gesetzen übereinstimmen müssen, erklärt werden können. Der Zustand und die Form eines Organismus wie auch die Veränderung dieses Zustandes bzw. der Form sind das Resultat

tat einer Anzahl von Kräften und Energien: "Die Morphologie ist nicht nur ein Studium materieller Dinge und der Form der materiellen Dinge, sondern besitzt auch die dynamischen Aspekte, in deren Rahmen wir uns mit Hilfe von Kräftebegriffen mit der Deutung von Energievorgängen befassen." [Thompson, 1966].

Die Vorstellung einer dynamischen Symmetrie, die auf den Gesetzen der Thermodynamik aufgebaut ist, war also schon vorhanden (siehe auch Jay Hambridges "Dynamic Symmetry" von 1920). Für die Entwicklung, das Wachstum und die Veränderung von Formen führte Thompson selbst mathematische Modelle ein, die berühmten cartesischen Transformationen. Unter den Fischen entdecken wir eine große Mannigfaltigkeit von solchen Deformationen.

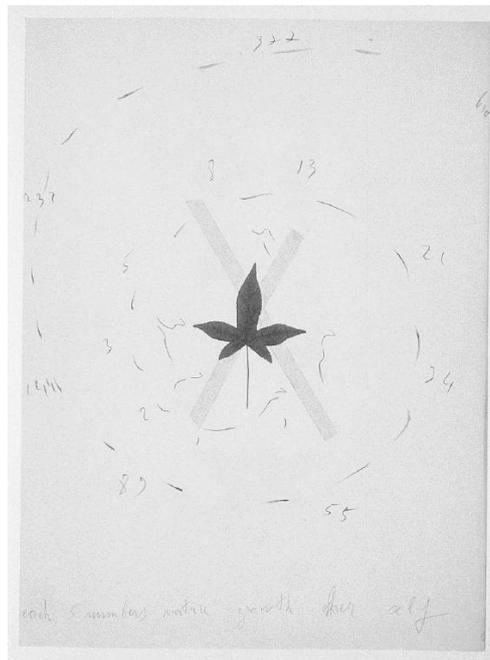


Abbildung 1.21. Mario Merz, each 5 numbers nature growth herself, 1979.

Thermodynamisch definierte Formen

Die mathematischen Aspekte und dynamischen Prinzipien der Morphologie, auf die Thompson hingewiesen hatte, verfolgte auch Matila Ghyka in

“The Geometry of Art and Life” (1946). Auch hier spielt die logarithmische Spirale eine Rolle. Ghyka verweist besonders auf die pentagonale Symmetrie als spezifisch für die vektorialen Pulsationen des Wachstums von lebenden Organismen. Er beruft sich dabei auf F.M. Jaegers “Lectures on the principle of symmetry and its applications in natural science” (1917). Aber diese Positionen, die Spiralen als Geometrie des Lebens selbst zu betrachten, sind schon von Thompson als “mystisch” widerlegt worden. Ghykas Verdienst ist es hingegen, thermodynamische Prinzipien zur Definition der Formen eingeführt zu haben. “The most general principle governing the states of equilibrium of physico-chemical systems is the ‘Criterium of Dirichlet’: in order that the equilibrium of a closed system should be stable it is sufficient that its potential energy should be (or should pass through) a minimum.”[Ghyka, 1977] Alfred J. Lotka hat 1922 auf die Bedeutung der Energieproblematik für gestaltbildende Reaktionen hingewiesen (The Energetics of Evolution).

Fraktales Wachstum: Symmetrie und Selbstsimilarität

Benoit Mandelbrots Untersuchungen zur “fraktalen Geometrie der Natur” [Mandelbrot, 1983] die aus seinem Essay “Les objets fractals: forme, hasard et dimension” (1975) hervorgingen, haben ebenfalls zum Ziel, Irregularitäten und Singularitäten der Formbildung mathematisch zu beschreiben. Die “mathematischen Monster”, die er Fraktale nannte, bilden einen Teil jener Geometrie des Wachstums, jener Formen in der Natur (Wolken, Berge, Wälder, Küsten), die besonders irregulär scheinen, doch in der Tat symmetrieähnlichen Gestaltprozessen unterliegen. Zum Beispiel könnte wegen der Identität der Irregularität in allen Skalierungen die fraktale Iteration des Sierpinski-Teppichs mit dem gnomonischen Wachstum der logarithmischen Spirale verglichen werden.

Fraktale Objekte entstehen durch sukzessive Einsetzungs- und Ersetzungsprozesse bzw. Umformungsregeln. In einen Initiator wird ein Generator eingesetzt, zum Beispiel in eine gerade Linie eine gebrochene Linie, wobei in die verbleibende gerade Linie der neuen Form wieder gebrochene Linien eingesetzt werden, also rekursiv vorgegangen wird. So entsteht eine weitere “Kurve des Lebens”, die Schneeflockenkurve, 1905 von Koch vorgeschlagen.

Mit fraktalen Objekten erreicht der esprit de la géométrie eine neue Nähe zum Wachstum der Natur, zum Beispiel zu den Pflanzen. Bäume und Pflanzen sind wegen ihrer rekursiven Verzweigungsstruktur bzw. wegen ihrer

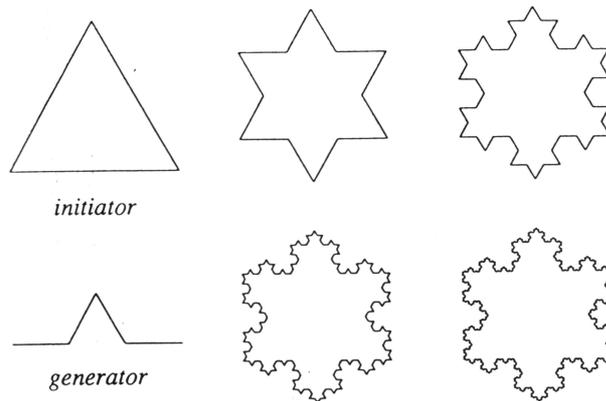


Abbildung 1.22. Schneeflockenkurve nach Koch

Bifurkation zum Teil als Fraktale darstellbar. Die fraktale Iteration, die rekursive Ersetzung gleicher Objekte, führt zu einer Struktur, wo alle Teile einer Gestalt eine geometrische Ähnlichkeit zum Ganzen haben. Diese Selbstsimilarität als Struktur zwischen den Teilen und dem Ganzen ist eine relevante Eigenschaft, wenn auch abstrahiert, von Pflanzen.

Fraktale können als vereinfachte abstrahierte Repräsentationen von realen Pflanzen betrachtet werden, wobei die rekursiven Entwicklungsmechanismen nicht mehr Symmetrie oder Spirale sind, sondern Selbstsimilarität und Entwicklungsalgorithmen. Symmetrie wird als Invarianz einer Konfiguration von Elementen gegenüber einer Gruppe von automorphen Transformationen definiert, welche Transformationen Kongruenzen sind, die durch Rotation, Reflexion und Translation erreicht werden. Könnte daher Selbstsimilarität als Spezialfall der Symmetrie unter Skalierungsoperationen betrachtet werden? Nein, aber wenn auch Selbstsimilarität schwächer ist als Symmetrie, ist Selbstsimilarität dennoch ein Paradigma für Strukturen in der natürlichen Welt, besonders in der Botanik.

Formale Grammatiken

Das rekursive Ersetzen eines Teils des Ausgangsobjektes durch (selbst)ähnliche Elemente kann auch als Umschreib- bzw. Umformungsprogramm definiert werden. Die fraktale Kurve einer Schneeflocke ist das Ergebnis einer Umschreibregel, wo rekursiv offene Polygone ersetzt werden. John Con-

ways populäres "game of life" von 1970 ist ein Umschreibemechanismus, der sich auf Felder bezieht.

Der norwegische Mathematiker Axel Thue hat 1914 mit seiner Arbeit "Probleme über Veränderungen von Zeichenreihen nach gegebenen Regeln" das erste Umformungs- bzw. Umschreibe-Programm geliefert, das über Buchstaben-Ketten operiert. Sein Problem war: Wie kann ich wissen, ob aus einem endlichen Alphabet X (zum Beispiel A, B, C, D, E, F) mit einigen Umformungsregeln der Gestalt $u \rightarrow v$ (zum Beispiel $ABA \rightarrow CA, DA \rightarrow D$) eine gegebene Zeichenkette ableitbar ist (ist $ABADAEBC$ in CCA überführbar)?

Emil Post hat 1947 die "rekursive Unlösbarkeit" des Thue-Wortproblems bewiesen, das heißt, daß es kein allgemeines Verfahren zur Lösung gibt. Semi-Thue-Systeme erwiesen sich aber in den fünfziger Jahren als Modelle zur Beschreibung formaler syntaktischer Strukturen der natürlichen Sprache sehr brauchbar (N. Chomsky). Backus und Naur verwendeten eine auf Umformung basierte Notation, um eine formale Definition der Programmiersprache ALGOL-60 (Algorithmic Language) zu besorgen. Umformungsprogramme, Umschreibesysteme von Zeichenketten waren also gleichsam linguistische Wachstumsprogramme. 1968 führte der Biologe Aristid Lindenmayer neue Zeichenketten-Umschreibemechanismen ein, die später sogenannten L-Systeme, wo alle Buchstaben eines Wortes parallel und simultan, nicht sequentiell wie bei Chomsky, ersetzt werden. Die einfachste Klasse der L-Systeme sind die deterministischen kontextfreien L-Systeme, DOL-Systeme genannt. Wenn wir eine Zeichenkette aus den Buchstaben a und b haben und die Umformungsregeln $a \rightarrow ab$ und $b \rightarrow a$ und das Axiom b gegeben sind, dann erhalten wir am Ende der Entwicklung die in der Abbildung 1.23 wiedergegebene Sequenz.

Dieser Formalismus kann benutzt werden, um die Entwicklung von Zellen zu simulieren. Die komplexen Formalismen der L-Systeme werden zum mathematischen Instrument, Wachstumsprozesse zu simulieren.

Przemslaw Prusinkiewicz und Aristid Lindenmayer haben die "algorithmische Schönheit der Pflanzen" [Prusinkiewicz und Lindenmayer, 1990] gezeigt. Die Sprache der Mathematik beschreibt besser denn je die Entwicklung von Pflanzen. Mit Hilfe der auf den L-Systemen aufgebauten formalen Produktionssystemen können Pflanzen im Computer perfekt simuliert werden. Geometrische Eigenschaften von Pflanzen wie bilaterale Symmetrie der Blätter, Rotationssymmetrie der Blumen etc. werden durch die Eigenschaft der Selbstsimilarität (eine schwächere Symmetrie) ergänzt. Selbstähn-

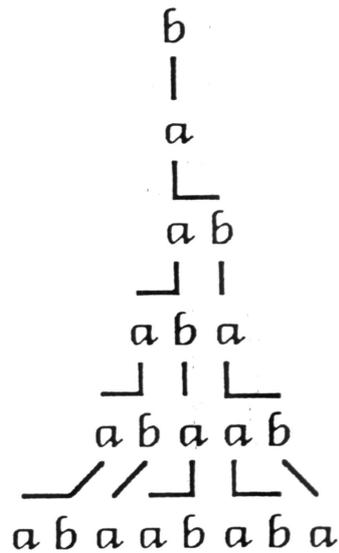


Abbildung 1.23. Beispiel einer Ableitung in einem DOL-System.

lichkeit entsteht dort, wo ein Teil einer Form geometrisch ähnlich dem Ganzen der Form ist. Selbstsimilarität verbindet daher Pflanzenstrukturen mit fraktaler Geometrie. Selbstsimilarität in Pflanzen ist das Ergebnis von Entwicklungsprozessen, welche "die Form eines Organismus zu einem Ereignis in Raum und Zeit" (Thompson) machen. Nicht mehr die Form von Organismen der Natur wird statisch visuell simuliert, sondern das Gesetz des dynamischen Wachstums dieser Organismen wird mathematisch simuliert. Wir können den Prozeß des Wachsens simulieren, nicht nur dessen Ergebnis: die Form. Diese Entwicklungsalgorithmen werden vom Formalismus der L-Systeme erfaßt, die ein Teil der Theorie der formalen Sprache sind. Sie spielen heute im computergestützten Film, in der Kunst, in der Architektur bei morphologischen Prozessen eine zentrale Rolle.

Die Sprache der Natur wird in der Tat zu einer Sprache, zu einer formalen Sprache. Symmetrie, Selbstsimilarität, Fraktale etc. werden zu Elementen einer solchen formalen Sprache, die dem Verständnis dynamischer Prozesse des Wachstums und der Selbstorganisation dient.



Abbildung 1.24. Christa Sommerer und Laurent Mignonneau: Phototrophie, Eine interaktive Echtzeit-Computerinstallation, 1994

Anmerkungen

- ¹ Das altchinesische "sakrale" Rechnen begann mit 2 (- -, weibliche oder weiche Linie) und 3 (-, männliche oder feste Linie).

Daraus leitet sich auch die Bewertung der Linien bei der Deutung des Hexagramms ab, den "acht Urbildern nach König Wen".

Leibniz hat übrigens 1703 die Vermutung aufgestellt, die altchinesischen Zahlentheorien seien ein sinnvolles Ordnungsprinzip des Weltbildes auf binärarithmetischer Basis: "Erklärung der binären Arithmetik, die sich einzig der Zahl-Zeichen 0 und 1 bedient; mit Bemerkungen über ihre Nützlichkeit und über den Sinn, den sie den alten chinesischen Zeichen Fo-his verleiht". (G. W. Leibniz: Zwei Briefe über das binäre Zahlensystem und die chinesische Philosophie. Belsler Presse 1969). "Das Überraschende daran ist, daß diese Arithmetik mit 0 und 1 den Schlüssel liefert zum Geheimnis der Linien-Zeichen eines alten Königs und Philosophen, genannt FO-HI, der vor mehr als viertausend Jahren gelebt haben soll und den die Chinesen als den Gründer ihres Reiches und ihrer Wissenschaft betrachten. Es gibt einige Linien-Zeichen, die man ihm zuschreibt" (Leibniz bezieht sich auch die Pa-kua, die "Acht Urbilder" oder Trigramme des FU HSI. Der legendäre Kulturschöpfer Chinas FU HSI soll zwischen 2953 und 2838 v. Chr. gelebt haben).

Leibniz fährt fort: "Sie haben alle Bezug auf diese Arithmetik; man braucht nur das sogenannte Acht-Cova-Zeichen einzusetzen, das als Grundzeichen gilt, und die Erklärung anzufügen, die ins Auge springt, nämlich daß erstens eine durchgehende Linie (—) eine Einheit oder 1 bedeutet und daß zweitens eine unterbrochene Linie (— —) für Null oder 0 steht. Die Chinesen wissen seit tausend Jahren nicht mehr, was die Cova-oder Linien-Zeichen des FO-HI bedeuten; sie haben Kommentare darüber verfaßt, in denen sie einen, ich weiß nicht wie weit hergeholten Sinn für diese Zeichen suchten, so daß die RICHTIGE ERKLÄRUNG JETZT VON DEN EUROPÄERN KOMMEN MUSSTE".

Maria-Louise von Franz, Schülerin von C. G. Jung, vertritt die Theorie, daß die altchinesische Zahlenauffassung mit der Idee des Zahlenfeldes verknüpft ist, in dem die einzelnen Zahlen als "rhythmische Konfigurationen" auftreten: "In den entsprechenden 'Weltmodellen' und mathematischen Gottesbildern dominiert die Bedeutung der ersten vier Zahlen in besonderem Maße, ebenso in den systematisierten Divinationstechniken der Vergangenheit". Im Klappentext ihres Buches "Zahl und Zeit. Psychologische Überlegungen zu einer Annäherung von Tiefenpsychologie und Physik" (Stuttgart [Klett] 1970) wird bemerkt: "Der Archetypus (im Sinne C. G. Jungs) wird in seinem Ordnungspaket, als welcher sich die Zahl erweist, zu einer neuen naturwissenschaftlich beschreibbaren Grundlage, die einer Reihe von Disziplinen gemeinsam ist". Übrigens stammt auch das älteste magische Quadrat aus China. Es wurde dem Kaiser Yü durch die heilige Schildkröte vom Flusse Lo zugebracht:

4	9	2
3	5	7
8	1	6

- ² Beachtenswert sind die Überlegungen, die Bombieri hierzu vorbringt: "There are very many old problems in arithmetic whose interest is practically rich, e.g. the existence of infinitely many Fermat primes $2^{2^n} + 1$ etc. Some of the questions may very well be undecidable in arithmetic; the construction of arithmetical models in which questions of this type have different answers would be of great importance". (in: Browder, Hg.,: Mathematical Developments arising from Hilbert's Problems. Proc. Symp. Pure Math. 28 (1976), American Mathematical Society, II, A, S. 36).
- ³ Doch auch in der Gegenwart zeitigten die Eigenschaften der Fibonacci-Zahlen noch sehr brauchbare Resultate. Matyasevic hat in seiner berühmten (negativen) Lösung des 10. Hilbertschen Problems die Reihe der Fibonacci-Zahlen wesentlich benützt, da diese exponentiell (also stark) wächst und diophantisch definierbar ist. Die Fibonacci-Reihe ist die erste bekannt gewordene exponentiell wachsende Folge in der Literatur. Diesem historischen Faktum verdankt sie viel von ihrer Faszination und Stellung.
- ⁴ "Musik ist eine verborgene arithmetische Übung des seines Zählens unbewußten Geistes", lautet die Definition der Musik bei Leibniz. Bachs Goldberg-Variationen (1742) für Cembalo, dreißig Variationen, auf einem durchgehenden Passacaglia-Bass aufgebaut, sind ein berühmtes Beispiel für musikalisch-mathematische Proportionen. Vergleiche auch die rationale Ästhetik der homophonen Kompositionstechniken der monodischen Harmonielehre, welche die Einführung des Generalbasses in die Musikgeschichte mit sich brachte, bei Monteverdi, dem Begründer der modernen Musik.
- ⁵ Da wir die Geduld des Lesers nicht weiter beanspruchen wollen, beenden wir hier diese kleine Exkursion in die Zahlentheorie mit einem Ausblick auf die noch möglichen Felder und Probleme, als da sind: die Abbildung von verschiedenen Quadraten in einem Quadrat, die numerischen magischen Quadrate, die Inversion des Kreises, die Fermatschen Zahlen, Transformationen, Partial-Funktionen, numeri idonei, die Wurzelschnecke, das harmonische Dreieck, die Pascal'sche Schnecke, das Diagonalverfahren, die Zahl der Universalbibliothek, die Reihen, die logarithmische Spirale, die Symmetrie, die Möbius-Schleife, usw. All diese numerischen und geometrischen Merkwürdigkeiten haben in der Geschichte der bildenden Kunst, Architektur und Musik ihre tiefgreifende Wirkung gezeitigt, besonders im 20. Jahrhundert.

Literatur

- [Basieux, 1999] Basieux, P. (1999). *Abenteuer Mathematik*, Rowolt.
- [Coleman und Holmes, 1988] Coleman, W. und Holmes, F. L. (1988). *The Investigative Enterprise. Experimental Physiology in Nineteenth Century Medicine*. Univ. California Press.
- [Conway und Guy, 1996] Conway, J. H. und Guy R. K. (1996). *The Book of Numbers*. Springer, New York.
- [Cook, 1979] Cook, T.A. (1979). *The Curves of Life*. Nachdruck der Ausgaben 1914 bei Dover Publ., N.Y.
- [Derlin, 1998] Derlin K. (1998). *Life by the Numbers*. Wiley.
- [Deutsch, 1985] Deutsch, D. (1985). Quantum theory, the Church-Turing principle and the universal quantum computer. *Proc. of the Royal Society of London*, 400, S. 97.
- [Frängsmyr et al., 1984] Frängsmyr, T., Heilbron, J. L. und Rider, R. E. (1984). *The Quantifying Spirit in the Eighteenth Century*. Univ. of California Press.
- [Ghyka, 1977] Ghyka, M. (1977). *The Geometry of Art and Life*. Dover, N.Y. Seite 85.
- [Hangenmaier, 1977] Hangenmaier, O. (1977). *Der Goldene Schnitt*, Moos Verlag, München 1977.
- [Hoffman, 1998] Hoffman P. (1998). *The Man Who Loved Only Numbers*. Fourth Estete, London.
- [Hug-Hellmuth, 1916] Hug-Hellmuth, H.v. (1916). *Einige Beziehungen zwischen Erotik und Mathematik*, S. Freud, Hrsg., Hugo Heller Verlag, Bd. 4, S. 52.
- [Mandelbrot, 1983] Mandelbrot, B.B. (1983). *The Fractal Geometry of Nature*. Freeman, N.Y.
- [Prusinkiewicz und Lindenmayer, 1990] Prusinkiewicz, P. und Lindenmayer, A. (1990). *The Algorithmic Beauty of Plants*. Springer, N.Y.
- [Singh, 1998] Singh, S. (1998). *Fermats letzter Satz*. Maurer, München.

- [Stewart, 1995] Stewart, I. (1995). *Nature's Numbers*. Basic Books.
- [Thom, 1972] Thom, R. (1972). *Théorie Dynamique de la Morphogenèse*. Benjamin, N.Y.
- [Thom, 1974] Thom, R. (1974). *Modèles Mathématiques de la Morphogenèse*. Paris.
- [Thompson, 1966] Thompson, D'Arcy (1966). *Über Wachstum und Form*. Basel 1973. Seite 223. Übersetzung der Ausgabe von 1966, Cambridge Univ. Press.
- [Turing, 1952] Turing, A. M. (1952). The Chemical Basis of Morphogenesis. *Phil. Trans. R. Soc. London, Ser. B* 327, 37.

Die Kunst, die Gunst des Zufalls zu erzwingen

Raimund Seidel

Einleitung

Der Zufall spielt eine erhebliche Rolle in unserem Leben. Von der Genese unserer eigenen Person (welche Samenzelle befruchtet das Ei) über die Entwicklung unserer Persönlichkeit (welche Menschen treffen wir, welche Bücher und Artikel fallen uns in die Hände) bis hin zum Tode (zur falschen Zeit am falschen Ort), fast immer scheinen Zufälle entscheidend mitzuwirken. Wir leben mit dieser Tatsache, aber es ist uns nicht immer wohl dabei, und wir wehren uns dagegen. Ein Teil des menschlichen Fortschritts kann als erfolgreicher Versuch gesehen werden, den Zufall zu beseitigen oder zumindest seine unpässlichen Konsequenzen zu lindern.

Wie kann man "Zufall beseitigen"? Das geht dann, wenn die Zufälligkeit eines Phänomens nur auf der Unwissenheit des Beobachters beruht oder auf dessen Unvermögen, tiefere Ursachen zu erkennen oder ihre Konsequenzen abzuschätzen. So wird zum Beispiel einem Binnenländer, der ahnungslos zum Meer kommt, die Gezeitenstärke vorerst zufällig und unvorhersagbar erscheinen. Sobald ihm aber der Zusammenhang zu Mond und Sonne hergestellt wird, ist die scheinbare Zufälligkeit dieses Phänomens vollkommen verschwunden.

Als zweites Beispiel nehmen wir eine Person, die sich auf eine Waldlichtung stellt und den Vogelflug beobachtet. Anfangs werden die beobachteten Flüge wohl zufällig erscheinen. Aber mit längerer Beobachtung werden sich Regelmäßigkeiten herauskristallisieren. Wenn die Person nun auch immer mehr über die Lage von Nestern und Futterquellen herausfindet und auch die grundsätzlichen Lebensgewohnheiten der einzelnen Vogelarten kennenlernt, dann wird ihr es immer leichter fallen, richtige Voraussagen über anstehende Vogelflüge zu machen.

Wie soll es aber möglich sein, "den Zufall zu nutzen"? Im vorangegangenen Beispiel könnte ein Langzeitbeobachter der Waldlichtung durch seine guten Vorhersagen einen neu angekommenen Beobachter, dem die Vogelflüge noch zufällig erscheinen, verblüffen und der Langzeitbeobachter

könnte sich so vielleicht auch Vorteile erzielen. Den Auguren der alten Römer wird das wohl nicht ganz unbekannt gewesen sein. Aber das kann man kaum "Nutzung des Zufalls" nennen, sondern eher Übertölpelung.

Tatsächliche Nutzung, so wie wir sie meinen, zeigt sich eher im folgenden Beispiel: Ein junger Mann muß zwischen zwei Freundinnen wählen, aber er kann sich nicht entscheiden. Er fragt seine Mutter um Rat. Sie aber bevorzugt insgeheim eine der beiden Freundinnen und sagt zum Sohn: Geh nachmittags zur U-Bahnstation. Deine beiden Freundinnen wohnen doch in verschiedenen Richtungen. Nimm den ersten Zug, der kommt. Die Freundin, zu der er dich bringt, nimm zur Frau. Dem Sohn scheint das einleuchtend. Nachmittags verkehren die Züge in beide Richtungen im 20 Minuten Takt. Wenn er also zu einem zufälligen Zeitpunkt zur Station käme, wäre das Zuerstkommen eines Zuges für beide Richtungen gleichwahrscheinlich, und damit käme es so zu einer fairen zufälligen Wahl zwischen den beiden Alternativen. Dem ist aber nicht so. Denn die Züge in Richtung A kommen 13, 33 und 53 Minuten nach der vollen Stunde in der Station an, und die Züge in Richtung B um 15, 35 und 55 Minuten nach der vollen Stunde. Damit gibt es in der Stunde nur sechs Minuten (13–15, 33–35 und 53–55), in denen zuerst ein Zug in Richtung B zur Station kommt, die übrigen 54 Minuten kommt zuerst ein Zug in Richtung A. Die Freundin in Richtung A hat also neunmal bessere Chancen gewählt zu werden. Die Mutter wusste genau, warum sie zu dieser Art des zufälligen Wählens riet.

Die Mutter nutzt hier den Zufall aus. Aber ist das nicht wieder einfach Übertölpelung so wie vorher beim Vogelflugbeispiel? Bei genauerem Hinsehen merkt man aber, daß die Sache hier ganz anders ist. Es dreht sich überhaupt nicht darum, ein zufälliges Ereignis, nämlich den Ankunftszeitpunkt des Sohnes bei der U-Bahnstation, vorauszusagen. Das Wesentliche ist, daß die Entscheidungsmethode so entworfen wurde, daß bei einer zufälligen (gleichverteilten) Ankunftszeit des Sohnes eine bestimmte der beiden Alternativen wesentlich wahrscheinlicher zum Ergebnis der Entscheidung wird.

Dies ist genau die Art und Weise, wie in der Informatik "die Gunst des Zufalls genutzt" wird. Innerhalb von Methoden und Algorithmen lässt man es zu Zufallsentscheidungen kommen, aber so, daß es mit hoher Wahrscheinlichkeit zu einem positiven Gesamtergebnis kommt. Man nennt diesen kontrollierten Einsatz von Zufall das *Randomisieren*. Wie das in etwa wirklich passiert, soll in den folgenden Kapiteln exemplarisch gezeigt werden.

Dieser Aufsatz richtet sich an ein breites Publikum und ist daher zum

größten Teil bewusst einfach gehalten. Für eine technisch anspruchsvollere, sowie auch breitere Behandlung des Einsatzes des Zufalls in der Informatik sei der interessierte Leser auf das Buch von Motwani und Raghavan [Motwani und Raghavan, 1995] oder auf den Übersichtsartikel von Karp [Karp, 1991] verwiesen.

Den Zufall kontrollieren: ein Brückenproblem

In diesem Beispiel soll klar werden, was mit "kontrolliertem Einsatz von Zufall" gemeint ist. Wir wollen uns zwar auf den Zufall verlassen, seine Quelle darf aber nicht ganz beliebig sein, sie muß unserer Kontrolle unterliegen.

Als sehr idealisiertes Beispiel, oder vielleicht Gedankenexperiment, stelle man sich eine große Stadt vor, an der ein Fluss vorbeiführt. Über diesen Fluss gebe es zwei Brücken, nennen wir sie B_0 und B_1 , und auf der anderen Seite des Flusses liege ein großes Stadion. Es sei Samstagnachmittag, eine Veranstaltung finde im Stadion statt, 20 000 Autos wollen aus der Stadt zum Stadion.

Problem: Wie kann man erreichen, daß beide Brücken gleich ausgelastet werden, also jeweils von ungefähr gleich vielen Autos benutzt werden?

Leser, denen dieses Problem zu wirklichkeitsfremd erscheint, mögen Stadt und Stadion jeweils durch ein Datennetz ersetzen, die beiden Brücken durch zwei die beiden Datennetze verbindende Kabel, und die Autos durch Nachrichten von einem Netz ins andere, die sich jetzt jeweils für eines der beiden Verbindungskabel entscheiden müssen.

Lösung 1 (laissez faire): Man macht nichts und überlässt es einfach jedem Autofahrer, nach seiner Wahl Brücke B_0 oder Brücke B_1 zu befahren.

Diese Lösung verlässt sich auf den Zufall und funktioniert wohl auch erfahrungsgemäß recht gut. Aber kann man irgendeine Art von Garantie über die gleiche Auslastung der Brücken abgeben? Nur schwer, oder nur, wenn man statistische Annahmen über das Verhalten der Autofahrer macht. Solche Annahmen mögen wohl fundiert sein, sie können aber wegen Baustellen, Gerüchten und aus anderen Gründen leicht verletzt werden. Bei der Datennetzversion unseres Beispiels kann es sogar schwierig sein, vernünftig fundierte statistische Annahmen zu machen. Fazit ist, daß man sich hier auf den Zufall verlässt, seine Quelle aber nicht unter Kontrolle hat und damit nur sehr beschränkt Garantien über die Gleichauslastung der Brücken abgeben kann.

Lösung 2 (deterministisch–diktatorisch): Vor seiner Abfahrt muß jeder Autofahrer eine zentrale Leitstelle anrufen, die ihm mitteilt, ob er Brücke B_0 oder B_1 verwenden soll. An diese Anweisung muß sich der Autofahrer auch halten.

Mit dieser Lösung kann sicher eine gleiche Auslastung der beiden Brücken erreicht werden. Aber diese Lösung ist offensichtlich nicht ideal, um nicht zu sagen, sehr teuer: jeder muß telefonieren, und es muß diese zentrale Leitstelle geben.

Lösung 3 (randomisiert): Vor seiner Abfahrt muß jeder Autofahrer eine Münze werfen. Kommt Zahl, muß er die Brücke B_0 verwenden, kommt Wappen, die Brücke B_1 .

Diese Lösung verursacht offensichtlich kaum Kosten, wie es die deterministische Lösung 2 tut. Man muß auch kein Wahrscheinlichkeitstheoretiker sein, um zu sehen, daß, wenn sich jeder an diese Münzwurfregel hält, es mit hoher Wahrscheinlichkeit zu einer fast gleichen Brückenauslastung kommt. Und mit ein klein wenig Mühe kann man auch quantifizierte Aussagen machen, wie z.B. die Wahrscheinlichkeit ist kleiner als 0.024 (oder 0.0000092), daß eine der Brücken von mehr als 10 300 (oder 10 500) der 20 000 Autos verwendet wird. Diese Wahrscheinlichkeitsgarantien beruhen hier auf keinerlei statistischen Annahmen über Autofahrer (außer, daß sie sich an die vorgegebene Regel halten), sondern nur auf der Annahme des fairen Münzwurfes, daß also Zahl und Wappen jeweils mit Wahrscheinlichkeit $1/2$ auftreten.

Diese randomisierte Lösung verläßt sich auf den Zufall, aber die Quelle des Zufalls ist genau bekannt und unter Kontrolle.

Zufällige Ordnung und ein geometrisches Optimierungsproblem

Das Brückenbeispiel des vorigen Kapitels hat vielleicht für manche wenig mit Berechnen zu tun, sondern "nur" mit der Koordination von unabhängigen Agenten. In diesem Kapitel wollen wir ein Beispiel geben, bei dem es sich ganz klar ums Berechnen handelt, und wo der Zufall verwendet wird, um die Berechnung zu beschleunigen.

Problem: Finde Für eine Menge S von n Punkten in der Ebene den kleinsten umschriebenen Kreis, also den kleinsten Kreis, der alle Punkte in S enthält.

Für manchen mag jetzt gar nicht klar sein, wo denn das Problem überhaupt steckt. Betrachtet man zum Beispiel die Punktmenge in Abbildung 2.1(a), dann "sieht man doch ganz klar", daß der kleinste umschriebene

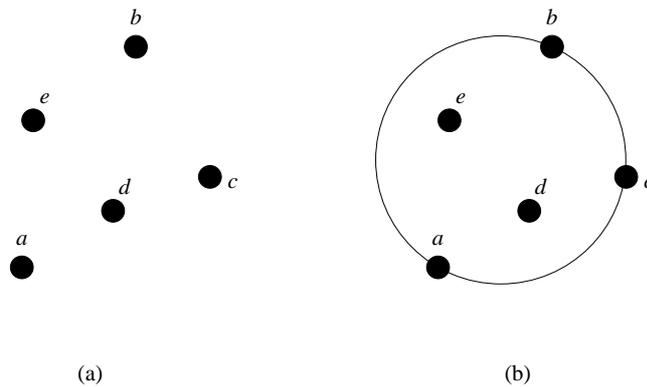


Abbildung 2.1.

Kreis der durch die Punkte a , b , und c definierte ist, so wie es in Abbildung 2.1(b) gezeigt ist.

Aber so einfach ist es nun doch nicht. Denn typischerweise bekommt man die Punktmenge S nicht gezeichnet, sondern man bekommt die einzelnen Punkte durch ihre Koordinaten angegeben, also ein Zahlenpaar pro Punkt. Hier wäre ein Beispielmenge von 20 Punkten:

(34.20, 9.90) (41.85, 12.60) (45.45, 15.30) (47.70, 19.35)
 (48.60, 23.85) (48.60, 32.85) (46.80, 36.90) (46.35, 40.50)
 (40.50, 42.75) (34.65, 44.10) (29.25, 44.10) (24.75, 42.75)
 (18.45, 38.70) (14.85, 31.95) (13.95, 28.35) (14.40, 22.50)
 (15.75, 19.35) (18.00, 14.40) (23.85, 11.25) (26.55, 9.90)

Wenn man sich die Mühe macht, diese Punkte aufzuzeichnen, wie in Abbildung 2.2 getan, kommt man drauf, daß die Punkte alle schon fast auf einem Kreis liegen, aber eben nicht genau auf einem Kreis. Welches jetzt *genau* der gesuchte kleinste umschriebene Kreis ist, das sieht man gar nicht mehr ganz klar.

Mathematisch geneigte Leser mögen jetzt vielleicht sagen, so schwer sei das Problem aber trotzdem nicht: Drei nicht auf einer Geraden liegende Punkte a, b, c , definieren genau einen Kreis $K_{a,b,c}$, auf dem sie alle drei liegen (den Umkreis des Dreiecks a, b, c). Zwei Punkte a, b definieren genau einen "Diametralkreis" $D_{a,b}$, auf dem sie beide liegen und dessen Durchmesser sie aufspannen (siehe Abbildung 2.3). Der kleinste umschriebene Kreis der Ausgangspunktmenge S kann nur ein Kreis der Form $K_{a,b,c}$ sein (wie in Abbildung 2.1) oder der Form $D_{a,b}$ (wie in Abbildung 2.3), mit a, b, c Punkte

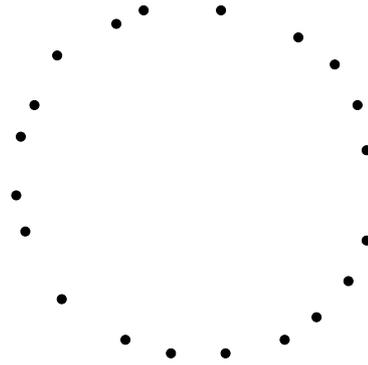


Abbildung 2.2.

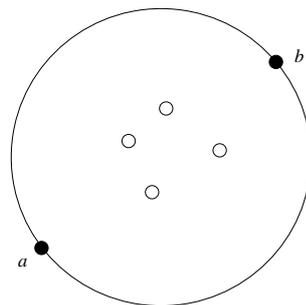


Abbildung 2.3.

in S . Nennen wir diese zwei oder drei Punkte, die den kleinsten umschriebenen Kreis von S bestimmen, die *Umpunkte* von S .

Um den kleinsten umschriebenen Kreis (bzw. die Umpunkte) von S zu finden, genügt es also, für jedes mögliche Paar a, b und jedes mögliche Tripel a, b, c von Punkten in S zu testen, ob der Kreis $D_{a,b}$ bzw. $K_{a,b,c}$ die Menge S umschreibt (alle Punkte von S enthält), und von den so gefundenen umschriebenen Kreisen den kleinsten auszuwählen.

Der mathematisch geneigte Leser hat damit recht, daß sich der kleinste umschriebene Kreis so bestimmen lässt. Aber geht das wirklich "leicht"? Bei 20 Punkten gibt es schon 190 Paare a, b und 1140 Tripel zu betrachten. Das ist für die Geduld eines Menschen zu viel. Aber für einen modernen Computer sollte das doch kein Problem sein. Das stimmt. Aber wenn man das Problem für 20 000 Punkte lösen möchte, dann gibt es rund 200 Millionen Paare zu betrachten und mehr als eine Billion Tripel! Nehmen wir

an, ein Computer könnte pro Sekunde ein Million Tripel abarbeiten — derzeit eine zweifelhafte Annahme, da ja für jedes Tripel a, b, c getestet werden muß, ob die anderen 19 997 Punkte alle innerhalb des Kreises $K_{a,b,c}$ liegen — so bräuchte der Computer allein für das Abarbeiten aller Tripel schon mehr als eine Million Sekunden. Das sind mehr als elf Tage! So lange möchte man wohl nicht warten, um den kleinsten umschriebenen Kreis von 20 000 Punkten zu finden.

Wie kann nun der Zufall helfen, dieses Problem schneller zu lösen? Die Idee ist folgende: Machen wir einmal das 20 000 Punkte Problem ein bisschen einfacher, indem wir einen Punkt, nennen wir ihn p , wegnehmen. Nehmen wir an, Kreis K' wäre die Lösung für das etwas einfachere 19 999 Punkte Problem und wir wüssten K' . Wie können wir nun den Lösungskreis K für das 20 000 Punkte Problem erhalten? Da gibt es nun zwei Fälle: Der *gute* Fall wäre, daß der weggenommene Punkt p im Kreis K' liegt. Dann ist nämlich K' nicht nur der kleinste umschriebene Kreis der 19 999 Punkte, sondern auch der kleinste umschriebene Kreis aller 20 000 Punkte. Also haben wir mit $K = K'$ die Gesamtlösung gefunden. Der *schlechte* Fall wäre, daß der Punkt p nicht im Kreis K' liegt. Dann haben wir zwar nicht den Gesamtlösungskreis K gefunden, aber wir haben wichtige Informationen über ihn gewonnen, denn p muß jetzt auf seinem Rand liegen. Das heisst, p ist ein Umpunkt der 20 000 Punkte und K kann jetzt nur noch von der Form $D_{a,p}$ oder $K_{a,b,p}$ sein.

Jetzt kommt der Zufall ins Spiel. Wenn p zufällig aus den 20 000 Punkten ausgewählt wird, dann ist es recht unwahrscheinlich, daß der *schlechte Fall* eintritt. Denn eine "schlechte" Wahl für p wären nur die Umpunkte, und davon gibt es unter den 20 000 entweder zwei oder drei, also höchstens¹ drei. Das heisst, der schlechte Fall tritt mit einer Wahrscheinlichkeit von höchstens $3/20\,000$ ein.

Diese Überlegungen führen zu folgender *randomisierter* Methode zur Bestimmung des kleinsten umschriebenen Kreises K einer Menge S von n Punkten.

1. (Randomisiere) Bringe die Punkte von S in eine zufällige Reihenfolge p_1, p_2, \dots, p_n .
2. (Initialisiere) Sei K der Diametralkreis von p_1 und p_2 .
3. (Inkrementiere) Für jedes i von 3 bis n mache folgendes:

Falls p_i nicht im Kreis K liegt, dann löse das Unterproblem, den kleinsten umschriebenen Kreis von $\{p_1, \dots, p_{i-1}\}$ zu finden, der p_i auf dem Rand hat. Von nun an sei K der Lösungskreis dieses Unterproblems.

Wie löst man das Unterproblem? Ganz analog zum ursprünglichen Problem, nur mit zwei kleinen Änderungen: erstens die Initialisierung, da K jetzt den Punkt p_i auf dem Rand haben muß, und zweitens die Inkrementierung, bei der sich das Unter-Unterproblem ergibt, den kleinsten umschriebenen Kreis zu finden, der zwei vorgegebene Punkte auf dem Rand hat. Man kann sich leicht davon überzeugen, daß dieses Unter-Unterproblem leicht zu lösen ist.

Die interessante Frage ist aber jetzt, wie schnell die vorgeschlagene Methode ist. Das hängt natürlich ganz stark davon ab, wieviel Zeit man für die Lösung eines Unterproblems braucht, und natürlich auch davon, wie oft der schlechte Fall eintritt, daß ein Unterproblem gelöst werden muß.

Nehmen wir an, daß ein Computer zur Unterproblemlösung für i Punkte $3i$ Mikrosekunden braucht. Wie oft Unterprobleme gelöst werden müssen, hängt natürlich von der im Schritt 1 gewählten Reihenfolge der Punkte ab. Ist sie schlecht, dann könnte es eintreten, daß für jedes i ein Unterproblem gelöst werden muß. Ist sie gut, muß vielleicht überhaupt kein Unterproblem gelöst werden. Das Interessante ist nun, daß man von einer zufällig gewählten Reihenfolge erwarten kann, daß sie ziemlich gut ist.

Hier ist die Argumentation: Wir fragen uns, was ist der *erwartete* Zeitaufwand für den i -ten Durchgang der Iteration? Wenn die Reihenfolge zufällig gewählt wurde, dann ist es für jeden der ersten i Punkte in der Reihenfolge gleichwahrscheinlich an i -ter Stelle zu stehen. Das heisst, mit Wahrscheinlichkeit höchstens $3/i$ steht einer der höchstens drei Umpunkte von $\{p_1, \dots, p_i\}$ an letzter Stelle, und damit tritt der schlechte Fall, daß ein Unterproblem gelöst werden muß, höchstens mit Wahrscheinlichkeit $3/i$ ein. Der Zeitaufwand für die Lösung dieses Unterproblems ist dann, wie oben angenommen, $3i$ Mikrosekunden. Der *erwartete* Aufwand ist damit $(3i) \cdot (3/i)$, also 9 Mikrosekunden. Zählen wir jetzt noch (konservativ) eine Mikrosekunde für den Test, ob p_i im Kreis K liegt, dazu, ergibt sich als *erwarteter* Zeitaufwand für den i -ten Durchgang der Iteration 10 Mikrosekunden. Alle $n - 2$ Durchgänge zusammen brauchen daher *im Erwartungswert* weniger als $10n$ Mikrosekunden. Nehmen wir noch n Mikrosekunden für das Herstellen der zufälligen Reihenfolge in Schritt 1 dazu, dann ist die *erwartete*

Gesamtzeit dieser Methode für das Bestimmen des kleinsten umschriebenen Kreises von n Punkten weniger als $11n$ Mikrosekunden.

Für 20 000 Punkte braucht diese randomisierte Methode "im Durchschnitt" also nur 220 000 Mikrosekunden, also etwas weniger als eine Viertelsekunde, und nicht über 11 Tage.

Der aufmerksame Leser wird jetzt fragen, wie realistisch die Annahme sei, daß ein Unterproblem für i Punkte in $3i$ Mikrosekunden gelöst werden könne. Das ist ziemlich realistisch, wenn man mit "Lösungszeit" die *erwartete* Lösungszeit meint. Denn für die vorgeschlagene Lösung des Unterproblems kann man die gleiche Analyse anwenden wie eben, nur daß jetzt die Wahrscheinlichkeit höchstens $2/i$ ist, daß der schlechte Fall eintritt und ein Unter-Unterproblem gelöst werden muß. Dieses Unter-Unterproblems kann man für i Punkte realistisch sehr leicht in i Mikrosekunden lösen.

Genauere Einzelheiten findet man in der Arbeit von Welzl [Welzl, 1991], in der diese Methode vorgestellt wurde. Dort wird auch gezeigt, daß sich diese Methode für das analoge Problem im 3-dimensionalen Raum, finde die kleinste umschriebene Kugel von n Punkten, bzw. auch im d -dimensionalen Raum verwenden läßt. Allerdings ist der erwartete Zeitaufwand dann proportional zu $d!n$, was nur für recht kleine d erträglich ist. Verfeinerte und allgemeinere Methoden findet man in Arbeiten von Matoušek, Sharir und Welzl [Matoušek et al., 1996] und von Gärtner [Gärtner, 1995]. Der Letztere stellt auch unter [Gärtner, 1999] Software zur Verfügung.

Zufällige Zeugen und Primzahlbestimmung

Eine natürliche Zahl $N > 1$ heißt *Primzahl*, wenn sie nur durch 1 und durch sich selbst teilbar ist. Eine Zahl, die keine Primzahl ist, heißt *zusammengesetzt*, denn sie ist darstellbar als $N = A \cdot B$ mit $1 < A, B < N$.

In diesem Abschnitt geht es um das Problem, für eine vorgegebene Zahl N festzustellen, ob sie eine Primzahl ist oder nicht.

Die Definition von Primzahl gibt eigentlich schon eine Möglichkeit an, wie man das Problem lösen kann: man probiert einfach für jede Zahl $A = 2, 3, 4, \dots, N - 1$, ob A die Zahl N teilt. Man kann das auch noch etwas beschleunigen. Erstens braucht man nach $A = 2$ nur die ungeraden Zahlen zu probieren. Zweitens reicht es, nur Zahlen bis höchstens \sqrt{N} zu probieren. Denn wenn N zusammengesetzt ist, also $N = A \cdot B$, dann ist sowohl A als auch B Teiler von N und es können nicht beide größer als \sqrt{N}

sein, denn dann gälte $A \cdot B > N$.

Selbst diese beschleunigte Methode ist leider nicht sehr schnell, außer N ist relativ klein. Wenn man sie auf die 30-stellige Zahl $N = 10^{30} - 11$ anwendet, dann müssen ungefähr $0.5 \cdot 10^{15}$ Teilbarkeitstests durchgeführt werden, um festzustellen, daß N prim ist. Nehmen wir an, ein Computer könne ein Million "primitive" Operationen pro Sekunde durchführen, wobei primitive Operationen so etwas sind wie Multiplikation zweier Zahlen oder deren Division mit Restbildung (de facto der Teilbarkeitstest) und wobei die Operanden, sagen wir, höchstens 100-stellige Zahlen sind. Auf so einem Computer bräuchte dann diese "beschleunigte" Methode zum Testen von $N = 10^{30} - 11$ ungefähr $0.5 \cdot 10^9$ Sekunden. Das sind mehr als 15 Millionen Jahre! Da würde es auch nicht viel nützen, wenn der Computer tausendmal schneller wäre. In der modernen Kryptographie verwendet man aber routinemäßig Primzahlen mit mehreren hundert Stellen. Wie kann man so große Zahlen auf Primalität testen?

Zuerst einmal ein klein wenig Zahlentheorie: Ein Satz von Fermat besagt, wenn N eine Primzahl ist, dann gilt für jede ganze Zahl X mit $1 \leq X < N$

$$X^{N-1} \bmod N = 1.$$

Dabei bedeutet $U \bmod V$ der Rest bei der ganzzahligen Division von U durch V . Zur Illustration das Beispiel $N = 5$ (eine Primzahl)

X	1	2	3	4
X^4	1	16	81	256
$X^4 \bmod 5$	1	1	1	1

und zum Gegensatz $N = 8$ (eine zusammengesetzte Zahl)

X	1	2	3	4	5	6	7
X^7	1	128	2187	16384	78125	279936	823543
$X^7 \bmod 8$	1	0	3	0	5	0	7

Hat man also ein X mit $X^{N-1} \bmod N \neq 1$, dann kann N keine Primzahl sein. Nennen wir so ein X einen *Zeugen* für die Zusammengesetztheit von N .

Nehmen wir an, daß für jede zusammengesetzte Zahl N mindestens die Hälfte der Kandidatenzahlen in $Z_N = \{1, 2, 3, 4, \dots, N - 1\}$ Zeugen für die Zusammengesetztheit von N sind.

Betrachten wir nun folgende randomisierte Methode zum Testen, ob eine Zahl N zusammengesetzt oder prim ist:

Versuche 60 mal durch *zufällige* Wahl einer Zahl aus Z_N einen Zeugen für die Zusammengesetztheit von N zu finden.

Falls einer der Versuche erfolgreich ist, erkläre N für zusammengesetzt.

Wenn alle 60 Versuche fehlschlagen, erkläre N für prim.

Ist diese Methode korrekt? Wenn sie zum Schluss kommt, daß N zusammengesetzt ist, dann stimmt das auch, denn es wurde ja ein Zeuge dafür gefunden. Wenn sie aber zum Schluss kommt, daß N prim ist, dann könnte sich die Methode auch geirrt haben. Denn N könnte zusammengesetzt sein, aber bei jedem der 60 Versuche wurde aus Z_N zufällig eine Zahl gewählt, die kein Zeuge für die Zusammengesetztheit von N ist.

Die Wahrscheinlichkeit, daß es zu diesem Irrtum kommt, ist allerdings sehr klein. Nach unserer Annahme sind bei zusammengesetztem N mindestens die Hälfte der Zahlen in Z_N Zeugen. Die Wahrscheinlichkeit, bei zufälliger Auswahl aus Z_N einen Nichtzeugen zu ziehen, ist daher kleiner als $1/2$, und die Wahrscheinlichkeit, dies 60 mal hintereinander zu tun, ist kleiner als $1/2^{60}$ (das ist kleiner als 10^{-18}).

Stellt also unsere randomisierte Methode fest, eine Zahl ist zusammengesetzt, dann stimmt das. Stellt sie fest, eine Zahl ist prim, dann kann sie sich irren, und zwar mit einer Wahrscheinlichkeit kleiner als 10^{-18} .

Nun wird sicher für manchen eine Feststellung der Art "Diese Zahl ist *höchstwahrscheinlich* prim" nicht zufriedenstellend sein. Dazu sind drei Dinge zu sagen: Erstens kann man durch eine größere Anzahl von Wiederholungen die Fehlerwahrscheinlichkeit beliebig klein machen. Zweitens ist die Wahrscheinlichkeit dieses Fehlers vergleichbar mit der Wahrscheinlichkeit anderer Fehler, die beim Betreiben und Benutzen von Rechnern auftreten, oder sogar kleiner. Zum Beispiel braucht ein 10^9 -Operationen-pro-Sekunden-Rechner mehr als 31 Jahre für 10^{18} Operationen; die Wahrscheinlichkeit ist erheblich, daß es in diesem Zeitraum zu einem Hardwarefehler kommt. Drittens sind zur Zeit leider keine *effizienten* fehlerfreien Methoden zur Primzahlbestimmung bekannt, mit Ausnahme der Methode von Miller [Miller, 1976], deren Korrektheit allerdings wiederum auf einer *un-*

bewiesenen Verallgemeinerung der *unbewiesenen* Riemannschen Vermutung beruht.

Zwei Dinge bedürfen nun aber noch der Klärung:

1. Ist die Annahme berechtigt, bei zusammengesetztem N seien mindestens die Hälfte der Kandidaten in Z_N Zeugen?
2. Wie stellt man wirklich fest, ob eine Zahl $X \in Z_N$ ein Zeuge für die Zusammengesetztheit von N ist?

Die Antwort zu Frage 2) erscheint zuerst klar: nach Definition von Zeugen berechnet man einfach $X^{N-1} \bmod N$ und testet, ob das Ergebnis 1 ist. Wie macht man aber dies auf effiziente Art und Weise? Die Zahl X einfach $N-1$ mal mit sich selbst zu multiplizieren braucht sicher zu lange, wenn N groß ist (z.B. eine 30-stellige Zahl). Durch den Trick des sogenannten "wiederholten Quadrierens" kommt man aber um dieses Problem herum: zum Beispiel lässt sich X^{16} durch nur 4 Multiplikationen berechnen, nämlich

$$\left(\left((X^2)^2 \right)^2 \right)^2,$$

oder, als etwas komplizierteres Beispiel, lässt sich X^{26} durch

$$\left(\left(\left((X^2) \cdot X \right)^2 \right) \cdot X \right)^2$$

mit nur 6 Multiplikationen berechnen. Im Allgemeinen lässt sich mit Hilfe von wiederholtem Quadrieren ein Ausdruck der Form X^M bei einer n -stelligen Zahl M durch höchstens ungefähr $7n$ Multiplikationen berechnen. Bei unserer 30-stelligen Beispielzahl $10^{30} - 11$ bräuchte man also weniger als 210 Multiplikationen für einen Zeugentest und nicht 10^{30} .

Selbst mit wiederholtem Quadrieren werden allerdings die Zwischenergebnisse so groß, daß die Annahme, man könne eine Multiplikation in einer konstanten Zeiteinheit (in unserem Beispiel 10^{-6} Sekunden bei 100-stelligen Operanden) durchführen, nicht mehr gerechtfertigt ist². In unserem Fall, wo $X^{N-1} \bmod N$ gefragt ist, kann man aber die Zwischenergebnisse klein halten, indem jedes Zwischenergebnis Y durch den Rest $Y \bmod N$ ersetzt wird und damit weitergerechnet wird. Dies ist wegen der Rechenregel

$$(A \cdot B \cdot C) \bmod N = (((A \cdot B) \bmod N) \cdot C) \bmod N$$

zulässig.

Bei unserer Beispielzahl $10^{30} - 11$ braucht dann die randomisierte Methode nur höchstens $60 \cdot 2 \cdot 210 = 25200$ primitive Operationen (die Berechnung des "modularen Produkts" $A \cdot B \bmod N$ braucht 2 Operationen). Auf unserem Beispielcomputer bräuchte dies etwas mehr als eine vierzigstel Sekunde, und nicht 15 Millionen Jahre.

Schließlich müssen wir noch die Frage 1) klären. Die Antwort dazu ist zwar für fast alle zusammengesetzten Zahlen N positiv, im Allgemeinen aber negativ. Die Antwort wird aber für *alle* N positiv, wenn man die Definition von " X ist Zeuge für die Zusammengesetztheit von N " etwas erweitert, und zwar gilt X auch als Zeuge, wenn während der Berechnung von $X^{N-1} \bmod N$ durch wiederholtes Quadrieren eine nichttriviale Wurzel von 1 entdeckt wird, d.h. eine Berechnung $Y^2 \bmod N$ ergibt 1, wobei $Y \neq 1$ und $Y \neq N - 1$. Für eine Primzahl N kann so etwas nämlich nicht auftreten. Einen Beweis, daß bei zusammengesetztem N immer mindestens die Hälfte der Kandidaten Zeugen in diesem erweiterten Sinne sind, findet man in [Rabin, 1980].

Man beachte die unterschiedlichen Nutzungen von Zufall in diesem und im vorherigen Kapitel. Die randomisierte Optimierungsmethode in Kapitel 3 liefert immer eine korrekte Antwort, nur die Zeit, die gebraucht wird, um zu dieser Antwort zu kommen, hängt vom Zufall ab und ist im Erwartungswert klein. Diese Art von randomisierter Methode wird *Las Vegas Methode* genannt. Dies steht im Gegensatz zur *Monte Carlo Methode* zum Primzahltesten in diesem Abschnitt: sie hat eine vom Zufall im Wesentlichen unabhängige Laufzeit, dafür liefert sie aber nicht immer die korrekte Antwort. Die Korrektheit der Antwort hängt also vom Zufall ab, aber auf eine so günstige Art und Weise, daß die Wahrscheinlichkeit eines Fehlers in der Antwort verschwindend klein ist.

Schnelle Monte Carlo Methoden zum Primzahltesten wurden von Solovay und Strassen [Solovay und Strassen, 1977] und von Rabin [Rabin, 1976] vorgestellt. Die hier umrissene Methode folgt Ansätzen von Rabin [Rabin, 1980] und Miller [Miller, 1976]. Es ist hochinteressant, daß alle diese Methoden eine Zahl N als zusammengesetzt erkennen, ohne einen Teiler von N zu liefern. Es wird allgemein angenommen (und es ist ein Grundpfeiler der modernen Kryptographie), daß das Faktorisieren zusammengesetzter Zahlen, also das Bestimmen der Teiler, *wesentlich* schwieriger und zeitaufwendiger ist als das reine Erkennen solcher Zahlen. Diese Annahme ist allerdings unbewiesen und interessanterweise für das (jetzt noch) futuristische Modell des Quantenrechners sogar falsch [Shor, 1997].

Daß beim ungefähren Zählen, also beim Schätzen, der Zufall von Nutzen sein kann, entspricht unserer Erfahrung. Wollten wir wissen, wieviele Worte in einem 500 Seiten Buch vorkommen, dann könnten wir ein paar, sagen wir fünf, zufällige Seiten aussuchen und die darauf vorkommenden Worte zählen. Das Ergebnis mal hundert wäre dann ein guter Schätzwert für die Zahl aller Worte in dem Buch. Wollten wir wissen, wie oft das Wort "Zufall" in diesem Buch vorkommt, könnten wir genau so vorgehen. Intuitiv wissen wir aber, daß im zweiten Fall der so erhaltene Schätzwert nicht so vertrauenswürdig ist wie im ersten, denn es wird eine größere Streuung geben, das heißt, das Wort "Zufall" wird kaum über alle Seiten des Buches annähernd gleich verteilt sein. So kommt es zum Beispiel auf den Seiten dieses Aufsatzes wohl um einiges öfter vor als auf den Seiten anderer Aufsätze dieses Bandes. Aber um unseren Schätzwert zu verbessern, könnten wir eine größere Anzahl von Seiten betrachten. Die Wissenschaft der Statistik gibt uns quantitativ an, wie viele Seiten wir bei welcher Streuung betrachten müssen, um eine bestimmte Güte des Schätzwerts zu erreichen.

Im Falle der Zahl der Worte in einem Buch wäre das Schätzen eigentlich nicht notwendig. Man könnte die Worte ja einfach abzählen. Für einen Menschen wäre das etwas mühsam, für einen Computer wäre es aber nur eine Sache weniger Sekunden, sollte das Buch in elektronischer Form vorhanden sein. In diesem Abschnitt wollen wir Zählprobleme behandeln, die, obwohl einfach formulierbar, so kompliziert sind, daß selbst mit den besten Computern, auch wenn man sie noch millionenfach beschleunigte, keine genauen Lösungen in vernünftiger Zeit errechnet werden können. Ja, selbst die Frage, ob man in vernünftiger Zeit halbwegs gute Schätzwerte errechnen kann, ist erst teilweise gelöst und ist gegenwärtig ein interessanter Forschungsgegenstand.

Wir wollen uns mit dem *Paarungsanzahl*-Problem beschäftigen, bei dem die Anzahl der *vollständigen Paarungen* in einem *bipartiten Graphen* gezählt werden soll. Was ist ein bipartiter Graph? Man malt sich n schwarze und n weiße Knoten auf und Kanten, die manche schwarze mit manchen weißen Knoten verbinden. Ein Beispiel mit $n = 4$ ist in Abbildung 2.4 zu sehen.

Was ist eine vollständige Paarung? Das sind n Kanten, die jeden Knoten mit genau einem Knoten der anderen Farbe verbinden, anders gesagt, n schwarz-weiße Knotenpaare, in denen alle Knoten vorkommen. Im Graphen von Abbildung 2.4 gibt es genau drei solche vollständige Paarungen.

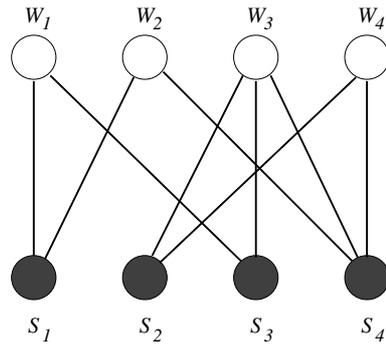


Abbildung 2.4.

Sie werden in Abbildung 2.5 gezeigt.

Der Leser möge sich davon überzeugen, daß es in dem Beispielgraphen keine weiteren vollständigen Paarungen gibt. Das wird wahrscheinlich nicht ganz leicht fallen und zeigt, daß selbst schon bei solch kleinen Beispielen das Paarungsanzahlproblem nicht einfach ist. Bei größeren Graphen wird das noch durch die riesige mögliche Anzahl der vollständigen Paarungen erschwert. In einem Graphen, in dem jeder der n schwarzen Knoten mit jedem der n weissen Knoten verbunden ist, gibt es n Partnerkandidaten für W_1 , für jede Wahl bleiben für W_2 dann $n - 1$ Partnerkandidaten übrig, $n - 2$ Kandidaten für W_3 , und so fort. Insgesamt gibt es dann $n \cdot (n - 1) \cdot (n - 1) \cdot \dots \cdot 2 \cdot 1 = n!$ viele vollständige Paarungen. Für $n = 10$ sind das etwas mehr 3,6 Millionen, für $n = 20$ schon über 2,4 Trillionen ($2,4 \cdot 10^{18}$).

Für das Paarungsanzahlproblem ist derzeit keine vernünftige, exakte und deterministische Lösungsmethode bekannt. Es ist nicht zu erwarten, daß je eine gefunden wird, da das Problem #P-vollständig ist. Das bedeutet, fände man eine vernünftig rasche Lösungsmethode für dieses Problem, würde man damit gleichzeitig Lösungsmethoden für eine ganze Horde anderer Zählprobleme finden, die allesamt als sehr schwierig gelten. Ein tatsächlicher Beweis, daß es so eine "vernünftige" Lösungsmethode nicht geben kann, steht allerdings noch aus.

Man versucht daher, zumindest approximative Lösungsmethoden für das Paarungsanzahlproblem zu finden. Es ist noch nicht klar, ob das immer mit vernünftigem Aufwand und auch vernünftiger Approximationsgüte möglich ist. Es sind aber schon gewisse Teilerfolge erzielt worden. Zwei

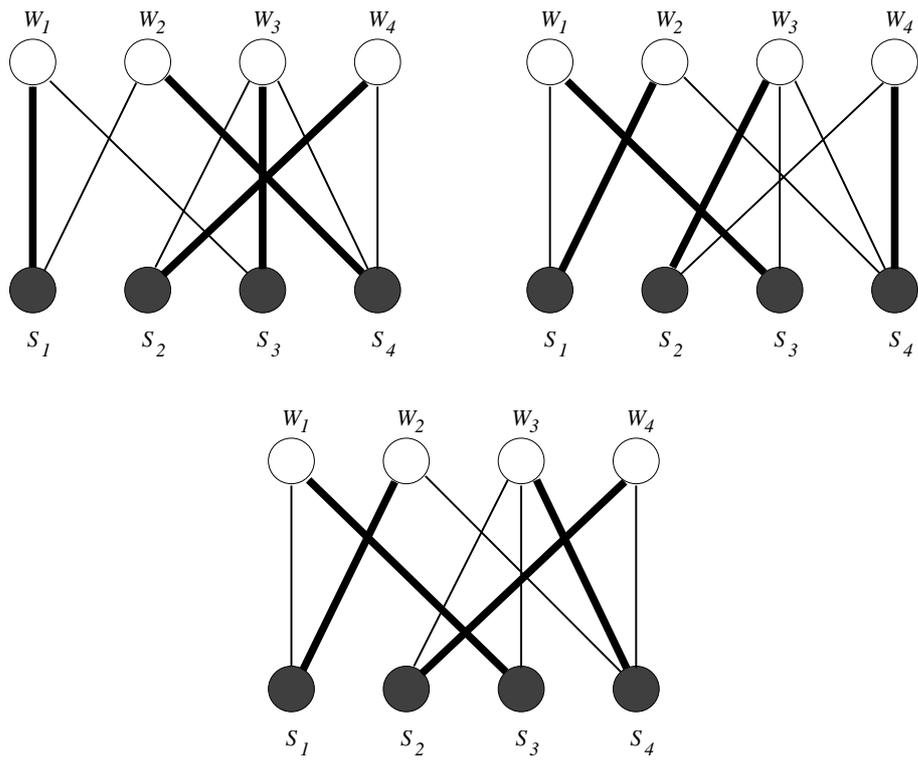


Abbildung 2.5.

Ansätze sollen hier kurz vorgestellt werden. Beide verwenden das Randomisieren, also den kontrollierten Einsatz von Zufall, auf sehr raffinierte Art und Weise. Die erste Methode beruht auf der Algebra, die zweite auf sogenannten zufälligen Wanderungen oder Irrfahrten.

Zuvor aber noch zur Beruhigung des Lesers: dies ist nicht alles *l'art pour l'art*. In der statistischen Physik möchte man genau solche Paarungsanzahlprobleme lösen.

Eine algebraische Methode

Man kann einen bipartiten Graphen mit n weissen Knoten W_1, \dots, W_n und n schwarzen Knoten S_1, \dots, S_n auf natürliche Weise durch seine *Adjazenzmatrix* darstellen. Das ist eine $n \times n$ Matrix A , in der $A_{i,j}$, der Eintrag in der i -ten Zeile und j -ten Spalte, zu 1 gemacht wird, wenn die Knoten W_i und S_j

durch eine Kante verbunden sind, und zu 0 sonst. Die Adjazenzmatrix des Graphen von Abbildung 2.4 sieht dann folgendermaßen aus:

$$\begin{array}{cccc}
 & S_1 & S_2 & S_3 & S_4 \\
 W_1 & \left(\begin{array}{cccc} 1 & 0 & 1 & 0 \end{array} \right) \\
 W_2 & \left(\begin{array}{cccc} 1 & 0 & 0 & 1 \end{array} \right) \\
 W_3 & \left(\begin{array}{cccc} 0 & 1 & 1 & 1 \end{array} \right) \\
 W_4 & \left(\begin{array}{cccc} 0 & 1 & 0 & 1 \end{array} \right)
 \end{array}$$

Eine vollständige Paarung entspricht dann einer Auswahl von n Einsen in dieser Matrix, sodaß in jeder Zeile und in jeder Spalte genau eine Eins gewählt wurde. Die drei möglichen vollständigen Paarungen unseres Beispiels sehen dann so aus:

$$\left(\begin{array}{cccc} \boxed{1} & 0 & 1 & 0 \\ 1 & 0 & 0 & \boxed{1} \\ 0 & 1 & \boxed{1} & 1 \\ 0 & \boxed{1} & 0 & 1 \end{array} \right) \quad \left(\begin{array}{cccc} 1 & 0 & \boxed{1} & 0 \\ \boxed{1} & 0 & 0 & 1 \\ 0 & \boxed{1} & 1 & 1 \\ 0 & 1 & 0 & \boxed{1} \end{array} \right) \quad \left(\begin{array}{cccc} 1 & 0 & \boxed{1} & 0 \\ \boxed{1} & 0 & 0 & 1 \\ 0 & 1 & 1 & \boxed{1} \\ 0 & \boxed{1} & 0 & 1 \end{array} \right)$$

Um dies nun algebraisch zu fassen, brauchen wir den Begriff der Permutation. Eine Permutation der Größe n ist eine Auflistung der Zahlen von 1 bis n , sodaß jede Zahl in dieser Auflistung genau einmal vorkommt. Zum Beispiel sind $\rho = (1, 4, 3, 2)$ und $\sigma = (3, 1, 2, 4)$ Permutationen der Größe 4, aber $\tau = (2, 4, 1, 2)$ ist keine. Wenn π eine Permutation ist, dann bezeichnet $\pi(i)$ einfach die Zahl, die an i -ter Stelle der Auflistung π steht; z.B. $\sigma(3)$ ist 2. Wenn π eine Permutation der Größe n ist und A eine $n \times n$ Matrix, dann ist

$$A_{1,\pi(1)}, A_{2,\pi(2)}, A_{3,\pi(3)}, \dots, A_{n,\pi(n)}$$

eine Auswahl von n Elementen der Matrix, sodaß aus jeder Zeile und aus jeder Spalte genau ein Element gewählt wurde. Jede Auswahl mit dieser Eigenschaft lässt sich durch genau eine Permutation identifizieren. So entsprechen die oben angegebenen Permutationen ρ und σ genau zwei der in unserem Beispiel getroffenen Auswahlen (welchen, das möge der Leser sich überlegen). Die Permutation $\pi = (1, 2, 3, 4)$ entspricht der Auswahl $A_{1,1}, A_{2,2}, A_{3,3}, A_{4,4}$, die aber in unserem Beispiel wiederum keiner vollständigen Paarung entspricht, da $A_{2,2}$ gleich 0 ist.

Wir sehen nun, daß vollständige Paarungen genau solchen Permutationen π entsprechen, für die gilt $A_{i,\pi(i)} = 1$ für alle i , oder anders ausgedrückt,

für das Produkt gilt

$$A_{1,\pi(1)} \cdot A_{2,\pi(2)} \cdot A_{3,\pi(3)} \cdots A_{n,\pi(n)} = 1.$$

Für Permutationen, die keinen vollständigen Paarungen entsprechen, ist dieses Produkt null. Damit kann man jetzt die Anzahl der vollständigen Paarungen in einem bipartiten Graphen schön algebraisch ausdrücken als Summe über alle möglichen permutationsentsprechenden Produkten:

$$\sum_{\substack{\pi \text{ ist Permutation} \\ \text{der Größe } n}} A_{1,\pi(1)} \cdot A_{2,\pi(2)} \cdot A_{3,\pi(3)} \cdots A_{n,\pi(n)}.$$

Dies mag zwar hübsch aussehen, ist aber vorerst nicht besonders hilfreich, denn es gibt $n!$ viele Permutationen der Größe n . Es gibt also viel zu viele Summanden, als daß diese Summe in vernünftiger Zeit ausgewertet werden könnte.

Aber diese Summe ähnelt stark der Definition der *Determinante* $\det(A)$ einer Matrix A :

$$\det(A) = \sum_{\substack{\pi \text{ ist Permutation} \\ \text{der Größe } n}} s(\pi) \cdot A_{1,\pi(1)} \cdot A_{2,\pi(2)} \cdot A_{3,\pi(3)} \cdots A_{n,\pi(n)},$$

wobei $s(\pi)$ ein von der Permutation abhängiges Vorzeichen ist, also $+1$ oder -1 . Obwohl die Determinantenfunktion durch diese Vorzeichen in ihrer Summendefinition komplizierter aussieht, lässt sie sich mit Hilfe alternativer Definitionen relativ leicht berechnen. Diese Tatsache wird nun unter Zuhilfenahme von Zufall ausgenutzt.

Godsil und Gutman [Godsil und Gutman, 1981] schlugen 1981 folgende Methode vor: Es sei A die Adjazenzmatrix eines bipartiten Graphens G . Ändere jede 1 in A zufällig, mit Wahrscheinlichkeit $1/2$ (und unabhängig von den anderen Einsen) zu -1 . Berechne die Determinante der so erhaltenen zufälligen Matrix und nimm ihr Quadrat als Schätzwert für die Anzahl der vollständigen Paarungen im Graphen G .

Warum soll das ein guter Schätzwert sein? Es stellt sich heraus, daß, wenn man alle möglichen Muster von $+1/-1$ der Matrix A betrachtete und man für jedes Muster den Schätzwert wie angegeben berechnete, dann wäre der Durchschnitt der so erhaltenen Schätzwerte genau die Anzahl der vollständigen Paarungen. Mathematisch ausgedrückt: der Erwartungswert des Schätzwertes ist die Paarungsanzahl.

Das ist nicht sehr schwer zu beweisen. Das Quadrat der Determinante von A lässt sich schreiben als

$$\left(\sum_{\substack{\pi \text{ ist Permutation} \\ \text{der Größe } n}} s(\pi) \cdot A_{1,\pi(1)} \cdots A_{n,\pi(n)} \right)^2,$$

was mit Hilfe des Distributivgesetzes zu

$$\sum_{\substack{\pi, \rho \text{ Permutationen} \\ \text{der Größe } n}} (s(\pi) \cdot A_{1,\pi(1)} \cdots A_{n,\pi(n)}) (s(\rho) \cdot A_{1,\rho(1)} \cdots A_{n,\rho(n)})$$

wird. Wir wollen den Erwartungswert dieser Summe wissen. Der ist die Summe der Erwartungswerte der einzelnen Summanden. Betrachten wir also einen Summanden

$$(s(\pi) \cdot A_{1,\pi(1)} \cdots A_{n,\pi(n)}) (s(\rho) \cdot A_{1,\rho(1)} \cdots A_{n,\rho(n)}).$$

Sollte π oder ρ keiner vollständigen Paarung entsprechen, d.h. $A_{i,\pi(i)} = 0$ oder $A_{i,\rho(i)} = 0$ für irgendein i , dann ist dieser Ausdruck (und damit auch sein Erwartungswert) 0.

Im Fall $\rho \neq \pi$ muß es ein i geben mit $\pi(i) \neq \rho(i)$. Der Term $A_{i,\pi(i)}$ kommt dann in dem Produkt

$$(s(\pi) \cdot A_{1,\pi(1)} \cdots A_{n,\pi(n)}) (s(\rho) \cdot A_{1,\rho(1)} \cdots A_{n,\rho(n)})$$

genau einmal vor. Aufgrund seiner gleichwahrscheinlichen $-1/ +1$ Belegung ist der Erwartungswert von $A_{i,\pi(i)}$ gleich 0, und damit (und aufgrund der Unabhängigkeit der zufälligen $-1/+1$ -Zuweisungen) ist auch der Erwartungswert des gesamten Produkts gleich 0.

Schließlich im Fall, daß $\rho = \pi$, ergibt dieser Ausdruck

$$(s(\pi) \cdot A_{1,\pi(1)} \cdots A_{n,\pi(n)})^2,$$

und das ist immer 1, ganz gleich wie das Vorzeichen $s(\pi)$ aussieht und welche $+1/-1$ Werte den Matrixelementen zugewiesen wurden. Damit bekommen wir genau für jede Permutation, die einer Paarung entspricht, einen Beitrag von 1 zum Erwartungswert der gesamten Summe, und damit ist der Erwartungswert dieser Summe gleich der gesuchten Paarungsanzahl.

Dieser Schätzwert von Godsil und Gutman hat nun zwar den richtigen Erwartungswert, nur ist leider seine Streuung so groß, daß man etwas mehr

als $3^{n/2}$ viele unabhängige Schätzwerte ermitteln muß, damit ihr Durchschnitt mit vernünftiger Wahrscheinlichkeit eine vernünftige Approximation der wahren Paarungszahl ist.

Dies wurde inzwischen etwas verbessert. 1993 zeigten Karmakar et al. [Karmakar et al., 1993], daß folgende leichte Veränderung des Verfahrens von Godsil und Gutman zu einem guten Schätzverfahren mit geringerer Streuung führt und somit nur etwas mehr als $2^{n/2}$ Schätzwerte ermittelt werden müssen.

Ändere jede 1 in A zufällig, mit $1/4$ Wahrscheinlichkeit (und unabhängig von den anderen Einsen) zu $-1, +1, I, -I$, wobei $I = \sqrt{-1}$. Berechne die Determinante der so erhaltenen zufälligen Matrix und nimm das Quadrat ihres Absolutbetrags als Schätzwert für die Anzahl der vollständigen Paarungen im Graphen G .

Schließlich zeigte Rasmussen [Rasmussen, 1998] in 1998, aufbauend auf Arbeiten von Barvinok [Barvinok, 1999], daß man über die komplexen Zahlen hinweg noch einen Schritt weiter machen kann, um die Streuung und damit die Anzahl der nötigen Schätzwerte zu verringern. Er verwendet zufällige Ersetzungen der Einsen durch die acht Quaternionen $\pm 1, \pm I, \pm J, \pm K$ und reduziert damit die Anzahl der zu ermittelnden Schätzwerte auf etwas über $(3/2)^{n/2}$. Dies ist zwar noch immer exponentiell, also viel zu groß und führt nicht zu einer Schätzmethode mit vernünftiger Laufzeit, aber es legt nahe, daß durch die Verwendung von höher-dimensionalen Algebren, sogenannten Clifford Algebren, die Streuung auf ein vernünftiges Maß reduziert werden könnte. Ob dies wirklich möglich ist, ist zur Zeit nicht bekannt und Gegenstand der Forschung.

Eine Irrfahrtmethode

Hier ist eine vollkommen andere randomisierte Methode, um die Paarungszahl in einem bipartiten Graphen zu schätzen. Diese Methode ist viel weniger direkt als die gerade gezeigte.

Definieren wir eine k -Paarung in einem bipartiten Graphen als eine Menge P von k Kanten, sodaß jeder Knoten im Graphen an höchstens einer der Kanten in P anliegt. Im folgenden sei nun G ein bipartiter Graph mit $2n$ Knoten. Es seien M_k die Menge aller k -Paarungen in G und m_k die Anzahl der verschiedenen k -Paarungen in G . Eine vollständige Paarung in G ist jetzt nichts anderes als eine n -Paarung, also ein Element von M_n , und die Paarungszahl in G zu bestimmen oder zu schätzen ist nichts anderes als das Bestimmen oder Schätzen von m_n .

Für $i = 2, \dots, n$ sei $r_i = m_i/m_{i-1}$. Dann gilt

$$m_1 \cdot r_2 \cdot r_3 \cdots r_{n-1} \cdot r_n = m_1 \cdot \frac{m_2}{m_1} \cdot \frac{m_3}{m_2} \cdots \frac{m_{n-1}}{m_{n-2}} \cdot \frac{m_n}{m_{n-1}} = m_n.$$

Die Idee ist nun diese: m_1 ist bekannt, es ist die Anzahl der Kanten von G ; wenn man jedes der r_i 's schätzt, dann ergibt das Produkt von m_i und von allen diesen Schätzungen nach der angegebenen Gleichung eine Schätzung der gesuchten Zahl m_n .

Wie kann man aber ein r_i schätzen, und zwar mit hinreichender Genauigkeit? Die Grundidee hier ist wieder einfach: Angenommen, man hätte einen "Zufallspaarungsgenerator", der auf Anfrage eine zufällige Paarung in $M_{i-1} \cup M_i$ erzeugt, jede mit gleicher, oder zumindest fast gleicher Wahrscheinlichkeit. Dann könnte man eine ganze Reihe von Anfragen an diesen Generator richten und zählen, wie oft er eine $(i-1)$ -Paarung und wie oft er eine i Paarung liefert, sagen wir jeweils c_{i-1} und c_i mal. Wenn r_i nicht zu groß oder zu klein ist (was derzeit nur garantiert werden kann, wenn jeder Knoten in G an hinreichend viele, sagen wir mindestens $n/2$, Kanten anliegt) und wenn hinreichend viele Anfragen gestellt wurden, dann liefert das Verhältnis c_{i-1}/c_i mit hoher Wahrscheinlichkeit eine gute Schätzung für das gesuchte r_i .

Jetzt brauchen wir also nur mehr so einen "Zufallspaarungsgenerator". Dieser funktioniert folgendermaßen: Man beginnt mit irgendeiner Paarung P_0 in $M_{i-1} \cup M_i$. Diese verändert man auf zufällige Art und Weise durch eine kleine "Mutation" zu einer Paarung P_1 in $M_{i-1} \cup M_i$. Durch eine weitere kleine zufällige Änderung bekommt man davon Paarung P_2 , und so fort. Nach t solchen "zufälligen, kleinen" Veränderungen erreicht man eine Paarung P_t in $M_{i-1} \cup M_i$, und es ist plausibel, daß bei hinreichend großem t die so in vielen zufälligen kleinen Schritten erreichte Paarung P_t mit fast gleicher Wahrscheinlichkeit jede der Paarungen in $M_{i-1} \cup M_i$ sein kann.

Dies ist bei geeigneter Spezifikation von "kleiner zufälligen Veränderung" (im Prinzip ändert sich die Paarung in höchstens einer zufälligen Kante) tatsächlich der Fall. Der Beweis dafür ist aber äußerst kompliziert und weit außerhalb des Rahmens dieses Aufsatzes.

Der gerade skizzierte Zufallspaarungsgenerator basiert auf der Methode der "Irrfahrt" oder "zufälligen Wanderung". Diese Namen erklären sich dadurch, daß man sich vorstellen kann, daß man auf zufällige Art und Weise entlang der Kanten eines Graphens U wandert, also eine Irrfahrt macht, wobei in unserem Fall die Paarungen in $M_{i-1} \cup M_i$ die Knoten von U sind, und

zwei Knoten (also Paarungen) durch eine Kante in U verbunden sind, wenn sie durch "kleine Änderungen" auseinander hervorgehen. Diese Sichtweise und die damit verbundene Theorie der Markoff-Ketten hat sich bei einigen Zähl- und Schätzproblemen als sehr fruchtbar erwiesen. Wir verweisen den Leser auf das Buch [Motwani und Raghavan, 1995] von Motwani und Raghavan und auf die Monographie [Sinclair, 1992] von Sinclair.

Zusammenfassend können wir sagen, daß die hier gerade skizzierte randomisierte Methode für das Schätzen der Paarungsanzahl eines bipartiten Graphen G im Fall, daß jeder Knoten von G an hinreichend vielen Kanten anliegt, nachweislich mit "vernünftig" viel Rechenaufwand mit hoher Wahrscheinlichkeit zu einer "vernünftig" genauen Lösung führt. Was "vernünftig" hier genau bedeuten soll, lassen wir offen. Der technisch versierte Leser möge das Wort "polynomiell" dafür einsetzen. Man soll aber keinesfalls glauben, daß "vernünftig viel Rechenaufwand" hier mit praktischer Effizienz gleichzusetzen sei.

Schlußbemerkungen

Rechnen, Berechnen und Problemlösen tragen üblicherweise die Attribute "exakt", "vorhersagbar" und "deterministisch", also eigentlich gerade das Gegenteil von "zufällig". Die Beispiele in diesem Aufsatz haben den Leser hoffentlich davon überzeugt, daß der Zufall sehr wohl fürs Rechnen und Problemlösen nützlich sein kann, wenn er nur richtig eingesetzt wird. Die dargestellten Beispiele reichten vom Koordinieren von unabhängigen Agenten über rasches Lösen von Optimierungsproblemen bis hin zum Testen von speziellen Eigenschaften von Zahlen und zum algorithmischen Schätzen komplizierter Größen. Dies ist aber wahrlich nur ein sehr kleiner Ausschnitt aus dem Gebiet der randomisierten Algorithmen. Beim Verwalten von Information findet Randomisierung ebenso Anwendung wie beim Lösen geometrischer oder zahlentheoretischer Probleme, oder bei Algorithmen der Textverarbeitung. Der interessierte Leser sei an das Buch von Motwani und Raghavan [Motwani und Raghavan, 1995] verwiesen.

Randomisierte Algorithmen zeichnen sich oft durch ihre Einfachheit aus. Wenn überhaupt etwas dabei kompliziert wird, dann sind es typischerweise die Beweise ihrer Korrektheit- oder Laufzeiteigenschaften. Dies macht sie für den praktischen Einsatz oft besonders attraktiv.

In randomisierten Methoden muß oft "eine Münze geworfen" werden, oder es soll eine zufällige ganze Zahl zwischen 0 und n generiert werden.

Mancher Leser wird fragen, durch welchen Mechanismus dies im Ablauf von Computerprogrammen tatsächlich realisiert wird. Dieses Problem ist schon in sich selbst einen Aufsatz wert. Wir verweisen dafür auf das Buch von Luby [Luby, 1996] für einige theoretische Grundlagen der sogenannten Pseudozufallszahlenerzeugung und auf den zweiten Band des großen Werkes von Knuth [Knuth, 1997].

Danksagung

Die Abschnitte *Den Zufall kontrollieren* und *Zufällige Zeugen und Primzahlbestimmung* dieser Arbeit erschienen schon in einem früheren Aufsatz des Autors mit dem Titel "Der Zufall in der Informatik". Der Abdruck erfolgt mit freundlicher Genehmigung der Deutschen Akademie der Naturforscher Leopoldina aus: Nova Acta Leopoldina NF Bd. 79, Nr. 308 "Der Zufall", S. 127–139 (1999), Herausgeber: K. Krickeberg, H. Bauer, H. Föllmer, J. Moser und V. Strassen.

Anmerkungen

- ¹ Wir nehmen hier an, daß die Punktmenge S in "allgemeiner Lage" ist, in dem Sinne, daß keine vier Punkte in S auf einem gemeinsamen Kreis liegen. Aber selbst wenn dies nicht der Fall ist, stimmt die Wahrscheinlichkeitsaussage noch immer, ja, die Wahrscheinlichkeit für den schlechten Fall wird sogar kleiner.
- ² Mathematische Leser mögen bitte verzeihen, daß darauf verzichtet wurde, hier Restklassenringe einzuführen.

Literatur

- [Barvinok, 1999] Barvinok, A. I. (1999). Polynomial time algorithms to approximate permanents and mixed discriminants within a simply exponential factor. *Random Structures & Algorithms* 14, 29–61.
- [Birkhoff, 1946] Birkhoff, G., (1946). Tres observaciones sobre el algebra lineal. *Rev. univ. nac. Tucum n (A)* 5, 147–151.
- [Brogan, 1989] Brogan, W. L., (1989). Algorithm for ranked assignments with application to multiobject tracking. *Journal of Guidance* 12, 357–364.
- [Gärtner, 1995] Gärtner, B. (1995). A subexponential algorithm for abstract optimization problems. *SIAM Journal on Computing* 24, 1018–1035.
- [Gärtner, 1999] Gärtner, B. (1999). Smallest enclosing balls – fast and robust in C++.
<http://www.inf.ethz.ch/personal/gaertner/miniball.html>

- [Godsil und Gutman, 1981] Godsil, C. und Gutman, I. (1981). On the matching polynomial of a graph. *Algebraic Methods in Graph Theory I*. Lovász, L. und Sós, V., Hrsg., Colloq. Math. Soc. János Bolyai, 25, North-Holland, Amsterdam, 241–249.
- [Karp, 1991] Karp, R. (1991). An introduction to randomized algorithms. *Discrete Applied Mathematics*, 34, 165–201.
- [Knuth, 1997] Knuth, D. (1997). *Seminumerical Algorithms*, Band 2 von *The Art of Computer Programming*. Addison-Wesley, Dritte Ausgabe.
- [Luby, 1996] Luby, M. (1996). *Pseudorandomness and Cryptographic Applications*. Princeton University Press.
- [Matoušek et al., 1996] Matoušek, J., Sharir, M. und Welzl, E. (1996). A sub-exponential bound for linear programming. *Algorithmica*, 16, 498–516.
- [Miller, 1976] Miller, G. (1976). Riemann’s Hypothesis and Tests for Primality. *Journal of Computer and System Sciences*, 13, 300–317.
- [Motwani und Raghavan, 1995] Motwani, R. und Raghavan, P. (1995). *Randomized Algorithms*. Cambridge University Press.
- [Rabin, 1976] Rabin, R. (1976). Probabilistic algorithms. In *Algorithms and Complexity, Recent Results and New Directions*, Traub, J., Hrsg., Academic Press, 21–39.
- [Rabin, 1980] Rabin, M. (1980). Probabilistic algorithm for testing primality. *Journal of Number Theory*, 12, 128–138.
- [Rasmussen, 1998] Rasmussen, L. E. (1998) *On Approximating the Permanent and other #P-complete Problems*. Ph.D. Thesis, Computer Science Division, Univ. of California, Berkeley.
- [Shor, 1997] Shor, P. (1997). Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26, 1484–1509.
- [Sinclair, 1992] Sinclair, A. (1992). *Algorithms for Random Generation and Counting: A Markov Chain Approach*. Progress in Theoretical Computer Science. Birkhäuser, Boston.

[Solovay und Strassen, 1977] Solovay, R. und Strassen, V. (1977, 1978). A fast Monte-Carlo test for primality. *SIAM Journal on Computing*, 6(1), 84–85, 1977; und auch *SIAM Journal on Computing*, 7(1), 118, 1978.

[Welzl, 1991] Welzl, E. (1991). Smallest enclosing disks (balls and ellipsoids). In *New Results and New Trends in Computer Science*, Maurer, H., Hrsg., Springer Lecture Notes in Computer Science, 555, 359–370.

Die Kunst im formalen Denken anhand zweier Beispiele aus der Zahlentheorie¹ Wolfgang M. Schmidt

Ganz im Gegensatz zur weithin verbreiteten Ansicht, formales Denken habe nichts mit Kunst zu tun, steht der Glaube an tiefliegende Parallelen zwischen mannigfachen Bereichen der Kunst und derjenigen Wissenschaft, deren wesentlichstes Hilfsmittel das formale Denken ist, der Mathematik. Man kann sogar noch einen Schritt weitergehen und mit dem Schlagwort "Mathematik als Kunst" eine neue, zur gemeinhin üblichen ganz konträre Sichtweise dieser Wissenschaft annehmen — ein Thema, das ich bereits in mehreren Vorträgen angesprochen habe.

Doch stellen wir einmal die Frage: was unterscheidet die Wissenschaft denn so grundlegend von der Kunst? Eine der augenscheinlichsten Antworten darauf lautet wahrscheinlich: die Anwendung im täglichen Leben. Während der Nutzen eines künstlerischen Werkes hauptsächlich in der Beglückung seiner Bewunderer zu suchen ist, sind es im Bereich der Wissenschaft hauptsächlich deren Anwendungen im Alltag, die uns wesentlich sind. Dies trifft ganz besonders auf alle Naturwissenschaften zu und somit auch indirekt auf die Mathematik, da deren Ergebnisse wiederum in all diesen anderen Bereichen eingesetzt werden.

Die direkten Anwendungen des formalen Denkens, denen wir begegnen, sind aber relativ selten, eine Tatsache, die Kunst und Mathematik bereits etwas näher rückt. Insbesondere mein Arbeitsgebiet, die Zahlentheorie, galt lange Zeit als besonders anwendungsarm, was einen ihrer berühmtesten Vertreter Carl Friedrich Gauß nicht hinderte, sie mit dem Worten: "Die Mathematik ist die Königin der Wissenschaften und die Zahlentheorie ist die Königin der Mathematik" zu preisen. Noch im 19. Jahrhundert war alles, was auf diesem Gebiet der Mathematik ersonnen wurde, lediglich als wissenschaftliche Erkenntnis von Bedeutung, was viele berühmte Mathematiker in den Bann gerade dieser Disziplin zog. Im 20. Jahrhundert allerdings scheint dieser Standpunkt nicht mehr ganz vertretbar zu sein, da sich auch für die Ergebnisse der Zahlentheorie interessierte Abnehmer gefunden haben. Wer hätte es für möglich gehalten, daß etwa das amerikanische Vertei-

digungsministerium die mathematische Forschung fördert, da viele Ergebnisse der Zahlentheorie für die Erstellung von Geheimcodes beziehungsweise deren Entschlüsselung von großer Bedeutung sind? Auf dieses, Kryptographie genannte Gebiet, möchte ich später näher eingehen, um weitere Parallelen zwischen Kunst und Mathematik aufzuzeigen.

Vorerst aber ein Beispiel dazu, dessen Wurzeln auf mathematischer Seite tief in Österreich verankert sind durch bekannte Vertreter wie Edmund Hlawka, Harald Niederreiter oder Robert Tichy. Es handelt sich dabei um die Theorie der Gleichverteilung; etwas banaler ausgedrückt um die Untersuchung von sogenannten Zufallsfolgen: das sind Folgen von Zahlen, die scheinbar willkürlich zustande kommen, hinter denen sich aber eine versteckte Gesetzmäßigkeit verbirgt. Ein quasi "geordnetes Chaos" also — und damit tut sich bereits eine Parallele zwischen Mathematik und Kunst auf, nämlich die simultane Präsenz von Ordnung und Chaos, sowie deren wechselseitige Beziehungen und Einflüsse. Besonders deutlich wird dies in der modernen Kunst des 20. Jahrhunderts, wo mit klassischen Methoden der Malerei Motive geschaffen werden, die scheinbar jeder Gesetzmäßigkeit entbehren.

Um diesen Gedankengang weiter zu vertiefen, möchte ich nun den Begriff der Zufallsfolgen näher erklären, was allerdings einiger Vorbemerkungen in elementarer Zahlentheorie bedarf.

Bereits in der Volksschule wird uns das Dividieren mit Rest beigebracht: nehmen wir als Beispiel die Zahl 7 her und dividieren eine beliebige Zahl durch sie, so bleibt ein ganzzahliger Rest, der zwischen 0 und 6 liegt. Wir schreiben

$$x \equiv y \pmod{7} \quad (\text{sprich } x \text{ kongruent } y \pmod{7})$$

wenn x und y denselben Rest bei Division durch 7 haben, wie etwa

$$44 \equiv 16 \pmod{7},$$

denn bei beiden bleibt 2 als Rest übrig.

Die ganzen Zahlen zerfallen so, entsprechend ihrem Rest nach Division durch 7 in 7 Restklassen, die man mit

$$\bar{0}, \bar{1}, \dots, \bar{6}$$

bezeichnen kann. 44 liegt somit in der Klasse $\bar{2}$ ebenso wie 16. Anstelle von 7 kann dieselbe Überlegung für jede Zahl m als Modul durchgeführt werden

und wir schreiben

$$x \equiv y \pmod{m},$$

falls x und y denselben Rest bei Division durch m ergeben und die ganzen Zahlen werden nun in m Restklassen bezüglich des Moduls m eingeteilt, die wir mit

$$\overline{0}, \overline{1}, \overline{2}, \dots, \overline{m-2}, \overline{m-1}$$

bezeichnen. Diese Reihenfolge der Restklassen scheint hier ganz natürlich, wir wollen aber noch eine weitere kennenlernen, hinter deren scheinbarer Zufälligkeit sich eine tiefere Gesetzmäßigkeit verbirgt.

Dazu betrachten wir die Multiplikation von Restklassen, wiederum am Beispiel des Moduls $m = 7$. Um etwa die Klassen $\overline{3}$ und $\overline{4}$ zu multiplizieren, werden einfach die Zahlen 3 und 4 multipliziert, dem Ergebnis dieser Operation wird dann die entsprechende Restklasse zugeordnet:

$$\overline{3} \cdot \overline{4} = \overline{3 \cdot 4} = \overline{12} = \overline{5}.$$

Mit Hilfe dieser Definition kann man nun untersuchen, wie sich Potenzen einer beliebig gewählten Zahl auf die einzelnen Restklassen verteilen. Für $m = 7$ und die Zahl 3 liefert dies

$$\begin{aligned} 3 &\in \overline{3} \\ 3^2 &= 3 \times 3 \in \overline{9} = \overline{2} \\ 3^3 &= 3 \times 3^2 \in \overline{3} \cdot \overline{2} = \overline{6} \\ 3^4 &= 3^2 \times 3^2 \dots \in \overline{4} \\ 3^5 &\in \overline{5} \\ 3^6 &\in \overline{1} \end{aligned}$$

und man erhält so die Restklassen $\overline{3}, \overline{2}, \overline{6}, \overline{4}, \overline{5}, \overline{1}$. Dem aufmerksamen Betrachter wird dabei auffallen, daß dabei alle Restklassen ungleich $\overline{0}$ genau einmal vorkommen. Das hängt natürlich mit der speziellen Wahl dieses Beispiels zusammen, jedoch hat Gauß gezeigt, daß für eine gewisse Klasse von Moduln m , nämlich alle Primzahlen (das sind alle ganzen Zahlen, die nur durch 1 und sich selbst teilbar sind, also 2, 3, 5, 7, 11, ...) immer eine Zahl g existiert, deren erste $m - 1$ Potenzen $\{\overline{g}, \overline{g^2}, \dots, \overline{g^{m-1}}\}$ genau die Restklassen $\overline{1}, \overline{2}, \dots, \overline{m-1}$ durchlaufen. Ordnet man diese in der Reihenfolge, wie sie in $\overline{g}, \overline{g^2}, \dots, \overline{g^{m-1}}$ auftreten, so erhält man eine scheinbar zufällige Anordnung der Zahlen von 1 bis $m - 1$. Für $m = 19$ und $g = 3$ erhielten wir beispielsweise:

$$\{\overline{3}, \overline{9}, \overline{8}, \overline{5}, \overline{15}, \overline{7}, \overline{2}, \overline{6}, \overline{18}, \overline{16}, \overline{10}, \overline{11}, \overline{14}, \overline{4}, \overline{12}, \overline{17}, \overline{13}, \overline{1}\}.$$

Nun aber zurück zum angestrebten Vergleich mit der Kunst. Mir scheint die Parallele zu den Gegensätzen zwischen der klassischen Malerei und deren neuzeitlichen Stilrichtungen im 20. Jahrhundert besonders erwähnenswert. Während erstere großen Bedacht darauf legte, Regelmäßigkeiten und Symmetrien aufzuzeigen, rückte in der Moderne immer mehr das zufällige Ereignis in den Mittelpunkt des Interesses, wobei allerdings nicht gesagt ist, daß hinter aller Zufälligkeit nicht eine versteckte Ordnung beziehungsweise Gesetzmäßigkeit steckt, wie dies auch bei der betrachteten Zufallsfolge $\{\bar{g}, \bar{g}^2, \dots, \bar{g}^{m-1}\}$ der Fall ist.

Dieser Vergleich läßt sich sogar noch etwas vertiefen: in der surrealistischen Kunst wird mit klassischen Methoden versucht, unerwartete Zusammenhänge darzustellen und ich möchte nun kurz erläutern, wie dieser Aspekt auch in einer Anwendung der Zufallsfolgen auftaucht.

Man stelle sich dazu ein Gebiet mit Flächeninhalt 1 vor und darin ein Quadrat, dessen Flächeninhalt es abzuschätzen gilt. Die sogenannte Monte Carlo Methode besteht nun darin, nach irgendeinem Schema eine Zufallsfolge von N Punkten zu erzeugen, die willkürlich in dem Gebiet verteilt sind. Die Erzeugung einer solchen ist schwierig und im Grund ein Paradoxon; aber die Quasizufallsfolgen, die von der Zahlentheorie herrühren, können anstelle von Zufallsfolgen benützt werden.

Das unerwartete Ergebnis ist nun, daß, obwohl bei einer Zufallsfolge die Zahl N' der Punkte in einem Quadrat des Flächeninhalts F annähernd gleich NF ist, egal wie die Punkte der Zufallsfolge auch verteilt sein mögen, immer eingeschriebene Quadrate existieren, deren Inneres entweder wesentlich zu wenige oder wesentlich zu viele Punkte enthält, sodaß N' verhältnismäßig stark von NF abweicht. Dies konnte von der holländischen Mathematikerin van Aardenne - Ehrenfest 1945 unter Verwendung ganz klassischer Methoden gezeigt werden, obwohl diese Erkenntnis so massiv unserer Anschauung widerspricht, wie etwa das Bild einer Treppe ohne oben und unten, Anfang und Ende, das von einem Maler des Surrealismus auch nur mit Pinsel und Farbe erschaffen wird.

Im Anschluß an diesen Vergleich möchte ich nun zu einem weiteren Anwendungsgebiet der Zahlentheorie kommen, das ich bereits kurz angesprochen habe, der Kryptographie. Auch hier werden einige Parallelen zur Kunst augenscheinlich: Geheimschrift, die deutsche Übersetzung von Kryptographie, läßt bereits erahnen, daß es sich um das Wechselspiel zwischen Verborgenen und Offenkundigen handelt, das ja auch in der Musik, der Malerei, usw. ständig präsent ist. Historisch gesehen war Julius Cäsar

der erste, der sich eines Geheimcodes bediente, um eine verschlüsselte Botschaft an seine Truppen zu senden. Das dafür verwendete Kryptosystem, das heißt die Methode nach der ein Text verschlüsselt und entschlüsselt wird, war denkbar einfach: jeder Buchstabe wurde dabei durch denjenigen ersetzt, der im Alphabet 3 Stellen hinter ihm lag, wenn eine Botschaft also den Text ANGRIFFF enthielt, so war codiert der Chiffretext DQJULII zu lesen.

Auf diese Weise waren Botschaften schon recht gut zu codieren, allerdings nur im Falle strengster Geheimhaltung, da bereits die Kenntnis der Verschlüsselungsmethode automatisch die Tür zur Entschlüsselung öffnet. Diese sogenannten Privatschlüssel-Kryptosysteme haben weiter den Nachteil, daß nur zwei eingeweihte auf diese Art kommunizieren können. Lange Zeit war es fraglich, wie man diesen Makel beheben kann, bis die Mathematiker Rivest, Shamir und Adelman 1976 eine Art öffentliches Kryptosystem ersannen, bei dem der Verschlüsselungsschlüssel öffentlich preisgegeben werden kann, ohne daß damit automatisch auch die Entschlüsselungstechnik offenbar würde. Der große Vorteil dieser Methode liegt darin, daß der Initiator eines Geheimcodes nun von jedermann Geheimbotschaften empfangen kann (die Codierungsweise ist ja öffentlich bekannt), die nur er in der Lage ist, zu dechiffrieren. Um dieses sogenannte RSA Schema zu erläutern, bedarf es wiederum einiger Vorbereitungen.

Zunächst wird jedem Buchstaben des Alphabets eine Zahl zugeordnet

$$A \leftrightarrow 00, B \leftrightarrow 01, \dots, Z \leftrightarrow 25$$

um das Problem den Methoden der Zahlentheorie zugänglich zu machen. Das grundlegende Ergebnis auf dem die Codierungstheorie beruht ist unter dem Namen "Kleiner Fermat" bekannt und lautet: ist p eine Primzahl und gilt $x \equiv 1 \pmod{p-1}$, so folgt daraus $a^x \equiv a \pmod{p}$ für jede ganze Zahl a .

Wie kann man dies nun ausnützen, um eine Geheimschrift zu erzeugen? Zunächst wählt Person A eine Primzahl p sowie zwei Zahlen V und E aus, die der Bedingung

$$VE \equiv 1 \pmod{p-1}$$

genügen. Sodann gibt A die Zahlen p und V als Verschlüsselungscode bekannt, behält aber E als Entschlüsselungsschlüssel geheim. Jede Person B , die (p, V) kennt, kann nun codierte Botschaften an A senden nach folgendem Schema:

Jeder Text wird zunächst als Zahl t geschrieben (wobei $0 \leq t < p$ vorausgesetzt wird) und dieser sodann die Zahl c zugeordnet, die durch die

Kongruenz

$$c \equiv t^V \pmod{p}$$

bestimmt wird.

Zur Dechiffrierung geht A nun vom chiffrierten Text in Form der Zahl c aus und bestimmt sodann die Lösung t^* , $0 < t^* < p$, der Kongruenz

$$t^* \equiv c^E \pmod{p}.$$

Der "Kleine Fermat" besagt nun:

$$t^* \equiv c^E \equiv (t^V)^E = t^{VE} \equiv t \pmod{p}$$

und da t^* und t beide zwischen 0 und $p - 1$ liegen, müssen sie gleich sein und die Botschaft ist vollständig wiederhergestellt. Dazu ein Beispiel zur Veranschaulichung:

Wir wählen die Primzahl $p = 11$ und müssen zwei Zahlen V und E finden, deren Produkt in der Restklasse $\bar{1}$ modulo 10 liegt, das heißt

$$EV \equiv 1 \pmod{10}.$$

$E = 7$ und $V = 3$ erfüllen diese Bedingung und wir können $(11, 3)$ als Verschlüsselungsschlüssel bekannt geben. Um die Zahl $t = 9$ zu codieren, berechnen wir c durch

$$c \equiv t^V = 9^3 \equiv 3 \pmod{11}$$

und dechiffriert wird dann mittels

$$t^* \equiv c^E = 3^7 = 2187 \equiv 9 \pmod{11}.$$

So weit, so gut; bleibt nur noch die Frage, ob dieses System die gestellten Anforderungen erfüllt, das heißt, ist es wirklich unmöglich, aus der Kenntnis des Verschlüsselungsverfahrens (p, V) auf die Entschlüsselung (p, E) zu schließen? Dies ist allerdings nicht der Fall wegen $EV \equiv 1 \pmod{p - 1}$, so daß E ein Reziprokwert von $V \pmod{p - 1}$ ist und dieser Vorschlag noch etwas zusätzlicher Verfeinerung bedarf, um effektiv zu sein.

Eine Folgerung aus dem kleinen Satz von Fermat besagt: sind $p \neq q$ Primzahlen und $m = pq$, so gilt:

$$x \equiv 1 \pmod{(p - 1)(q - 1)} \Rightarrow a^x \equiv a \pmod{m}$$

für jedes a und wir wollen nun versuchen, mit Hilfe dieses allgemeineren Ergebnisses die Strategie von vorher zum Erfolg zu führen.

Diesmal wählt Person A zwei Primzahlen $p \neq q$ aus, bildet deren Produkt m und sucht zwei Zahlen V und E , die der Kongruenz

$$VE \equiv 1 \pmod{(p-1)(q-1)}$$

genügen. Der öffentlich bekanntgegebene Verschlüsselungsschlüssel besteht nun aus dem Paar (m, V) , während E , aber auch p und q (!) geheim bleiben. Im weiteren verläuft alles analog zu vorher: der Text t mit $0 \leq t < m$ wird codiert via

$$c \equiv t^V \pmod{m}$$

und danach entschlüsselt durch

$$t^* \equiv c^E \pmod{m}$$

und wieder gilt $t^* = t$ aufgrund des verallgemeinerten "Kleinen Fermat", sodaß A in den Besitz des ursprünglichen Texts kommt.

Der Vorteil dieser Methode liegt nun darin, daß man aus der Kenntnis von (m, V) nicht mehr direkt auf E schließen kann. In der Tat sind E und V durch

$$VE \equiv 1 \pmod{(p-1)(q-1)}$$

verknüpft, das heißt, man müßte p und q kennen um E aus V zu berechnen. Dies führt zu dem Problem bei gegebenem Produkt auf die einzelnen Faktoren zu kommen und das *RSA* Schema lebt von der Tatsache, daß dies in der Praxis bei mehr als hundertstelligen Primzahlen nur mit gigantischem Rechen- und Zeitaufwand möglich ist.

Diese und ähnliche Aufgaben fallen unter den Begriff der Numerik, wo man sich mit den beiden Fragen

- ist eine gegebene Zahl m prim?
- finde die Primfaktorzerlegung von m

befaßt. Wesentlich ist dabei immer der dazu nötige Rechenaufwand, wie dies ja schon am Beispiel der Entschlüsselung eines *RSA* Codes klar wird.

Um zu zeigen, daß eine Zahl m nicht prim ist, genügt es, sie in ein Produkt $m = pq$ zu zerlegen; um diese Zerlegung zu verifizieren kommt man

dann mit einer einzigen Rechenoperation aus. Das ist prinzipiell von Interesse, allerdings kann auch das schon ziemlich aufwendig sein, wenn die auftretenden Zahlen sehr groß sind. Ein Maß dafür ist die Anzahl der nötigen Ziffernoperationen, das sind Operationen mit den Ziffern $0, 1, \dots, 9$. Bezeichnet $z(m)$ die Anzahl der Ziffern von m , so kommt man mit weniger als $z(p) \cdot z(q) \leq z(m)^2$ Ziffernoperationen aus; diese Schranke wurde 1971 von Schönhage und Strassen noch auf $z(m)^{1+\varepsilon}$ gedrückt, wobei ε eine beliebig kleine positive Zahl darstellt.

Viel schwieriger ist es allerdings von einer Zahl nachzuweisen, daß sie prim ist. Ein ebenso bemerkenswertes wie überraschendes Resultat aus der Logik besagt nun, daß dafür ebenfalls endlich viele (≤ 87) Rechenoperationen ausreichen. An dieser Stelle sei nochmals auf die bereits aufgezeigten Parallelen zum Surrealismus hingewiesen: mit den formalsten Techniken (nämlich der Logik) werden Resultate (das Analogon zu Effekten in der Kunst) erreicht, die völlig überraschend und unerwartet scheinen. Hier ist es das Resultat von Jones aus dem Jahre 1976, das besagt: es gibt ein Polynom in 26 Variablen, das bei positiven Argumenten nur Primzahlen als Wert annimmt, und zwar jede mindestens einmal. Wieder ist dieses prinzipielle Ergebnis für die Praxis nutzlos, da die fraglichen Zahlen gigantisch sein können. Fragt man daher abermals nach einer Schranke für die Anzahl der nötigen Ziffernoperationen, so konnte gezeigt werden, daß man größenordnungsmäßig mit $z(m)^4$ das Auslangen findet. Dahinter steckt natürlich einiges an Theorie, unter anderem das Ergebnis: ist von der Zahl m die Zerlegung von $m - 1$ in ein Produkt von Primfaktoren bekannt, das heißt $m - 1 = q_1 \dots q_k$ und findet man eine Zahl g , für die $g^{\frac{m-1}{q_i}} \equiv 1 \pmod{m}$ für sämtliche i , so ist m eine Primzahl.

Die konkreteste Problemstellung der Numerik ist es aber, eine gegebene Zahl m tatsächlich in Primfaktoren zu zerlegen. Während dies 1970 für 20-stellige Zahlen gerade noch möglich war, haben diverse neue Algorithmen (Kettenbruchalgorithmus, quadratisches Sieb, Zahlkörpersieb, und Lenstras elliptischer Kurven-Algorithmus) die Grenzen des Machbaren auf 130-stellige Zahlen ausgedehnt.

Dabei ist die Sache im Prinzip ganz einfach: man dividiere die gegebene Zahl durch alle darunterliegenden, um alle Teiler und somit alle Primfaktoren ausfindig zu machen. Nach kurzer Überlegung wird man außerdem draufkommen, daß man lediglich die Zahlen bis zur Quadratwurzel von m durchzuprobieren hat, doch auch das übersteigt für große Zahlen die Leistungsfähigkeit jeglicher Computer, sodaß nur die bereits angesprochenen

Algorithmen das Rechnen im Bereich von 100-stelligen Zahlen erlauben.

Dies ermöglicht es mir, ein letztes Mal die Brücke zur Kunst zu schlagen: die eben erwähnten ebenso geistreichen wie phantasievollen Methoden im Umgang mit großen Zahlen rechtfertigen es durchaus, von Rechenkunst zu sprechen und somit Mathematik selbst zur Kunst zu erheben, zur Kunst des formalen Denkens eben.

Anmerkungen

- ¹ Niederschrift eines Vortrags von Dr. Wolfgang M. Schmidt, Professor am Institut für Zahlentheorie der Universität Boulder, Colorado; redigiert von Dr. Leonhard Summerer.
Unterstützt vom Jubiläumsfonds der Österreichischen Nationalbank.

Graphentheorie: Strukturen und Symmetrien

Gottfried Tinhofer

Einleitung

Unter den Exponaten einer Ausstellung in einer Kirche in Stralsund fand ich eine Komposition, wie sie in Abbildung 4.1a (Farbabbildungen: S. 139) wiederzugeben versucht wurde. Die Farben waren natürlich anders, schöner, als sie mit Zeichenprogramm und Farbdrucker erzeugt werden können. Aber die Anordnung der Farben war genau dieselbe, und auf diese Anordnung möchte ich hier hinweisen. Der Künstler vermittelte dadurch (sehr wahrscheinlich unbewußt) mathematische Informationen über eine gewisse Struktur. Es handelt sich um die Beschreibung einer sogenannten *kohärenten Konfiguration*, welche die Symmetrieeigenschaften eines sehr einfachen Graphen, eines Zyklus über sieben Punkte, erfaßt. Ein solcher Zykel ist in Abbildung 4.1b (Farbabbildungen: S. 139) dargestellt.

Ein Zykel soll eine "Reihenfolge" von gewissen Objekten vermitteln, die im Bild nur durch kleine Vollkreise angedeutet und somit nicht näher erläutert sind. Ein Zykel vermittelt also ein "Hintereinander" und durch seine Geschlossenheit auch ein "immer wieder". Ein Zykel ist daher ein abstrakter Begriff, der durch Anwendung auf verschiedene Objekte verschiedene Konkretisierungen erlaubt. Daher sind seine Symmetrien auch abstrakte Symmetrien, die keiner Erklärung durch die Symmetrieeoperationen des Raumes (Drehungen, Spiegelungen und deren Kombinationen) bedürfen und daher auch unabhängig von einem Raumbegriff existieren. Wir nennen sie auch *kombinatorische* Symmetrien. Ihre Darstellung führt bei komplizierteren Graphentypen, als es der Zykel ist, auch zu komplizierteren Farbgebilden. Im letzten Abschnitt dieses vorliegenden kleinen Exkurses über die Inhalte der Graphentheorie kommen wir nochmals auf diese kombinatorischen Symmetrien von Graphen zu sprechen. Sie werden dort erklärt und es wird auch ein komplizierteres Beispiel angegeben.

Ein Zykel ist ein sehr einfacher Graph. *Graphen* im allgemeinen sind abstrakte mathematische Modelle zum Studium der Beziehungen zwischen den

Elementen einer unter Beobachtung stehenden Menge von Objekten (Personen, Dingen oder Begriffen). Diese Beziehungen definieren eine *Struktur* auf der Beobachtungsmenge. Wir können daher auch sagen: Graphen dienen zur Synthese oder Analyse von Strukturen auf Mengen.

Obwohl Graphen abstrakte Gebilde sind, kann man sie "bildhaft" denken, man kann sie auf jede nur denkbare Art "zeichnen" und dadurch Überlegungen abstrakter Natur durch anschauliche Analoga simulieren. Daher eignen sich graphentheoretische Überlegungen vorzüglich zur Demonstration formaler Denkprozesse. In den folgenden Abschnitten wird versucht, diese Möglichkeit anhand einer kleinen Einführung in eine Auswahl der Hauptinhalte der Graphentheorie aufzuzeigen. Im ersten Abschnitt wird der Begriff Graph erläutert. Dann folgt ein Abschnitt über Eulersche Linien und das Postbotenproblem. Im nächsten Abschnitt werden elektrische Netzwerke als Graphen interpretiert, es kommt der Baumbegriff zur Sprache und wir erwähnen das Problem der Zufallswanderungen auf Graphen. Der nächste Abschnitt handelt von planaren Graphen und vom Problem der Realisierung elektrischer Netze. Im fünften Abschnitt wird der Begriff der Zuverlässigkeit von Kommunikationsnetzen mit Hilfe von Zusammenhangseigenschaften der den Netzen zugrunde liegenden Graphen besprochen. Dann folgt ein Abschnitt über Rundreisen und Tourenplanung. Der siebte Abschnitt befaßt sich mit zulässigen Färbungen von Graphen und mit dem Vierfarbenproblem, er leitet über zur Diskussion von perfekten Graphen. Im letzten Abschnitt wird schließlich die Rolle der Graphentheorie bei der Beschreibung molekularer Strukturen in der Chemie angedeutet und der in der Einleitung erwähnte Begriff der kombinatorischen Symmetrie erklärt.

Graphen

Einfache Beispiele für Graphen finden wir bei der Betrachtung von dreidimensionalen Polyedern, also beim Würfel, Quader, Tetraeder, Oktoeder, etc. Ein Polyeder besitzt *Ecken*, *Kanten* und *Flächen*. Man denke an einen Würfel. Die Menge der Ecken zusammen mit der Menge der Kanten bilden einen Graphen, den man auch das *Skelett* des Polyeders nennt. In Abbildung 4.2a sieht man das Skelett eines Würfels. Die Ecken tragen Nummern, die als Namen verwendbar sind. Wir können von der Ecke 1, der Ecke 2, usw. reden. Die Kanten verbinden solche Ecken, sie drücken die Möglichkeit aus, auf dem Rand des Polyeders von einer Ecke i direkt, das heißt ohne eine

weitere Ecke, Kante oder Fläche zu treffen, nach einer Ecke j zu gelangen. Offenbar ist diese Möglichkeit z.B. zwischen den Ecken 1 und 2, 1 und 4, 1 und 5, aber nicht zwischen den Ecken 1 und 3, 1 und 6, usw. vorhanden. Die Gesamtheit dieser Möglichkeiten und Nicht-Möglichkeiten bildet eine (abstrakte) Struktur auf der Eckenmenge, eben die Struktur des Würfels.

Bei der mathematischen Modellierung und Lösung von Problemen in nicht-mathematischen Wissensbereichen (Physik, Chemie, Technik, Operations Research, Psychologie, Soziologie, usw.) treten Polyeder auf, die viel komplizierter sind als der Würfel oder die anderen oben angesprochenen Platonischen Körper. Erstens sind diese Polyeder meist *hochdimensional*, das heißt man kann sie nur in einem hochdimensionalen abstrakten Raum denken, und zweitens besitzen sie in der Regel sehr viele (tausende oder sogar hunderttausende und mehr) Ecken und Kanten. Der zu dem Polyeder gehörende Graph, das heißt sein Skelett, ist demnach ebenfalls sehr groß.

Bei einem sogenannten *Linearen Programm* liegt die Beschreibung eines Polyeders in Form eines linearen Ungleichungssystems vor. Die Ecken dieses Polyeders sind durch eine sogenannte *Zielfunktion* bewertet, wobei der Wert einer Ecke eine lineare Funktion ihrer Koordinaten ist. Die Aufgabe besteht darin, eine Ecke mit möglichst großem Wert zu finden. Anhand der Koordinaten kann man leicht erkennen, ob ein Raumpunkt eine Ecke des vorliegenden Polyeders ist oder nicht. Anhand der Koordinaten einer Ecke A kann man auch leicht die Koordinaten der Ecken bestimmen (durch eine Rechnung), die mit A durch eine Kante verbunden sind. Solche Ecken nennt man *Nachbarn* von A . Ein klassisches Lösungsverfahren für Lineare Programme startet in einer Ecke des gegebenen Polyeders und berechnet eine Folge von "immer besseren" benachbarten Ecken, bis die "beste" Ecke erreicht wurde. Man kann dieses Verfahren als eine *Wanderung* auf einem Graphen (dem Skelett des Polyeders) betrachten. Solche Wanderungen auf beliebigen Graphen, die möglicherweise nicht das Skelett eines Polyeders sind, sind auch bei Lösungsverfahren für zahlreiche andere Problemtypen notwendig.

Das Würfelskelett besteht aus einer Eckenmenge $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$ und aus einer Menge von Kanten E . Wir schreiben $[i, j]$ für eine Kante zwischen den Ecken i und j . Auf diese Weise ergibt sich

$$E = \{[1, 2], [1, 4], [1, 5], [2, 3], [2, 6], [3, 4], [3, 7], [4, 8], [5, 6], [5, 8], [6, 7], [7, 8]\}.$$

Dieser Sachverhalt ist charakterisierend für jeden Graphen. Man definiert:

Ein Graph G besteht aus einer Eckenmenge V und einer Menge E von Kanten zwischen diesen Ecken.

Damit wir mit diesen Mengen arbeiten (rechnen) können, benötigen wir Namen für deren Elemente. Oft numeriert man die Menge V einfach durch und verwendet die Nummern als Namen für die Ecken. Statt Nummern kann man natürlich auch beliebige Buchstaben a, b, c, \dots , oder indizierte Buchstaben v_1, v_2, v_3, \dots , verwenden. Sind n Ecken gegeben, so vereinbaren wir daher oft $V = \{v_1, v_2, \dots, v_n\}$, oder einfacher $V = \{1, 2, \dots, n\}$. Statt "Ecke" sagt man in der Graphentheorie auch "Knoten" und nennt V dann die *Knotenmenge* von G . Dies ist sogar die gebräuchlichere Bezeichnung, sie soll von nun ab auch in diesen Ausführungen verwendet werden.

Kanten bezeichnen wir wie oben durch Angabe der Knoten, die sie verbinden, also je nach Wahl der Namen für die Knoten durch $[i, j]$ oder $[v_i, v_j]$. Gibt es, was vorkommen kann, mehrere Kanten zwischen denselben Knoten i und j , so unterscheiden wir diese durch Indizes, schreiben also $[i, j]_1, [i, j]_2, \dots$, etc. Oft ist die Beziehung zwischen den Knoten einer Menge, die man mit Hilfe eines Graphen modellieren möchte, nicht symmetrisch. Dann ist es notwendig, die Kanten zu *orientieren*. Von den zwei Knoten i und j , zwischen denen eine orientierte Kante verläuft, ist dann einer der *Anfangsknoten* und einer der *Endknoten*. Statt orientierte Kante sagt man häufiger *gerichtete Kante*. Ein Graph mit gerichteten Kanten heißt auch *gerichteter Graph*. Eine von i nach j gerichtete Kante bezeichnet man durch (i, j) . Der Zykel in Abbildung 4.1b (Farbabbildungen: S. 139) ist ein gerichteter Graph.

Jeder Graph läßt sich zeichnen. Dazu wählt man eine Zeichenebene und dort für jeden Knoten i einen Punkt. Für jede Kante $[i, j]$ zeichnet man eine Verbindungslinie zwischen den Punkten, welche die Knoten i und j repräsentieren. In Abbildung 4.2a ist das Würfelskelett gezeichnet. Da es bei einem Graphen jedoch nur darauf ankommt, *welche Knoten durch eine Kante verbunden sind und welche nicht*, ist es beim Zeichnen dieses Skeletts nicht notwendig, die Würfelform anzudeuten. Die Zeichnung in Abbildung 4.2b zeigt denselben Graphen. Diese Zeichnung hat den Vorteil, daß sich ihre Kanten nirgends überschneiden. Wir sprechen daher von einer *planaren* Darstellung des Würfelgraphen. Schließlich stellt auch die Zeichnung in Abbildung 4.2c denselben Graph dar. Die drei Zeichnungen in Abbildung 4.2 sind also nur verschiedene Darstellungen eines einzigen Graphen, eben des Würfelskeletts. Schließlich sei erwähnt, daß Kanten nicht unbedingt durch gerade Linien dargestellt werden müssen, ein Bogen oder eine beliebige an-

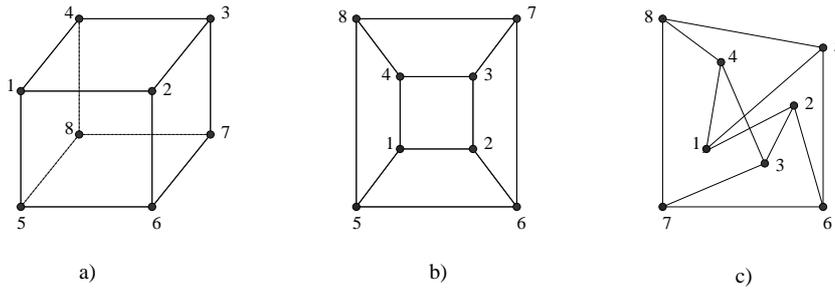


Abbildung 4.2.

dere stetige Kurve zwischen den Endknoten leistet dasselbe. In späteren Beispielen werden wir von dieser Möglichkeit Gebrauch machen.

Die Tatsache, daß man Graphen "zeichnen" kann und man dadurch eine visuelle Vorstellung von der Sachlage, die man untersuchen will, erhält, hat viel dazu beigetragen, daß die Theorie dieses mathematischen Modells, eben die Graphentheorie, bis zu dem heutigen enormen Umfang ausgebaut wurde. Wir haben gesehen, daß ein Graph unterschiedliche Zeichnungen zuläßt. Je nach Art des Zieles, das man verfolgt, wird diese oder jene Zeichnung vorteilhafter (anschaulicher) sein. Die verschiedenen Möglichkeiten bei der Darstellung eines Graphen durch ein Bild sind selbst ein wichtiger Untersuchungsgegenstand. Zum Beispiel hat nicht jeder Graph eine planare Darstellung. Wir kommen zu diesem Sachverhalt in einem späteren Abschnitt zurück.

Graphen, die man wie das Würfelskelett, in einer Ebene ohne Kantenüberschneidung zeichnen kann, nennt man *planare Graphen*. Sie stellen einen wichtigen Graphentyp dar. In der Informatik befaßt man sich mit der Frage, wie man planare Graphen möglichst schnell erkennen und unter Einhaltung von zahlreichen verschiedenen Nebenbedingungen zeichnen kann.

Die Hauptanwendungsbereiche der Graphentheorie sind Chemie, Physik, Operations Research, und vor allem Informatik. In der Chemie werden Graphen zum Studium von Molekül- und Kristallstrukturen verwendet. In der Biologie und in der Physik modelliert man Konfigurationsräume mit Hilfe von (in der Regel sehr großen und stark regulären) Graphen. Die Knoten dieser Graphen entsprechen dabei den Zuständen eines physikalischen oder biologischen Systems, die Kanten entsprechen den möglichen Übergängen zwischen diesen Zuständen. Die Eigenschaften solcher Konfigurationsgraphen gestatten dann Rückschlüsse auf die Art der physika-

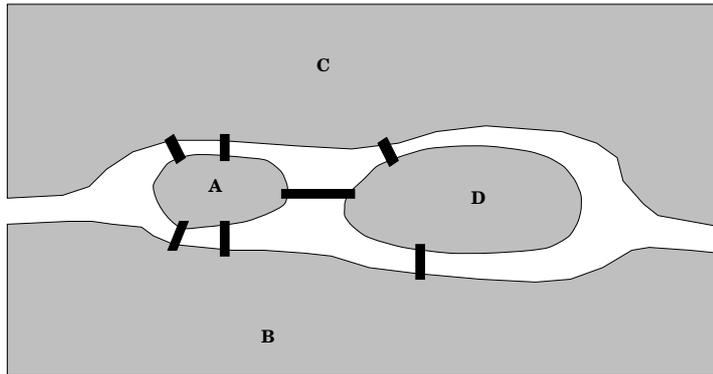


Abbildung 4.3.

lischen oder biologischen Vorgänge in den unter Untersuchung stehenden Systemen. Im Rahmen des Operations Research werden Graphen zur Modellierung und Planung von industriellen Prozessen herangezogen. In der Informatik benötigt man Graphen unter anderem bei der Untersuchung von Rechnerarchitekturen und von Datenstrukturen, in der Bildverarbeitung, bei der Analyse von Algorithmen und in vielen weiteren Teilbereichen.

Das Postbotenproblem

Zu den Anfängen der Graphentheorie rechnet man gewöhnlich die Formulierung und Lösung des sogenannten *Königsberger Brückenproblems* durch den berühmten Mathematiker L. Euler¹ [Euler, 1736 (1741)]. In Abbildung 4.3 ist schematisch der Lauf des Flusses Pregel durch die Stadt Königsberg angedeutet. Teile der Stadt liegen auf den zwei durch A und D markierten Inseln in der Pregel. Die anderen Stadtteile liegen auf den beiden durch B und C markierten Ufern. Vier Brücken verbinden die beiden Ufer mit der Insel A, zwei Brücken mit der Insel D. Ferner führt eine Brücke von A nach D.

Das Schema im Bild gibt den Zustand zur Zeit eines Aufenthaltes Eulers in Königsberg wieder. Eulers Frage war: *Kann man einen Rundgang finden, bei dem man von einem beliebigen Standpunkt (A, B, C oder D) aus jede der sieben Brücken genau einmal überquert und dann zum Ausgangspunkt zurückkehrt?*

Während einer Wanderung über die Brücken wechseln vier mögliche Positionen einander ab. Man kann sich im Stadteil A, B, C oder D befinden. Diese vier Positionen erklären wir als die vier Knoten eines Graphen,

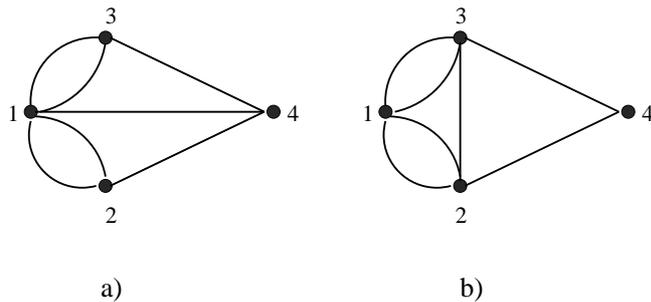


Abbildung 4.4.

die wir durch dieselben vier Großbuchstaben oder einfach auch die Nummern 1, 2, 3 und 4 unterscheiden. Wir verbinden zwei Knoten, wenn eine Brücke die dadurch erfaßten Positionen verbindet. So entsteht der in Abbildung 4.4a gezeichnete Graph. Da von A nach B und von A nach C jeweils zwei Brücken führen, enthält dieser Graph je zwei Kanten zwischen 1 und 2 und 1 und 3. Das Eulersche Problem lautet dann: *Kann man die Kanten des Graphen in Abbildung 4.4a so anordnen, daß sie einen geschlossenen Weg durch diesen Graphen bilden?*

Die letzte Frage hat auch einen Sinn, wenn statt des Graphen aus Abbildung 4.4a ein beliebiger anderer Graph vorliegt. Zur allgemeinen Formulierung des Eulerschen Problems und auch für spätere Zwecke führen wir zunächst einige notwendige Begriffe ein. Ein *offener Kantenzug* in einem Graphen ist eine Folge $[x_1, x_2], [x_2, x_3], \dots, [x_{k-1}, x_k]$ von paarweise verschiedenen Kanten, wobei je zwei aufeinander folgende Kanten einen Knoten gemeinsam haben. Dabei sind x_1, x_2, \dots, x_k Knoten, die nicht alle verschieden sein müssen. In Abbildung 4.5a ist der Kantenzug $[2, 3], [3, 5], [5, 6], [6, 4], [4, 3], [3, 1]$ enthalten. Hier ist $k = 7$ und $x_1 = 2, x_2 = 3, x_3 = 5, x_4 = 6, x_5 = 4, x_6 = 3, x_7 = 1$. Sind alle x_i verschieden, so heißt der Kantenzug ein *Weg* zwischen x_1 und x_k . Ist $x_1 = x_k$, so nennt man den Kantenzug *geschlossen*. Sind nur x_1 und x_k gleich, alle anderen Knoten aber paarweise verschieden, so heißt der Kantenzug ein *Kreis*. Wege und Kreise sind wichtige, zur Beschreibung von Struktureigenschaften benötigte Teile eines Graphen. In Abbildung 4.5a findet man zum Beispiel den Weg $[3, 5], [5, 6], [6, 7]$ und den Kreis $[1, 2], [2, 7], [7, 8], [8, 1]$. Ein Graph heißt *zusammenhängend*, wenn es in ihm zwischen je zwei Knoten mindestens einen Weg gibt. Der Graph in Abbildung 4.5a ist offensichtlich zusammenhängend.

Ein offener Kantenzug, der alle Kanten des Graphen (genau einmal)

enthält, heißt *offene Eulersche Linie*. Ein geschlossener Kantenzug mit dieser Eigenschaft heißt *geschlossene Eulersche Linie*. Unter Verwendung dieser Begriffe lautet nun die Verallgemeinerung des Königsberger Brückenproblems: *Gegeben ein Graph G , bestimme, ob G eine geschlossene Eulersche Linie hat oder nicht.*

Graphen, die eine geschlossene Eulersche Linie besitzen, wurden bereits von Euler charakterisiert. Ganz offensichtlich sind solche Graphen zusammenhängend. Geht man in einem Graph G eine vorhandene geschlossene Eulersche Linie in der Reihenfolge ihrer Kanten ab, so kommt man dabei in jedem Knoten genau so oft an, wie man davon abgeht. Da bei jeder Ankunft und bei jedem Abgang eine neue Kante benützt wird, müssen in jedem Knoten eine gerade Anzahl von Kanten "anliegen". Die Anzahl der Kanten, die ein Ende in einem Knoten v haben, nennt man den *Grad* $d(v)$ dieses Knoten. Hat also G eine geschlossene Eulersche Linie, so ist für jeden Knoten v dieses Graphen der Grad $d(v)$ eine gerade Zahl. Für einen zusammenhängenden Graphen erweist sich diese notwendige Bedingung auch als hinreichend. Somit können wir das Eulersche Ergebnis so formulieren: *Ein zusammenhängender Graph hat genau dann mindestens eine geschlossene Eulersche Linie, wenn alle seine Knotengrade gerade sind.*

Der Graph in Abbildung 4.4a ist zusammenhängend, er erfüllt aber nicht die Knotengradbedingung. Das Königsberger Brückenproblem ist also mit "Nein" zu beantworten. Es gibt mehrere Möglichkeiten, dieses "Nein" in "Ja" zu verwandeln. Zum Beispiel könnte man die Brücke zwischen den beiden Inseln (die Kante [1, 4]) entfernen, und dafür die beiden Ufer B und C direkt durch eine Brücke verbinden (die Kante [2, 3] einführen). Der Graph in Abbildung 4.4a geht dabei über in den Graph in Abbildung 4.4b, der die Knotengradbedingung erfüllt. Der Leser wird selber leicht eine geschlossene Eulersche Linie in diesem neuen Graphen finden.

Betrachten wir nun nochmals den Graph in Abbildung 4.5a. Man nehme an, daß die Kanten dieses Graphen ein System von Straßenzügen modellieren, das ein Postbote in einem Rundgang zu bedienen hat. Die Knoten bedeuten in diesem Graphen Straßenkreuzungen, also Punkte des Straßensystems, in denen man von einer Straße in eine andere wechseln kann. Ein Postbote möchte seinen Rundgang sicher so planen, daß er von einem festen Anfangspunkt ausgehend möglichst wenig Wegstrecken mehrmals (vor und zurück) zu gehen hat und dann wieder zum Ausgangspunkt zurückkommt. Ideal wäre natürlich, wenn er überhaupt keine Wegstrecke mehrfach bege-

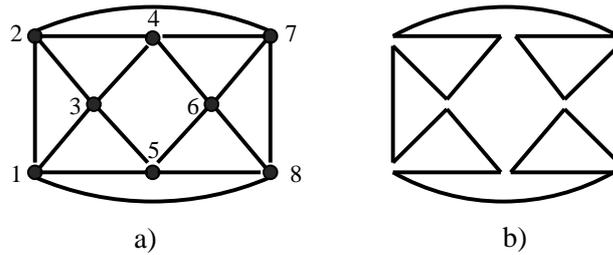


Abbildung 4.5.

hen müßte. Ist so ein Rundgang möglich? Der Graph in diesem Beispiel ist zusammenhängend und alle Knotengrade sind gerade. Daher besitzt er geschlossene Eulersche Linien. Eine davon ist in Abbildung 4.5b angedeutet.

In der Realität findet man natürlich auch Straßensysteme vor, deren Modellgraph nicht die Gradbedingung erfüllt. Dann existiert kein Rundgang, bei dem man jede Straße genau einmal abgeht. Will man trotzdem einen Rundgang durchführen, so muß man also zulassen, daß eine oder mehrere Straßen doppelt oder sogar mehrfach zu begehen sind. Man wird dann seinen Rundgang eben so planen, daß möglichst wenig Wiederholungen vorkommen. In der Sprache der Graphentheorie lautet das Planungsproblem dann so: *Zu einem gegebenen zusammenhängenden Graphen finde man einen geschlossenen Kantenzug aus möglichst wenig Kanten, der aber jede Kante mindestens einmal enthält.* Dieses Problem heißt das *Postbotenproblem*. Es ist eines der Grundprobleme aus einem Problemereich, den man unter dem Titel *Tourenplanung* zusammenfaßt.

Elektrische Netze

Auch einem elektrischen Netz läßt sich ein Graph zuordnen, mit dessen Hilfe die im Netz möglichen Ströme auf einfache Weise ermittelt werden können. Als Knoten dieses Graphen treten auf: die Pole der vorhandenen Spannungsquellen, die Erdungspunkte und die Netzpunkte, an denen eine Stromverzweigung stattfinden kann. Jeder Spannungsquelle wird eine von + nach - gerichtete Kante zugeordnet. Ferner wird jedem stromführenden Zweig des Netzes eine in Stromrichtung orientierte Kante zugeordnet. Abbildung 4.6a zeigt ein einfaches elektrisches Netz, Abbildung 4.6b dessen gerichteten Graph.

Jeder stromführende Zweig des Netzes stellt einen Widerstand dar. Strom, Spannung und Widerstände stehen zueinander in einer Relation,

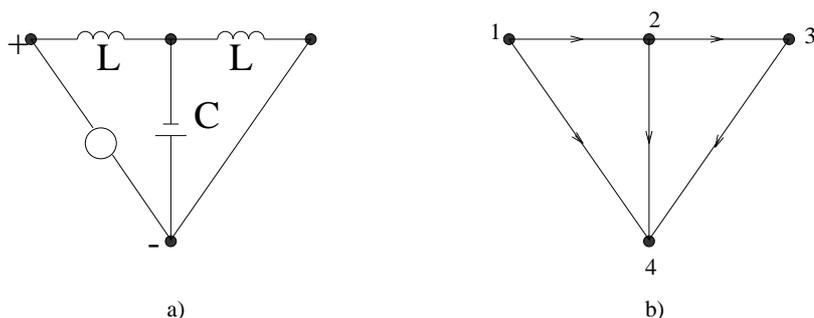


Abbildung 4.6.

die durch das Ohmsche Gesetz ausgedrückt wird. Außerdem gehorchen die Ströme in den einzelnen Zweigen des Netzes Regeln, die durch lineare Gleichungen ausgedrückt werden. Man nennt diese die *Kirchhoff'schen Regeln*.² Es sei $G = (V, E)$ der das elektrische Netz modellierende Graph und $V = \{1, 2, \dots, n\}$. Für $(i, j) \in E$ sei $I(i, j)$ der Strom, der die Kante (i, j) in Richtung von i nach j durchfließt. Die Kirchhoff'schen Regeln sind:

Knotenregel: Für jeden Knoten i gilt

$$\sum_{j:(i,j) \in E} I(i, j) - \sum_{j:(j,i) \in E} I(j, i) = 0,$$

das heißt in jedem Knoten i fließt dieselbe Stromgröße zu wie ab.

Maschenregel: Für einen Kreis C des Graphen, den man sich im Uhrzeigersinn durchlaufen denke, vereinbare man ein "Vorzeichen" $\epsilon(i, j)$ einer Kante (i, j) zu +1, wenn die Kante in Kreisrichtung liegt, und zu -1, wenn sie entgegengesetzt dazu liegt. Zum Beispiel enthält der Kreis $[1, 2], [2, 4], [4, 1]$ in Abbildung 4.6b die gerichteten Kanten $(1, 2)$ und $(2, 4)$ in seiner Richtung, die Kante $(1, 4)$ aber in Gegenrichtung. Daher ist $\epsilon(1, 2) = \epsilon(2, 4) = 1$ und $\epsilon(1, 4) = -1$. Die zweite Kirchhoff'sche Regel besagt, daß für jeden Kreis C von G

$$\sum_{(i,j) \in C} \epsilon(i, j) I(i, j) = 0,$$

das heißt längs jeden Kreises ist die Summe der Ströme, mit dem richtigen Vorzeichen versehen, gleich Null.

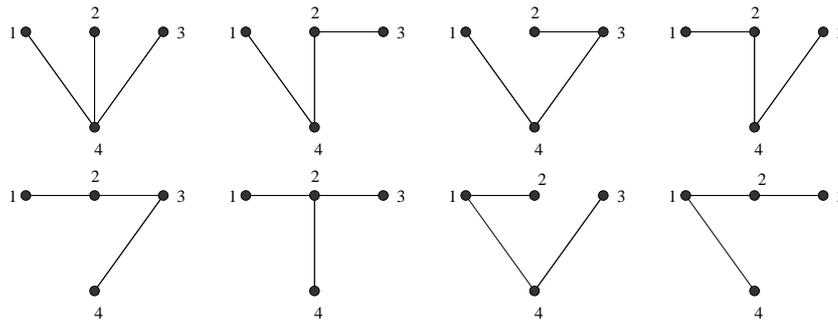


Abbildung 4.7.

Der Graph in Abbildung 4.6b ist gerichtet. Die Begriffe "Weg" und "Kreis", die in Rahmen des Eulerschen Problems eingeführt wurden, haben jedoch auch in gerichteten Graphen ihren Sinn. Man bildet Wege oder Kreise in gerichteten Graphen wie in ungerichteten, indem man einfach die Kantenrichtung ignoriert.

Enthält ein zusammenhängender Graph einen Kreis C , so kann man den Zusammenhang nicht dadurch verlieren, daß man eine Kante $[i, j]$ von C wegläßt. Denn jeder Weg, der diese Kante benützt, kann zu einem Weg zwischen denselben Knoten verändert werden, indem man andere Kanten von C verwendet (man kann von i nach j statt über $[i, j]$ auf dem restlichen Teil des Kreises C gelangen).

Nun sei $G = (V, E)$ ein beliebiger zusammenhängender Graph. Man entferne eine beliebige Kante eines Kreises von G . Hat der restliche Graph noch einen Kreis, so entferne man eine beliebige Kante davon. Diesen Prozeß setze man solange fort, bis man zu einem (immer noch zusammenhängenden) Graph $T = (V, E_T)$ gelangt, der keinen Kreis mehr enthält. Einen derartigen Graph nennt man einen *Baum*. Da T alle Knoten von G und eine Kantenmenge $E_T \subseteq E$ besitzt, heißt T auch *spannender Baum* von G .

Die verschiedenen spannenden Bäume des Graphen in Abbildung 4.6b sind in Abbildung 4.7 zu sehen.

Von R. Kirchhoff [Kirchhoff, 1847] stammt die Beobachtung, daß alle Stromgrößen in einem elektrischen Netz durch eine sehr einfache Rechnung ermittelt werden können, sobald man die Ströme in den Zweigen eines spannenden Baumes kennt. In der Tat, es sei T so ein spannender Baum und

$I(i, j)$ sei bekannt für alle Kanten (i, j) von T . Ferner sei (u, v) die letzte Kante, die bei der Ermittlung von T entfernt wurde und C der Kreis, den sie geschlossen hat. Dann gilt auf Grund der Maschenregel

$$\sum_{(i,j) \in C} \epsilon(i, j) I(i, j) = 0,$$

also

$$I(u, v) = - \sum \frac{\epsilon(i, j)}{\epsilon(u, v)} I(i, j),$$

wobei über alle $(i, j) \in C, (i, j) \neq (u, v)$ zu summieren ist. Genauso kann man die Stromgröße auf der vorletzten entfernten Kante aus den Stromgrößen des Kreises, den sie geschlossen hat, berechnen, da dort bereits alle Größen bekannt sind. So fortfahrend, in der umgekehrten Reihenfolge des Entfernens von Kanten, kann man alle noch unbekanntes Stromgrößen ermitteln.

In dem Beispiel von Abbildung 4.6 gilt, wenn wir zum Beispiel den dritten spannenden Baum verwenden:

$$I(1, 2) = -I(2, 3) - I(3, 4) + I(1, 4), \quad I(2, 4) = -I(2, 3) - I(3, 4).$$

Der hier beschriebene Sachverhalt ist wichtig bei der Behandlung von elektrischen Netzwerken. Er gestattet eine vorteilhafte Verwendung graphentheoretischer Begriffe und Methoden zur Synthese und Analyse von elektrischen Netzwerken. Für eine detaillierte Einführung in diesen Problemkreis sei der Leser auf [Chen, 1971] verwiesen.

Der Begriff "Baum" ist einer der wichtigsten graphentheoretischen Begriffe überhaupt, das heißt Bäume bilden einen der wichtigsten Graphentypen. Bäume treten bei der Behandlung verschiedenster Problembereiche auf, sei es bei der Modellierung von Alkanen (Kohlenwasserstoffen), im Umgang mit Grammatiken für Computersprachen, als Datenstrukturen, bei der Optimierung von Kommunikationsnetzwerken, und in vielen anderen Bereichen.

Der Leser betrachte nun Abbildung 4.8. Mit jedem der Graphen in den beiden Bildteilen ist ein mathematisches Problem verknüpft. Die beiden Probleme erscheinen zunächst zusammenhangslos. In Wirklichkeit kann man zeigen, daß beide Probleme insofern gleich sind, als sie dieselbe (eindeutige) Lösung besitzen.

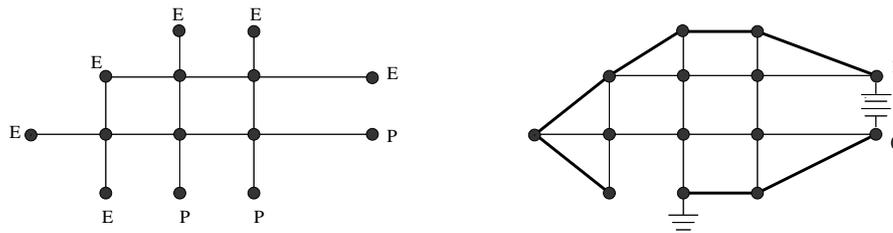


Abbildung 4.8.

Betrachten wir zuerst das linke Bild. Seine Knoten sollen Positionen in einem Straßennetz darstellen, die Kanten Straßen. Wir nehmen an, in diesem Netz hält sich in irgendeinem Knoten ein Flüchtling auf, der versucht, der Polizei zu entkommen. Ist er in einem der nicht-markierten Punkte, so wählt er mangels an Ortskenntnissen eine der vier hier einmündenden Straßen mit gleicher Wahrscheinlichkeit $\frac{1}{4}$ zur Fortsetzung seiner Flucht aus (er weiß nicht, welcher Weg günstig ist). Erreicht er einen der mit P markierten Punkte, so schnappt ihn die Polizei. Erreicht er aber einen der mit E markierten Punkte, so kann er das Netz verlassen und entkommt. Was ist nun seine Wahrscheinlichkeit für ein Entkommen, wenn der Flüchtling sich zu Beginn der Flucht im Knoten x befindet? (Dabei sei x ein beliebiger Knoten des Straßennetzes.)

Wir erkennen rasch folgenden Sachverhalt. Es sei $w(x)$ die Wahrscheinlichkeit für ein Entkommen, wenn die Flucht in x beginnt. Offenbar ist $w(x) = 0$, wenn x durch P markiert ist, und $w(x) = 1$, wenn x durch E markiert ist. Zu berechnen sind daher nur die übrigen Werte von $w(x)$. Für einen unmarkierten Knoten x sei $N(x)$ die Menge seiner Nachbarn. Von x aus kann der Flüchtling zu einem seiner Nachbarn y gehen und dort seine Flucht fortsetzen, wobei er dann mit Wahrscheinlichkeit $w(y)$ entkommt. Da er y mit Wahrscheinlichkeit $\frac{1}{4}$ auswählt, hat diese Möglichkeit zu entkommen die Wahrscheinlichkeit $\frac{w(y)}{4}$. Somit ist, wenn y_1, y_2, y_3, y_4 die vier Nachbarn von x sind

$$w(x) = \frac{1}{4}[w(y_1) + w(y_2) + w(y_3) + w(y_4)] = \frac{1}{4} \sum_{y \in N(x)} w(y).$$

Betrachten wir nun das rechte Bild. Es ist der Graph eines elektrischen Netzes. Dieser Graph unterscheidet sich von dem im linken Bild durch das Vorhandensein der dicken Kanten. Diese deuten an, daß die von ihnen berührten Punkte "kurz geschlossen" sind. Der untere Bereich ist geerdet.

Er erfaßt die im linken Abbildung durch P markierten Knoten. Diese sind hier auf Potential 0 gesetzt. Eine Spannungsquelle setzt die Knoten auf dem oberen Rand auf Potential 1. Diese Knoten entsprechen links den durch E markierten Knoten. Alle Widerstände R in den dünn gezeichneten Zweigen des Netzes seien identisch gleich 1. Die Ströme in den Zweigen des Netzes kann man mit Hilfe des Ohmschen Gesetzes $I = \frac{U}{R}$ durch die Spannung U beschreiben. Die Spannung U zwischen den Knoten x und y des Netzes läßt sich als Potentialdifferenz $P(x) - P(y)$ einer auf der Knotenmenge V definierten Potentialfunktion P ermitteln. Wir wissen bereits, daß $P(x) = 0$, wenn x ein (geerdeter) unterer Randknoten ist, und daß $P(x) = 1$, wenn x ein oberer Randknoten ist. Für jeden anderen Knoten x gilt, wenn seine Nachbarn wieder durch y_1, y_2, y_3, y_4 bezeichnet werden, auf Grund der Kirchhoff'schen Knotenregel (da $R = 1$ angenommen wurde)

$$(P(y_1) - P(x)) + (P(y_2) - P(x)) + (P(y_3) - P(x)) + (P(y_4) - P(x)) = 0.$$

Daraus folgt

$$P(x) = \frac{1}{4}[(P(y_1) + P(y_2) + P(y_3) + P(y_4))] = \frac{1}{4} \sum_{y \in N(x)} P(y).$$

Man erkennt daher: Die Bedingungen für die Festlegung der Funktionen w und P sind identisch. Jede Lösung für P ist auch eine Lösung für w , und umgekehrt.

Was sind nun diese Bedingungen, denen w und P genügen. Wir bezeichnen die markierten Knoten im linken Teil von Abbildung 4.8 und die entsprechenden Knoten im rechten Bildteil als *Randknoten*, die übrigen Knoten als *innere Knoten*. Dann können wir formulieren: Jeder Funktionswert (von w und auch von P) an einem inneren Knoten x ist gleich dem arithmetischen Mittel der Funktionswerte auf der Menge der Nachbarn von x , und die Funktionswerte auf den Randknoten sind gegeben. Jede Funktion mit diesen Eigenschaften auf der Knotenmenge V eines Graphen, die in innere Knoten und in Randknoten unterteilt ist, heißt *harmonische Funktion*. Von den Randknoten in unserem Beispiel wird folgende Bedingung erfüllt: Entfernt man in jedem Bildteil die Randknoten, so bleibt der restliche Graph zusammenhängend. Ferner hat jeder Randknoten mindestens einen inneren Knoten als Nachbarn. Man kann zeigen, daß unter dieser Bedingung zu gegebenen Funktionswerten auf dem Rand genau eine harmonische Funktion existiert. Es ist also in unserem Beispiel $w(x) = P(x)$ für alle Knoten x .³

In der ersten Fassung des obigen Problems vollzieht der Flüchtling eine Wanderung durch einen Graphen G von Knoten zu Knoten, wobei in jeder Position x der nächste Knoten gleichwahrscheinlich unter den Nachbarn von x ausgewählt wird. Eine solche Wanderung nennt man eine *Zufallswanderung* (*Random Walk*) auf dem Graphen G . Bei einer Zufallswanderung muß die Auswahl des jeweils nächsten Knoten nicht unbedingt gleichwahrscheinlich erfolgen. Statt dessen könnte man auch eine andere Wahrscheinlichkeitsverteilung verwenden. Aber auch dann ließe sich dem Problem ein analoges Problem in einem elektrischen Netzwerk zuordnen. Nur müßte man dann in diesem Netzwerk die Widerstände ebenfalls je nach Kante unterschiedlich wählen. Eine leicht lesbare Einführung in diesen Problemkreis findet man in [Doyle und Snell, 1984].

Die Tatsache, daß sich ein kombinatorisches Problem, nämlich die Ermittlung von Wahrscheinlichkeiten, durch Behandlung eines physikalischen Problems, die Ermittlung der Potentialverteilung in einem elektrischen Netz, lösen läßt, ist nur bei erster Begegnung mit solchen Sachverhalten verblüffend. Es gibt in der Mathematik viele andere Sachverhalte ähnlicher Natur. Auf der Möglichkeit, mathematische Probleme durch *analoge* physikalische Probleme (meist Probleme aus der Mechanik oder der Elektrizitätslehre) zu ersetzen, beruhen die sogenannten *Analogrechner*.

Planare Graphen

Elektrische Netze, die wir nach den Ausführungen im letzten Abschnitt als Graphen betrachten können, können auf verschiedenste Arten physikalisch realisiert werden. Dabei ist gemeint, daß die räumliche Anordnung der Netzelemente (Knoten und Kanten) sehr verschieden gewählt werden kann. Es wurde schon am Anfang dieses Aufsatzes darauf hingewiesen, daß ein Graph auf viele verschiedene Arten gezeichnet werden kann. Eine Zeichnung erfolgt in einer Ebene. Will man ein elektrisches Netz in einer Ebene realisieren, so muß es möglich sein, die Kanten seines Graphen ohne Überschneidungen zu zeichnen. Der Graph muß daher planar sein. Wann ist nun ein Graph planar?

Betrachten wir den linken oberen Teil von Abbildung 4.9. Dies ist ein Graph mit 6 Knoten und neun Kanten. In dem unteren Bildteil ist der Versuch demonstriert, diesen Graphen ohne Kantenüberschneidungen (planar) zu zeichnen. Kann man diesen Graphen planar zeichnen, so kann man ihn auch planar zeichnen, indem man die Kanten beliebig anordnet und dann Kante für Kante in der gewählten Anordnung einträgt. In unserem Beispiel

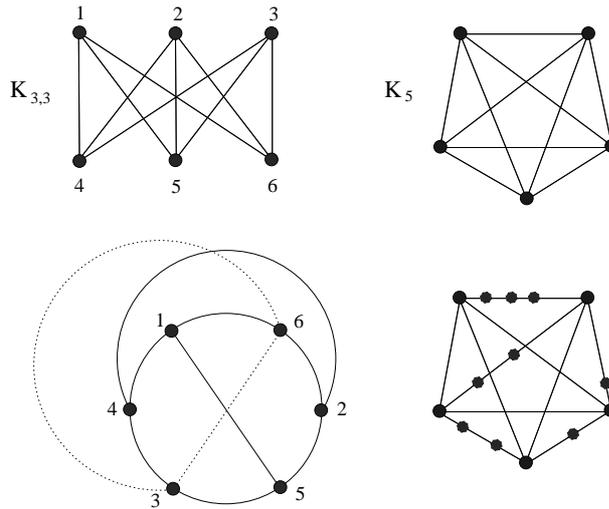


Abbildung 4.9.

kann man daher zuerst den Kreis $[1, 6], [6, 2], [2, 5], [5, 3], [3, 4], [4, 1]$ und dann die Kanten $[1, 5], [2, 4], [3, 6]$ zeichnen. Die beiden Kanten $[1, 5]$ und $[2, 4]$ kann man nicht auf derselben Seite des Kreises überschneidungsfrei zeichnen. Also muß eine innen und eine außen gezeichnet werden. Die letzte Kante $[6, 3]$ muß dann ebenfalls entweder innen oder außen gezeichnet werden. Wie man an der Figur erkennt, schneidet sie aber stets eine der beiden vorher gezeichneten Kanten.

Der erste Graph von Abbildung 4.9, der allgemein mit dem Namen $K_{3,3}$ bezeichnet wird, kann also nicht planar gezeichnet werden. Mit Hilfe einer analogen Betrachtung finden wir, daß auch der Graph daneben, der den Namen K_5 trägt, nicht planar gezeichnet werden kann. Der Buchstabe " K " bedeutet einen *vollständigen Graphen*, der Index bei K gibt an, wieviele Knoten er hat. K_5 ist also ein Graph mit fünf Knoten, in dem alle möglichen Kanten zwischen verschiedenen Knoten vorkommen. $K_{3,3}$ ist ein sogenannter *vollständiger bipartiter Graph*, in dem die Knoten in zwei Klassen von je drei Knoten unterteilt wurden, wobei Kanten nur zwischen Knoten verschiedener Klassen vorhanden sind, von denen aber alle.

$K_{3,3}$ und K_5 sind also nicht-planare Graphen. Wenn ein Graph planar ist, dann darf er natürlich keinen dieser beiden Graphen als einen Teil (wir sagen auch als Untergraph) enthalten. Ferner beobachten wir, daß wenn z.B. K_5 nicht überschneidungsfrei gezeichnet werden kann, dies natürlich

auch für den Graph in Abbildung 4.9 rechts unten gilt, der aus dem K_5 entsteht, wenn man in dessen Kanten neue Knoten hinzufügt. Ein Graph G' , der aus einem Graph G durch Einfügen von weiteren Knoten in seinen Kanten entsteht, wodurch diese Kanten unterteilt werden, heißt eine *Unterteilung* von G . Eine notwendige Bedingung dafür, daß ein Graph planar ist, ist also, daß er keine Unterteilung eines der Graphen $K_{3,3}$ oder K_5 enthält. Diese Bedingung ist aber auch hinreichend, was man nicht mehr so einfach zeigen kann. Zusammenfassend erhalten wir so das sogenannte *Theorem von Kuratowski*⁴ [Kuratowski, 1930]:

Ein Graph G ist genau dann planar, wenn er keine Unterteilung eines der Graphen $K_{3,3}$ oder K_5 enthält.

Man betrachte Abbildung 4.10. Es zeigt einen kleinen Teil einer Verdrahtungsebene des Speichers einer Transputeranlage, die real existiert und als Rechner eingesetzt ist. Man erkennt, daß alle Verbindungen von Knoten überschneidungsfrei gezeichnet sind. In manchen Knoten endet keine Kante, die Verbindungen zu diesen Knoten werden in einer anderen Ebene, die von der hier gezeigten räumlich getrennt und elektrisch isoliert ist, geführt. Insgesamt enthält der Speicher vier verschiedene Verdrahtungsebenen, in denen alle notwendigen elektrischen Leitungen überschneidungsfrei geführt werden können. Der Graph $G = (V, E)$, um den es sich hier handelt, ist nicht planar. Aber man kann seine Kantenmenge so in vier Teile E_1, E_2, E_3, E_4 zerlegen, daß jeder der Graphen $G_i = (V, E_i)$, $i = 1, 2, 3, 4$, planar ist. Wir sagen daher, man kann diesen Graphen in vier Ebenen überschneidungsfrei zeichnen.

Wenn man genügend Ebenen zu Hilfe nimmt, kann man jeden Graph überschneidungsfrei zeichnen. Man ordne dazu die Knoten längs einer beliebigen Geraden \mathcal{G} im Raum an und verwende z.B. für jede Kante eine eigene Ebene durch \mathcal{G} , in der man die Kante dann ungehindert von den übrigen Kanten zeichnen kann. Aus technischen und ökonomischen Gründen möchte man bei der Realisierung eines elektrischen Netzes aber mit möglichst wenig Ebenen auskommen, oder wenigstens mit einer vorgegebenen Anzahl k von Ebenen, die in der Regel viel kleiner als die Zahl der Kanten ist. Anstelle des Problems *Kann man G in einer Ebene zeichnen?* liegt dann also das etwas allgemeinere Problem *Kann man G in k Ebenen zeichnen?* vor, wobei k eine gegebene natürliche Zahl ist. Man ist versucht, die zweite Fragestellung als "leichter" zu bewerten? Das ist aber ein Irrtum. Gerade das Gegenteil ist der Fall, und zwar in dem folgenden Sinn. Wir kennen

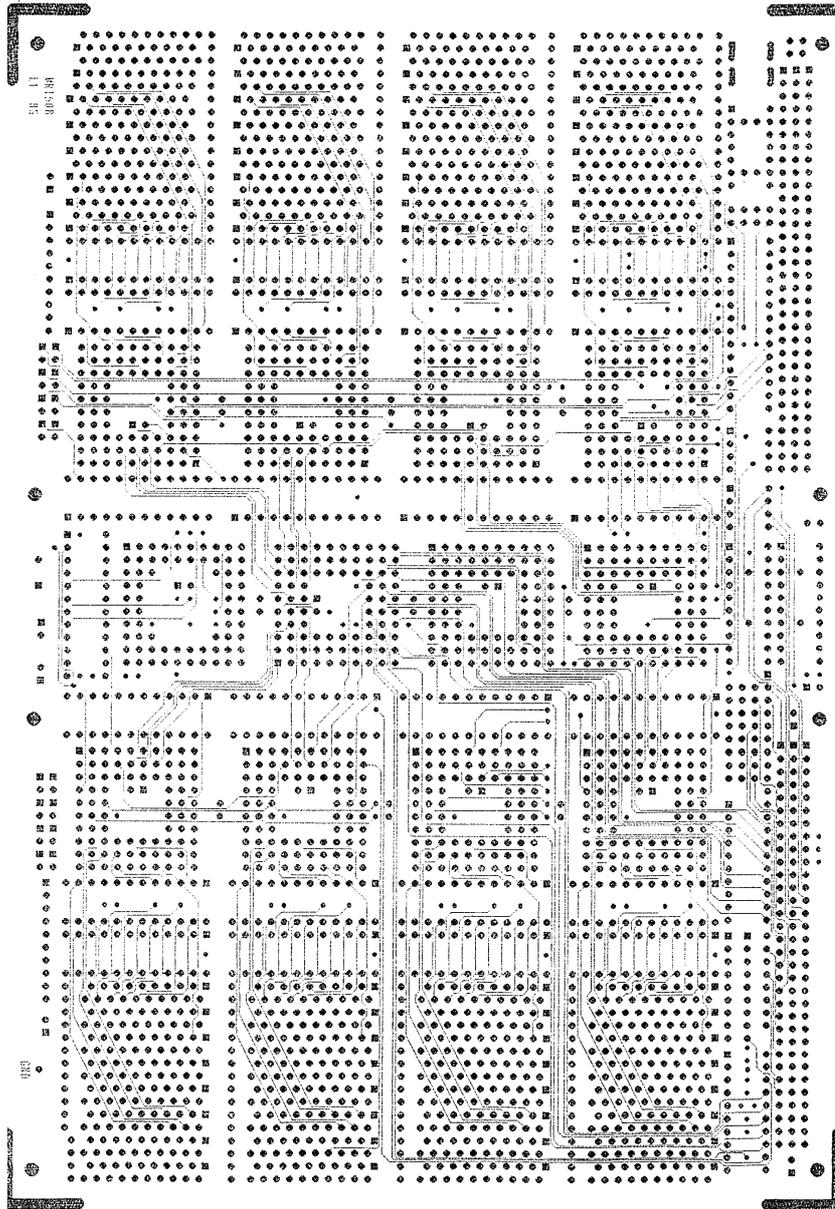


Abbildung 4.10.

heute Rechenmethoden, die es (unter Verwendung eines Computers) gestatten, sehr schnell zu entscheiden, ob ein Graph planar ist oder nicht. Aber wir kennen bisher keine Methode, die es bei allen Graphen G und bei allen vorgegebenen Zahlen k erlauben würde, ebenso schnell die zweite Frage zu beantworten, wenn die erste Frage mit "Nein" beantwortet wurde. Für die Beantwortung der ersten Frage existiert ein Algorithmus, der für einen Graph mit n Knoten die Antwort in einer Zeit liefert, die proportional zur Knotenanzahl n ist, während wir bisher für die Beantwortung der zweiten Frage nur Algorithmen kennen, deren Zeitbedarf exponentiell mit der Knotenzahl n anwächst. (Der erste Problemtyp ist daher *polynomial*, der zweite ist NP-vollständig. Es handelt sich hierbei um Begriffe aus der sogenannten Komplexitätstheorie. Eine Einführung in die Komplexitätstheorie gibt das Buch [Garey und Johnson, 1979].)

Bei einer weiteren Möglichkeit, ein nicht-planares elektrisches Netz zu realisieren, erfolgt die Verdrahtung doch in einer einzigen Ebene, indem man, wenn Überschneidungen notwendig sind, Brücken verwendet (das heißt eine der Leitungen wird mit Hilfe einer Brücke von der unten liegenden isoliert). Aus technischen und organisatorischen Gründen soll aber die Anzahl der zu verwendenden Brücken gering sein, soll also eine Höchstanzahl k nicht übertreffen. Damit treffen wir auf das Problem *Kann man G mit höchstens k Überschneidungen in einer Ebene zeichnen?*. Auch das ist ein NP-vollständiger Problemtyp.

Die minimale Anzahl von Überschneidungen, die notwendig ist, damit man G in einer Ebene zeichnen kann, heißt die *Überschneidungszahl* von G (*Crossing number*). Ihre Bestimmung ist ein schwieriges graphentheoretisches Problem.

Zusammenhang und Zuverlässigkeit

Graphen dienen auch als Modelle für Netze, in denen statt elektrischen Stroms eine andere Substanz transportiert wird. Das kann ein Straßennetz sein, in dem der Verkehr *fließt*, oder ein Leitungssystem zur zentralen Versorgung von Verbrauchern mit Gas oder Heizöl, oder ein Kommunikationsnetz, in dem Information (über ein geeignetes Medium) transportiert wird.

In Abbildung 4.11 (Farbabbildungen: S. 139) ist ein solches Kommunikationsnetz dargestellt, das real existiert, oder besser gesagt, in dieser Form real existiert hat. Heute ist dieses Netz bereits weit größer. Es handelt sich um ein Computernetz, mit dessen Hilfe weltweit Informationen ausgetauscht

werden können. Die Urform dieses Netzes bestand aus vier Knoten und wurde 1966 vom Verteidigungsministerium der USA in Auftrag gegeben. Abbildung 4.11 (Farbabbildungen: S. 139) zeigt die Ausbaustufe von etwa 1979. Im Jahre 1989 erfaßte dieses Netz, ursprünglich ARPA genannt, an die 30.000 Knoten. Heute ist es allgemein unter dem Namen INTERNET bekannt und erfaßt hunderttausende Knoten verteilt über die gesamte Erdkugel.

Im Zusammenhang mit der Planung, Errichtung und Aufrechterhaltung von solchen Netzen treten zahlreiche Fragen auf, die mit Hilfe von graphentheoretischen Methoden zu bearbeiten sind. Die Leistung eines Netzes wird gemessen an seiner Kapazität. Wieviel kann man gleichzeitig von Punkt A nach Punkt B transportieren? Wenn alle Leitungen (Kanten) gleiche Kapazität besitzen und alle Rechner (Knoten) dieselbe Leistungsstärke haben, so hängt die Antwort dieser Frage nur von der Struktur des Graphen ab, die es zu erfassen gilt. Eine andere Frage ist: Wie zuverlässig arbeitet ein Netz? Netzteile können ausfallen, dadurch können zwischen gewissen Knotenpaaren Verbindungen unmöglich werden. Unter der *Zuverlässigkeit* eines Netzes versteht man die Wahrscheinlichkeit dafür, daß alle Knotenpaare des Netzes durch einen Weg verbunden sind. Da ein Netz aus vielen einzelnen Bauteilen (Rechnern, Leitungen) besteht, ist seine Zuverlässigkeit meist eine sehr komplizierte Funktion der Zuverlässigkeiten seiner Bauteile. Unter der Zuverlässigkeit eines Bauteiles versteht man die Wahrscheinlichkeit dafür, daß dieses Teil intakt ist.

Zuverlässigkeitsfragen bei Netzen sind eng mit Zusammenhangsfragen in dem zugehörigen Modellgraph verknüpft. Die Wahrscheinlichkeit für die Intaktheit eines gewissen Bauteiles äußert sich dann als Wahrscheinlichkeit für das Vorhandensein eines gewissen Knotens oder einer gewissen Kante. Der Ausfall einer Kante kann einen Weg zwischen gewissen Knoten a und b zerstören. Die Verbindung im Graphen bleibt erhalten, wenn es einen weiteren Weg gibt, der die entnommene Kante nicht benutzt. Genauso kann der Ausfall von Knoten Wege zwischen a und b zerstören.

Es sei nun G ein beliebiger zusammenhängender Graph und a und b seien zwei seiner Knoten. Dann enthält G mindestens einen Weg zwischen a und b . Wir fragen nun nach der minimalen Anzahl von Knoten, die man aus G (samt den in ihnen anliegenden Kanten) entfernen muß, damit in dem verbleibenden Graph keine Verbindung zwischen a und b mehr möglich ist. Diese Anzahl nennt man die *Trennungszahl* $\tau_G(a, b)$ von G für das Knotenpaar a, b . Das Minimum dieser Zahlen, genommen über alle Knotenpaare

von G , die nicht direkt durch eine Kante verbunden sind, nennt man die *Trennungszahl* τ_G von G . Diese Größe ist ein strukturbezogenes Zuverlässigkeitsmerkmal. Intuitiv würden wir vermuten, daß der Graph (das Netz) umso zuverlässiger ist, je mehr man zerstören muß, um zwischen zwei Knoten alle Wege zu unterbinden. Der Zusammenhang mit der vorhin als Wahrscheinlichkeit eingeführten Zuverlässigkeit ist jedoch komplizierter. Eine Einführung in diesen Problembereich findet man in [Colbourn, 1984].

Zwei Wege zwischen denselben Knoten a und b in einem Graph G nennt man *kreuzungsfrei*, wenn sie außer a und b keine weiteren Knoten gemeinsam haben. In Abbildung 4.12 (Farbabbildungen: S. 140) ist links ein kleiner Graph und daneben zwei verschiedene Auswahlen von je zwei kreuzungsfreien Wegen zwischen a und b zu sehen. Die maximale Anzahl von kreuzungsfreien Wegen zwischen a und b heißt die *Zusammenhangszahl* $\mu_G(a, b)$ von G für das Knotenpaar a, b . Die Minimum dieser Zusammenhangszahlen, genommen über alle nicht durch eine Kante verbundenen Knotenpaare von G , heißt die *Zusammenhangszahl* μ_G von G .

Entnimmt man in Abbildung 4.12 (Farbabbildungen: S. 140) die durch ein Kästchen markierten Knoten, so zerfällt der Graph in zwei Teile. Die Knoten a und b sind nicht mehr verbunden.

Es ist klar, daß man in jedem Graph G , um jede Verbindung zwischen zwei Knoten a und b zu zerstören, auf jedem Weg einer Auswahl von $\mu_G(a, b)$ kreuzungsfreien Wegen mindestens einen Knoten entnehmen muß. Damit gilt $\mu_G(a, b) \leq \tau_G(a, b)$, und daher auch $\mu_G \leq \tau_G$. Nicht so leicht einzusehen ist, daß auch die umgekehrten Ungleichungen gelten und somit der folgende, für die Behandlung von Zusammenhangsfragen grundlegende Satz gilt:

Menger'scher⁵ Graphensatz [Menger, 1927]: *In jedem Graph G gilt für jedes Paar von Knoten a und b , die nicht durch eine Kante verbunden sind, daß $\mu_G(a, b) = \tau_G(a, b)$.*

Untersucht man die Zerstörbarkeit von Verbindungen in einem Graph durch Entfernen von Kanten statt durch Entfernen von Knoten, so kommt man zu ähnlichen Begriffen und Sachverhalten.

Rundreisen

Der Pförtner einer Industrieanlage soll täglich nach Arbeitsschluß einen Rundgang durch die Anlage machen und noch unversperrte Tore absperren. Der Graph in Abbildung 4.13a (Farbabbildungen: S. 140) beschreibt

einen schematischen Plan der Anlage. Der Pförtner befindet sich tagsüber an der Pforte, die durch den großen Knoten A ganz unten dargestellt ist. Die übrigen Knoten stellen die zu überprüfenden Tore dar. Jede Kante repräsentiert einen der im Gelände möglichen Wege zwischen den Toren selbst oder zwischen den Toren und der Pforte und ist mit einer Zahl markiert, die angibt, wieviel Minuten man für seine Zurücklegung benötigt. Der Pförtner kann die Anlage nur über die Pforte verlassen. Er muß also vom Punkt A aus den Rundgang beginnen und dort wieder enden. Der Arbeitgeber veranschlagt für die Durchführung des Rundganges eine halbe Stunde. Ist diese Zeit ausreichend?

Zur Unterstützung der Leser bei der Beantwortung dieser Frage ist in Abbildung 4.13b (Farbabbildungen: S. 140) ein möglicher Rundgang durch grüne Linien angedeutet. Für seine Durchführung benötigt man laut Angaben an den Kanten 36 Minuten. Bedenkt man, daß die Überprüfung und eventuelle Absperrung eines Tores zusätzliche Zeit erfordert, so ergibt sich die Folgerung, daß ein halber Stundenlohn für diese Aufgabe nicht ausreicht. Aber ist der vorgeschlagene Rundgang auch der kürzeste?

Was im letzten Teilabschnitt angesprochen wurde, ist ein sogenanntes *Rundreiseproblem*. Zu einem Rundreiseproblem gehört ein zusammenhängender Graph mit einer Kantenbewertung. Diese ordnet jeder Kante eine positive Zahl als Wert (Länge, Zeitdauer, Fahrtkosten) zu. Gesucht ist dann eine Rundreise durch den Graphen, das heißt ein Kreis, der alle Knoten genau einmal enthält. Einen solchen Kreis nennt man auch *Hamilton'schen Kreis*.⁶

Aus den Werten der zur Bildung des Kreises benutzten Kanten erhält man durch Aufsummieren den Gesamtwert der Rundreise. Wie im obigen Beispiel möchte man meist die Rundreise so wählen, daß dieser Gesamtwert möglichst klein ausfällt.

Der Problemtyp *Rundreiseproblem* ist einer der Grundtypen von Problemen aus der sogenannten *Kombinatorischen Optimierung*. Dieser Typ hat zahlreiche für die Anwendung wichtige Varianten. Als Anwendungen treten hauptsächlich Probleme bei industriellen Fertigungsprozessen auf (zum Beispiel günstige Führung eines Roboterarms bei Bearbeitung eines Werkstückes). Eine Übersicht über diesen Problemkreis findet man in [Lawler et al., 1985].

Nicht jeder Graph besitzt einen Hamilton'schen Kreis. Es gibt zahlreiche notwendige und hinreichende Bedingungen für die Existenz eines solchen Kreises. Zum Beispiel muß $\mu_G \geq 2$ gelten, denn wenn alle Knoten auf einem

Kreis liegen, dann gibt es zu je zwei nicht durch eine Kante verbundenen Knoten mindestens zwei kreuzungsfreie verbindende Wege, nämlich die beiden Kreisbögen, in die der Hamilton'sche Kreis durch die beiden Knoten zerlegt wird. Eine hinreichende Bedingung ist z.B., daß jeder Knoten v einen Grad $d(v) \geq \frac{n}{2}$ (n die Knotenzahl) besitzt. Jedoch ist bisher keine notwendige und hinreichende Bedingung für die Existenz eines Hamilton'schen Kreises bekannt, die sich mit geringem Zeitaufwand, das heißt einem Zeitaufwand, der mit der Knotenzahl nur polynomial anwächst, überprüfen ließe. Die Existenz eines derartigen Algorithmus hätte weitreichende Konsequenzen für die sogenannte *Komplexitätstheorie*, einem Teilgebiet der theoretischen Informatik. Denn der Problemtyp *Finde in G einen Hamilton'schen Kreis oder beweise, daß es keinen gibt* gehört zu den Grundproblemen dieser Theorie, mit denen man die "Schwierigkeit etwas zu berechnen" zu erklären versucht. Siehe dazu auch [Garey und Johnson, 1979].

Als beste Lösung für das Problem, alle Knoten eines Graphen mindestens einmal zu besuchen, muß nicht ein Hamilton'scher Kreis erscheinen (ein solcher muß ja nicht einmal existieren). Die Gestalt der Lösung hängt nicht nur von der Struktur des Graphen sondern auch von der Kantenbewertung ab. In Abbildung 4.14a (Farbabbildungen: S. 141) ist nochmals der gleiche Graph wie in Abbildung 4.13a (Farbabbildungen: S. 140), aber diesmal mit einer anderen Kantenbewertung angegeben. Wir finden einen optimalen Rundgang, der aus zwei Kreisen durch den Knoten A besteht. Dieser Rundgang ist in Abbildung 4.14b (Farbabbildungen: S. 141) durch die grünen Linien angedeutet.

Färbungen

In Abbildung 4.15a (Farbabbildungen: S. 141) ist eine Figur aus mehreren Kästchen auf einem quadratischen Hintergrund gezeigt. Wir wollen versuchen, den Hintergrund und die Kästchen so mit Farben auszufüllen, daß Kästchen mit gemeinsamen Rand verschiedene Farben erhalten. In Abbildung 4.15c (Farbabbildungen: S. 141) ist so eine Färbung gegeben. Dabei wurden fünf Farben verwendet.

Mit dem kleinen Beispiel von Abbildung 4.15a (Farbabbildungen: S. 141) wird ein Problem angesprochen, das sehr alt ist und viele Mathematiker beschäftigte, nämlich das sogenannte *Vierfarbenproblem*. Man stelle sich eine Landkarte eines Landes vor, in dem die politischen Einheiten (z.B. die Landkreise oder Verwaltungsbezirke) durch verschiedene Farbgebung kennt-

lich gemacht werden. Damit eine wirklich eindeutige und schnelle Unterscheidung durch das Auge möglich ist, sollen Bezirke, die eine gemeinsame Grenze haben, verschieden gefärbt sein. Die Frage ist nun: *Kann man bei dieser Aufgabe immer mit vier Farben auskommen?* Die Antwort ist: *Ja!* Nach vielen erfolglosen Versuchen bedeutender Mathematiker ist diese positive Antwort erst 1976 von K. Appel und W. Haken gefunden und in [Appel und Haken, 1976] veröffentlicht worden. Jedoch ist bis heute niemand in der Lage, diesen Beweis mit Bleistift und Papier auf seine Richtigkeit zu überprüfen. Appel und Haken verwendeten einen Computer. Es galt tausende von Spezialfällen zu überprüfen, dies konnte nicht von Menschenhand durchgeführt werden. Aus diesem Grund galt dieser "Beweis" des sogenannten *Vierfarbentheorems* als unbefriedigend und die Suche nach einem "echten" mathematischen Beweis für dieses Theorem ging unvermindert weiter. Heute kennt man eine Version des Beweises, die zwar immer noch schwierig, aber doch bedeutend einfacher ist als die ursprüngliche Version.

Dem Vierfarbentheorem kann eine graphentheoretische Formulierung gegeben werden. Man betrachte dazu Abbildung 4.15b (Farbabbildungen: S. 141). Hier wurde jedem Kästchen ein Knoten zugeordnet, ein weiterer Knoten (rechts unten) vertritt den Rahmen. Knoten werden verbunden, wenn die Bereiche, für die sie stehen, nicht gleich gefärbt werden dürfen. So entsteht der in Abbildung 4.15b gezeichnete Graph G (einige Kanten von Randknoten zu dem Knoten rechts unten sind nur angedeutet). Das Problem kann nun so formuliert werden:

Man färbe die Knoten von G so, daß durch Kanten verbundene Knoten stets verschiedene Farben erhalten. Eine Färbung mit dieser Eigenschaft nennen wir zulässig.

Der Graph in Abbildung 4.15b (Farbabbildungen: S. 141) ist planar. Im allgemeineren Fall einer beliebigen Landkarte mit unterschiedlichen politischen Regionen, wähle man für jede Region einen Knoten und verbinde Knoten, die Regionen mit gemeinsamer Grenze darstellen. Auf diese Weise entsteht immer ein planarer Graph. Das Vierfarbentheorem lautet nun:

Man kann jeden planaren Graphen mit vier Farben zulässig färben.

Eine Darstellung des Vierfarbentheorems und seines Beweises findet man im Internet unter der Adresse <http://www.math.gatech.edu/~thomas/FC/fourcolor.html>.

Aufgabe für den Leser: Finde eine zulässige Färbung in Abbildung 4.15a (Farbabbildungen: S. 141) mit nur vier Farben.

Der Begriff einer zulässigen Färbung ist nicht auf planare Graphen beschränkt. Wir können nach zulässigen Färbungen beliebiger Graphen mit möglichst wenig Farben suchen. Das ist das allgemeine *Färbungsproblem* für Graphen.

Zulässige Graphenfärbungen dienen zur Behandlung von *Konfliktsituationen*. Nehmen wir an, jeder Knoten eines Graphen G bedeute eine Aufgabe aus einer Menge von Aufgaben, die mit Hilfe eines gegebenen Vorrats von Betriebsmitteln durchzuführen sind. Einige dieser Aufgaben können gleichzeitig durchgeführt werden, andere nicht. Man denke z.B. an eine Menge von Prüfungen, die am Ende einer Schulungsperiode abgehalten werden sollen. Eine Prüfungseinheit, das heißt eine Aufgabe, besteht aus dem Prüfer, dem Prüfling, der Angabe des Ortes, an dem die Prüfung stattzufinden hat, der Hilfsmittel (Maschinen, Mikroskope, Filmkamera, etc.), die dabei mitwirken. Das Problem besteht in der Abfassung eines Zeitplans, der gestattet, alle vorzunehmenden Prüfungen an möglichst wenig Prüfungsterminen abzuhalten. Prüfungen können nur dann zum gleichen Termin erfolgen, wenn die beteiligten Prüfer und die beteiligten Prüflinge verschieden sind, wenn sie nicht am selben Ort stattfinden und wenn sie nicht dieselben Hilfsmittel erfordern. Es sei G der zugehörige *Konfliktgraph*, das heißt ein Graph, dessen Knoten die Prüfungseinheiten repräsentieren und der Kanten zwischen genau den Prüfungseinheiten hat, die nicht zum gleichen Termin geplant werden können. Dann benötigt man offensichtlich genau so viele verschiedene Prüfungstermine, wie man Farben in einer zulässigen Färbung von G mit möglichst wenig Farben benötigt. Diese minimale Anzahl von Farben heißt die *chromatische Zahl* χ_G des Graphen G .

In Abbildung 4.16 (Farbabbildungen: S. 142) sind zwei Beispiele für einen Konfliktgraphen angegeben. Der erste Graph besteht aus einem Kreis mit fünf Knoten, und einem sechsten Knoten, der mit allen Kreisknoten verbunden ist. Der zweite Graph ist nur ein Kreis mit fünf Knoten (üblicherweise durch C_5 bezeichnet). Man überlege sich: Der erste Graph hat chromatische Zahl 4, der zweite hat chromatische Zahl 3.

Eine Sonderstellung nehmen die Graphen ein, deren chromatische Zahl gleich 2 ist. Deren Knotenmengen zerfallen in zwei Klassen (die blauen und die roten Knoten), und Kanten verlaufen nur zwischen Knoten verschiedener Klassen. Man nennt solche Graphen auch *bipartite Graphen*. Bäume sind bipartit. Man kann zeigen: *Ein Graph ist genau dann bipartit, wenn er*

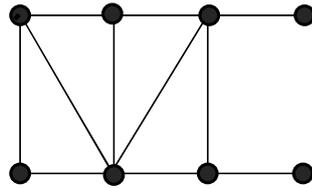


Abbildung 4.18.

keine Kreise mit ungerader Knotenanzahl enthält. (Zur Färbung solcher Kreise benötigt man drei Farben.)

Perfekte Graphen

Es ist leicht festzustellen, ob ein Graph mit zwei Farben gefärbt werden kann, ob er also bipartit ist. Dafür gibt es Algorithmen mit nur polynomialem Zeitaufwand. Dagegen ist es bereits schwer festzustellen, ob ein Graph mit drei Farben zulässig färbbar ist. Für dieses Problem ist bis jetzt kein Algorithmus mit nur polynomialem Zeitbedarf bekannt. Solche Algorithmen sind auch unbekannt zur Lösung von Problemen des Typs *Ist der Graph G mit k Farben zulässig färbbar?*, und zwar für jedes $k \geq 3$.

Ein Teil eines Graphen G , der aus Knoten besteht, die paarweise miteinander durch eine Kante verbunden sind, heißt eine *Clique* von G . Die größte Knotenzahl in einer Clique von G heißt die *Cliquezahl* von G , sie wird meist durch ω_G bezeichnet. Die Knoten einer Clique erhalten in jeder zulässigen Färbung paarweise unterschiedliche Farben. Daher benötigt jede zulässige Färbung mindestens ω_G Farben, das heißt es gilt für jeden Graph G , daß $\chi_G \geq \omega_G$.

In Abbildung 4.16 (Farbabbildungen: S 142) sehen wir zwei Graphen, deren Färbungszahlen jeweils um 1 größer sind als die Cliquezahl. Die größten Cliques des ersten Graphen sind seine Dreiecke. Er hat also Cliquezahl 3. Seine chromatische Zahl ist 4. Der zweite Graph hat nicht einmal ein Dreieck, seine Cliquezahl ist daher 2, aber seine chromatische Zahl ist 3. In Abbildung 4.17 (Farbabbildungen: S 142) sehen wir ein Beispiel für einen Graph ohne Dreiecke, dessen chromatische Zahl 4 ist. Heute weiß man: *Zu jeder natürlichen Zahl k gibt es dreiecksfreie Graphen (das heißt Graphen mit Cliquezahl 2), deren chromatische Zahl k ist.* Die Lücke zwischen χ_G und ω_G kann daher beliebig groß werden.

Ein Graph G heißt *perfekt*, wenn folgendes gilt: Es ist $\omega_G = \chi_G$, und ferner gilt $\omega_{G'} = \chi_{G'}$ für jeden Graph G' , den man aus G erhält indem man einen Teil seiner Knoten samt den daran anliegenden Kanten entfernt.

Da ein bipartiter Graph keine Dreiecke enthält, ist seine Cliquezahl 2, und damit gleich seiner chromatischen Zahl. Bipartite Graphen, und damit insbesondere Bäume, sind also Beispiele für perfekte Graphen. Die Graphen aus Abbildungen 4.16 (Farbabbildungen: S. 142) und 4.17 (Farbabbildungen: S. 142) sind nicht perfekt.

Perfektheit ist eine interessante Grapheneigenschaft. Für zahlreiche Typen von perfekten Graphen sind viele graphentheoretische Aufgaben (z.B. optimale zulässige Färbung) einfach, das heißt mit einem schnellen Algorithmus durchführbar. Einer der prominentesten Typen von perfekten Graphen ist der Typ der chordalen Graphen. Ein Graph G heißt *chordal*, wenn er keinen sehenlosen Kreis mit mindestens vier Knoten enthält. Zum Beispiel ist der Graph in Abbildung 4.18 chordal. Der rechte Graph in Abbildung 4.16 (Farbabbildungen: S. 142) ist selbst ein sehenloser Kreis, nämlich der C_5 , und somit nicht chordal. Aber auch der linke Graph in diesem Bild ist nicht chordal. Entfernt man den mittleren Knoten, so bleibt der Kreis C_5 übrig.

Die Erkennung allgemeiner perfekter Graphen, das heißt solcher, von denen man nicht eine weitere besondere Eigenschaft kennt, welche die Perfektheit nach sich zieht, ist schwierig. Man kennt bis heute keinen schnellen Algorithmus für diese Aufgabe. Manchmal kann man Perfektheit dadurch erkennen, daß man den komplementären Graphen betrachtet. Es sei $G = (V, E)$ ein beliebiger Graph. Der zu G *komplementäre Graph* \bar{G} , auch das *Komplement* zu G genannt, ist ein Graph mit derselben Knotenmenge V wie G , der eine Kante zwischen zwei verschiedenen Knoten genau dann enthält, wenn G dort keine hat. Formal: $\bar{G} = (V, \bar{E})$ mit $\bar{E} = \{[x, y] : x \neq y \wedge [x, y] \notin E\}$. Es gilt der folgende Satz, der auch als *Perfect Graph Theorem* bezeichnet wird:

Ein Graph G ist genau dann perfekt, wenn es sein Komplement \bar{G} ist.

Zahlreiche wichtige Graphentypen sind dadurch definiert, daß man die Existenz speziell geformter Teile (Untergraphen) ausschließt. Z.B. sind Bäume durch das Fehlen von Kreisen, bipartite Graphen durch das Fehlen von Kreisen ungerader Knotenzahl, planare Graphen durch das Fehlen von Unterteilungen des $K_{3,3}$ und des K_5 , chordale Graphen durch das Fehlen von

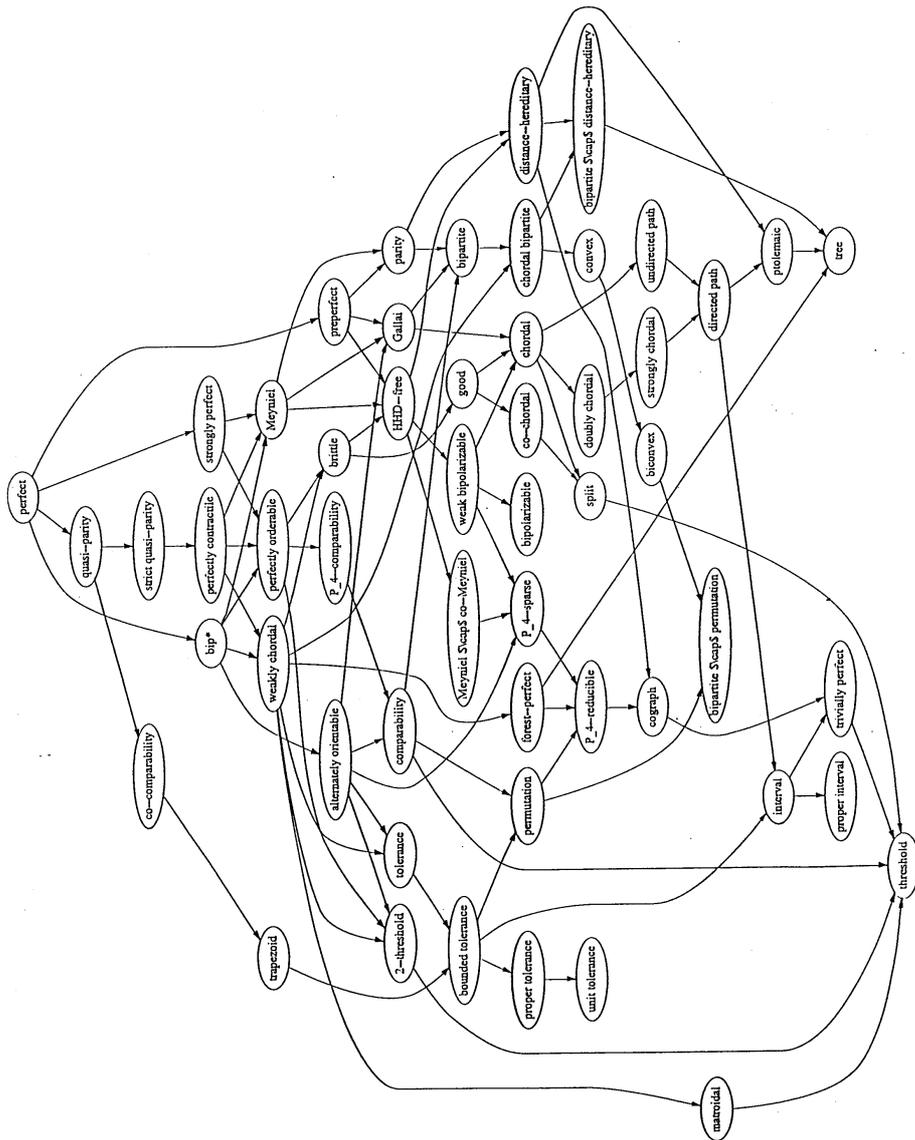


Abbildung 4.19.

sehenlosen Kreisen einer Knotenzahl von mindestens 4 charakterisiert. Eine derartige Charakterisierung nennt man *Charakterisierung durch verbotene Untergraphen*. Diese Art der Charakterisierung ist aber nur dann möglich, wenn die zu beschreibende Eigenschaft nicht nur der Graph G selbst, sondern auch jeder Untergraph G' besitzt, der durch Entfernen von Knoten und den an diesen anliegenden Kanten entsteht. Denn ein in G nicht enthaltener Untergraph kann ja dann auch nicht in G' enthalten sein. Durch Entfernen von Kanten allein kann man allerdings möglicherweise eine durch verbotene Untergraphen bestimmte Eigenschaft zerstören. Ein Kreis mit Sehnen z. B. kann durch Entfernen von Kanten zu einem sehenlosen Kreis werden. Ein perfekter Graph ist so definiert, daß auch alle seine wie oben gebildeten Untergraphen perfekt sind. Es stellt sich daher die Frage, ob man perfekte Graphen durch verbotene Untergraphen charakterisieren kann. Sicher darf ein perfekter Graph keine sehenlosen Kreise mit ungerader Knotenzahl enthalten, denn diese sind nicht perfekt. Es wird heute vermutet, daß es genügt, wenn man diese Kreise und deren Komplemente ausschließt. Dies ist der Inhalt der viel diskutierten sogenannten *Strong Perfect Graph Conjecture*:

Ein Graph G ist genau dann perfekt, wenn er keine sehenlosen Kreise mit ungerader Knotenzahl und keine Komplemente solcher Kreise besitzt.

Der Begriff *perfekter Graph* wurde von C. Berge zu Beginn der 60-er Jahre eingeführt [Berge, 1961]. Berge vermutete bereits, daß Perfektheit durch Komplementbildung nicht verloren geht. Der Beweis für das Perfect Graph Theorem gelang L. Lovász 1972 in [Lovász, 1972]. Die Strong Perfect Graph Conjecture ist trotz vieler Anstrengungen von namhaften Mathematikern bisher unbewiesen. Eine gute Einführung in die Theorie und Praxis perfekter Graphen ist [Golumbic, 1980].

Es gibt eine Vielzahl von Grapheneigenschaften, welche die Perfektheit nach sich ziehen. Daher kennt man zahlreiche Klassen von perfekten Graphen. Diese Klassen sind in Abbildung 4.19 als die Knoten eines gerichteten Graphen dargestellt, die Knoten enthalten den Klassennamen. Die Abbildung ist mit freundlicher Genehmigung der Verfasser dem Buch [Brandstädt et al., 1999] entnommen. Eine gerichtete Kante (A,B) bedeutet, daß die Klasse B in der Klasse A enthalten ist. Das Bild demonstriert einprägsam die große Aktivität der Graphentheoretiker im Bereich der perfekten Graphen. Die meisten der hier angeführten Klassen wurden erst in den letzten beiden Jahrzehnten gefunden oder näher untersucht.

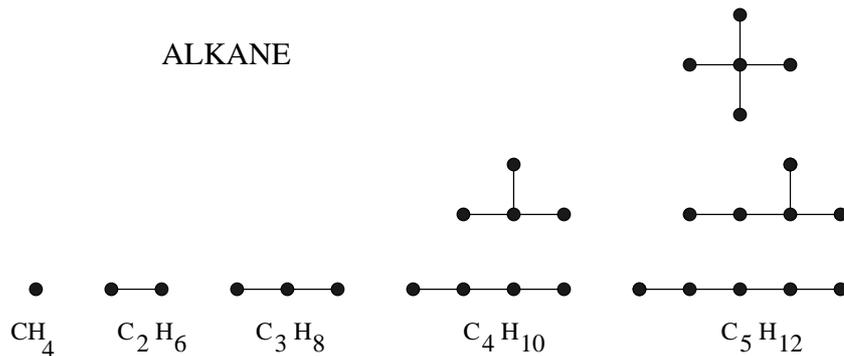


Abbildung 4.20.

Chemische Graphen

In der Chemie wird ein Molekül durch seine *Strukturformel* charakterisiert, die angibt, wieviele Atome von welcher Art am Bau des Moleküls beteiligt sind. So ist z.B. C_4H_{10} ein Molekül aus vier Kohlenstoff- und 10 Wasserstoffatomen. Auf Grund der verschiedenen Bindungsmöglichkeiten zwischen den Kohlenstoffatomen unterscheidet man verschiedene Molekülarten von Molekülen mit derselben Strukturformel (Isomere). Die Bindungsmöglichkeiten lassen sich durch einen ungerichteten Graph darstellen, den wir als *Strukturgraph* bezeichnen. Im Falle der Alkane (Kohlenwasserstoffe), das sind Moleküle mit der Strukturformel $\text{C}_n\text{H}_{2n+2}$, sind die Strukturgraphen Bäume, deren Knotengrade durch 4 beschränkt sind. In Abbildung 4.20 sind diese Bäume für $n \leq 5$ dargestellt.

Das erste Molekül dieser Art, das zwei verschiedene Isomere hat, ist demnach C_4H_{10} . C_5H_{12} hat bereits drei Isomere.

Bei komplizierteren Molekülen treten auch Kreise auf. Außerdem können Doppelbindungen auftreten, die wir dann durch parallele Kanten andeuten. Schließlich können noch mehrere (von H verschiedene) Atomarten am Bau des Moleküls beteiligt sein. Daher haben wir im Strukturgraphen die Knoten durch das zugehörige Atomsymbol zu markieren, um durch ihn die gesamte für die Charakterisierung des Moleküls benötigte Information darzustellen. In Abbildung 4.21 sieht man als Beispiel den Strukturgraph von Dioxin.

Für den Chemiker, und besonders für den Kristallographen, ist nicht jede Zeichnung eines Strukturgraphen gleichwertig. Die Eigenschaften eines

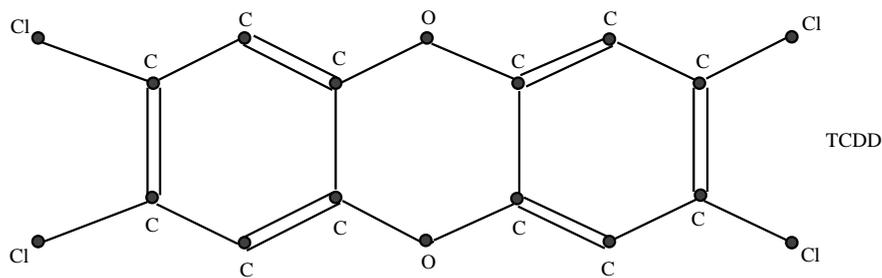


Abbildung 4.21.

Moleküls hängen auch von seiner Einbettung in den Raum ab. Auf Grund dieser Eigenschaften lassen sich gleichwertige Atome (eines Atomtyps) erkennen, die im Molekül wie im Molekülverband dieselbe Wirkung zeigen und daher im Molekül dieselbe Funktion ausüben. Diese Gleichwertigkeit zweier Atome A und A' erkennt man dadurch, daß die räumliche Verteilung der Atome im Molekül eine räumliche Symmetrie aufweist, und es daher eine Symmetrieoperation gibt, die A in A' überführt. Das kann eine Reflexion sein, eine Inversion, eine Rotation, oder eine Kombination dieser Operationen. In Abbildung 4.21 zum Beispiel kann man das Molekül um die Mittellinie als Achse um 180° drehen, wobei die Figur in sich übergeht. Auch eine Drehung um eine Achse durch die beiden Sauerstoffatome ist möglich. Wir stellen also fest, daß Graphen eine *Symmetrie* aufweisen können.

Die eben angesprochene Symmetrie von Graphen hat geometrischen Charakter. Sie ist von der räumlichen Darstellung des Graphen abhängig und verschwindet, wenn man diese Darstellung entsprechend ändert. Daneben gibt es sogenannte, bereits in der Einleitung erwähnte *kombinatorische* Symmetrien, die unabhängig von der Darstellung des Graphen sind. Eine kombinatorische Symmetrie wird auch als *Automorphismus* bezeichnet. Der Automorphiebegriff und die damit zusammenhängenden Sachverhalte sind wichtig bei der Kodierung von Graphen zum Zwecke der Katalogisierung (Datenbanken).

Ein *Automorphismus* eines Graphen $G = (V, E)$ ist eine Permutation (Umordnung) der Knotenmenge V , also eine bijektive Abbildung $\pi : V \rightarrow V$, mit der Eigenschaft, daß $[i, j] \in E$ genau dann gilt, wenn $[\pi(i), \pi(j)] \in E$. Wir sagen kurz: Ein Automorphismus von G ist eine Permutation der Kno-

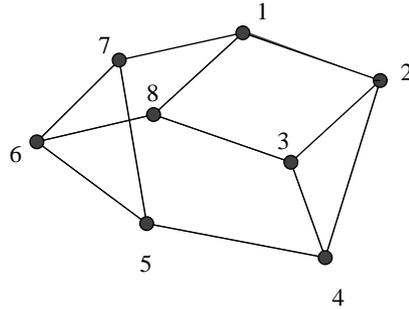


Abbildung 4.22.

ten, bei der die Kanten erhalten bleiben. Sind die Knoten oder die Kanten von G markiert, so verlangt man von einem Automorphismus zusätzlich, daß die Marken erhalten bleiben sollen. Man darf also nur Knoten mit gleichen Marken untereinander permutieren, und zwar nur so, daß dabei Kanten mit gleichen Marken ineinander übergehen. Im Falle von Strukturgraphen für Moleküle heißt das, daß der Atomtyp und die Bindungsart erhalten bleiben sollen.

In Abbildung 4.22 ist ein Graph gezeichnet, dessen Automorphismen wir noch leicht durch Probieren bestimmen können. Zum Beispiel ist die Permutation $\pi_1 : (1, 8)(2, 3)(4, 5)(6, 7)$ ein Automorphismus, das heißt vertauscht man 1 mit 8, 2 mit 3 und 6 mit 7 und läßt man 4 und 5 fest, so gehen Kanten in Kanten und Nicht-Kanten in Nicht-Kanten über. Ebenso sind $\pi_2 : (1)(2, 7)(3, 6)(4, 5)(8)$ und $\pi_3 : (1, 8)(2, 6)(3, 7)(4, 5)$ Automorphismen. $\pi_0 : (1)(2)(3)(4)(5)(6)(7)(8)$ wird auch als Automorphismus aufgefaßt, es ist der *triviale Automorphismus*. (Eine Zeichenfolge (i_1, i_2, \dots, i_k) in der Beschreibung einer Permutation π bedeutet, daß $\pi(i_1) = i_2, \pi(i_2) = i_3, \dots, \pi(i_{k-1}) = i_k, \pi(i_k) = i_1$).

Die Automorphismen eines Graphen G bilden mit der Hintereinanderausführung von Permutationen seiner Knotenmenge als Operation eine Gruppe von Permutationen, die *Automorphismengruppe* $Aut(G)$. In unserem Beispiel ist $Aut(G) = \{\pi_0, \pi_1, \pi_2, \pi_3\}$ mit der identischen Permutation π_0 als neutrales Element. Diese Gruppe ist isomorph zur sogenannten *Kleinschen Vierergruppe* (eine der zwei nicht-isomorphen Gruppen mit vier Elementen). Es gilt für die Hintereinanderausführung $\pi_i \pi_j(u) = \pi_i(\pi_j(u))$. Damit wird $\pi_i \pi_0 = \pi_0 \pi_i = \pi$ für $0 \leq i \leq 3$. Ferner ist $\pi_1 \pi_1 = \pi_2 \pi_2 = \pi_3 \pi_3 = \pi_0, \pi_1 \pi_2 = \pi_3, \text{ usw.}$

Die beiden Knoten 1 und 8 in Abbildung 4.22 sind (kombinatorisch) *äquivalent*, das soll heißen, daß es Automorphismen gibt, die einen Knoten in den anderen überführen, und umgekehrt. Alle untereinander äquivalenten Knoten bilden einen *Orbit* der Gruppe $Aut(G)$. Also ist z. B. $\{1, 8\}$ ein solcher *Orbit*. Die übrigen Orbits sind $\{4, 5\}$ und $\{2, 3, 6, 7\}$. Man beachte, daß die Knoten eines Orbits paarweise äquivalent sind. Jeder kann durch einen geeigneten Automorphismus in jeden anderen übergeführt werden. In Abbildung 4.23a (Farbabbildungen: S. 143) sind die drei verschiedenen Orbits durch verschiedene Farben verdeutlicht.

Es gibt auch äquivalente Kanten und äquivalente Nicht-Kanten. Zum Beispiel sind die beiden Kanten $[2, 3]$ und $[6, 7]$ äquivalent. Denn sie gehen bei dem Automorphismus π_2 ineinander über. $[1, 8]$ ist zu keiner anderen Kante äquivalent, ebenso $[4, 5]$ nicht. Die Nicht-Kante $[1, 3]$ ist äquivalent zur Nicht-Kante $[1, 6]$, aber auch zu $[7, 8]$. Um aber die "Bewegungen" der elementaren Bestandteile unserer Figur unter der Wirkung der Automorphismengruppe exakt erfassen zu können, sollten wir nicht die Äquivalenz von Kanten oder Nicht-Kanten, sondern die Äquivalenz von geordneten Knotenpaaren (x, y) erfassen. So ist z.B. $(1, 2)$ äquivalent zu den Paaren $(8, 3)$, $(1, 7)$ und $(8, 6)$, aber nicht zu $(2, 1)$, obwohl $[1, 2]$ eine Kante ist. Eine maximale Menge von paarweise äquivalenten geordneten Knotenpaaren, nennt man ein *Orbital* (oder einen *2-Orbit*) von $Aut(G)$. In unserem Beispiel ist $\{(1, 2), (8, 3), (1, 7), (8, 6)\}$ ein *Orbital*, $\{(2, 1), (3, 8), (7, 1), (6, 8)\}$ ist ein zweites. Insgesamt gibt es bei unserem Beispiel 15 verschiedene Orbitals. Wollte man alle diese in einer Zeichnung des Graphen wie in Abbildung 4.22 durch verschiedene Farbgebung unterscheiden, so benötigte man 15 verschiedene Farben für 56 verschiedene geordnete Knotenpaare. Das ergibt kein übersichtliches Bild. Daher werden in Abbildung 4.23b und 4.23c (Farbabbildungen: S. 143) nur einige der Orbitals in zwei verschiedenen Kopien des Graphen dargestellt.

Will man eine Übersicht über alle Orbits und Orbitals in einem Bild geben, so ist es besser, wenn man jedem geordneten Knotenpaar (i, j) eine Position in einem zweidimensionalen Schema von zeilen- und spaltenweise angeordneten Quadraten zuordnet. Das Quadrat in der i -ten Zeile und in der j -ten Spalte repräsentiert dann das geordnete Paar (i, j) . Paare der Form (i, i) liegen in der Diagonale und können dazu verwendet werden, die Knoten des Graphen darzustellen. Die Quadrate kann man dann unterschiedlich einfärben, um die Zugehörigkeit zu einem Orbit oder Orbital auszudrücken. Im Falle unseres Beispiels ergibt sich das Farbmuster in Ab-

bildung 4.24 (Farbabbildungen: S. 144).

Der Graph in Abbildung 4.21 hat vier Orbits und 82 Orbitals. Eine optische Darstellung würde also 84 verschiedene Farben benötigen.

Mit Hilfe der Bestimmung der Orbits und der Orbitals markierter Graphen ist es möglich, Kodierungsverfahren für Graphen zu konstruieren, die eine kompakte Darstellung solcher Graphen zum Zwecke der Katalogisierung erlauben, was besonders wichtig für die Arbeit mit Strukturgraphen von Molekülen ist. Bei der Implementierung solcher Kodierungsverfahren werden jedoch meist die Orbitals nicht explizit bestimmt, weil bereits eine nur implizit gegebene Information darüber ausreicht, um die Struktur der zu behandelnden Graphen eindeutig bis auf Isomorphie (also bis auf die Art der Darstellung) zu charakterisieren.

Bei gerichteten Graphen $G = (V, E)$ wird von einem Automorphismus π auch verlangt, daß er die Kantenrichtung erhält, daß also $(i, j) \in E$ genau dann, wenn $(\pi(i), \pi(j)) \in E$. Der Graph in Abbildung 4.25 (Farbabbildungen: S. 145) ist ein *zyklischer Graph* (auch einfach *Zykel* genannt) mit sieben Knoten. Seine Automorphismen sind gegeben durch die „Drehungen“ π_i , $i = 0, 1, \dots, 6$, (im Uhrzeigersinn), wobei $\pi_i(x)$ der Knoten ist, den man erreicht, wenn man längs des Zyklus um i Kanten weitergeht. Also ist z. B. $\pi_3(1) = 4$, $\pi_5(6) = 4$, usw. Man sieht also, daß jeder Automorphismus π definierbar ist, ohne daß der Zykel wie in Abbildung 4.25a so symmetrisch längs einer Kreislinie gezeichnet ist. Zu diesem Zykel, bzw. zu seiner Automorphismengruppe, gehört ein einziger Orbit (da man jeden Knoten in jeden anderen mit Hilfe einer „Drehung“ überführen kann) und 6 Orbitals, die in Abbildung 4.25b gezeigt sind.

Wenn man diese Orbitals, wie im vorletzten Absatz erklärt wurde, mit Hilfe einer Matrix darstellt, so ergibt sich das Farbmuster in Abbildung 4.1a (Farbabbildungen: S. 139).

Schluß

Wer gerne mehr über Graphentheorie lesen möchte, findet dazu eine reichhaltige Literatur vor. Eine gute Einführung in diesen Bereich der Kombinatorik ist [Harary, 1974]. Weitere Standardwerke mit spezieller Ausrichtung wurden bereits früher in den einzelnen Abschnitten genannt. Aktuelle Informationen, Problemsammlungen und Lösungsalgorithmen findet man auch im Internet, zum Beispiel unter der Adresse <http://www.informatik.hu-berlin.de/~weinert/graphs.html>.

Anmerkungen

- ¹ Leonhard Euler, 1707 - 1783 .
- ² Robert Kirchhoff, 1824 - 1887.
- ³ Das hier behandelte Problem ist das diskrete Analogon zu dem sogenannten *Dirichlet'schen Problem* der Analysis, bei dem die Lösung der Laplace'schen Differentialgleichung $\Delta f(x) = 0$ für x im Inneren eines Bereiches B und mit gegebenen Werten für f auf dem Rand von B gesucht wird.
- ⁴ Kazimierz Kuratowski, 1896 - 1980.
- ⁵ K. Menger, 1902 - 1985, österreichischer Mathematiker, war von 1938 an bis zu seinem Tode in den USA tätig.
- ⁶ William Rowan Hamilton, 1805-1865.
Hamilton erfand ein Spiel, zu dem ein Spielbrett gehörte, auf dem das Skelett eines Dodekaeders (also ein Graph) aufgezeichnet war. Ferner wurden Spielmarken benötigt, die von 1 bis 20 durchnummeriert waren. Der erste Spieler markierte ein Anfangsstück einer Wanderung durch den Graphen, der zweite Spieler sollte dieses Anfangsstück zu einer Rundwanderung durch alle Knoten längs eines geschlossenen Kreises ergänzen. In einer anderen Version war das Spielbrett durch eine hölzerne Weltkugel ersetzt, auf der Städte und Reisewege durch kleine Stifte markiert waren. Die Aufgabe war eine "Reise um die Welt" zu finden (und mit Hilfe eines Fadens zu markieren), wobei jede Stadt genau einmal zu besuchen war.

Literatur

- [Appel und Haken, 1976] Appel, K., Haken, W. (1976). Every planar map is four colorable. *Bull. Amer. Math. Soc.* 82, 711–712.
- [Berge, 1961] Berge, C. (1961). Färbung von Graphen, deren sämtliche bzw. deren ungeraden Kreise starr sind. *Wiss. Z., Martin-Luther-Univ., Halle-Wittenberg Math.-Naturw. Reihe*, 114–115.
- [Brandstädt et al., 1999] Brandstädt A., Le V. B. und Spinrad J. P. (1999). Graph Classes - A Survey. *SIAM Monographs on Discrete Mathematics and Applications*.
- [Chen, 1971] Chen, W. K. (1971). *Applied Graph Theory*. North-Holland Publishing Company.
- [Colbourn, 1984] Colbourn, C. J. (1984). *The Combinatorics of Network Reliability*. Oxford University Press.
- [Doyle und Snell, 1984] Doyle, P., Snell und J. L. (1984). Random walks and electric networks. *The Carus Mathematical Monographs*, 22, Mathematical Association of America.

- [Euler, 1736 (1741)] Euler, L. (1736 (1741)). Solutio problematis ad geometriam situs pertinentis. *Commentarii Academiae Petropolitanae* 8, 128–140.
- [Garey und Johnson, 1979] Garey M. R. und Johnson D. S. (1979). *Computers and Intractability*. Freeman.
- [Golumbic, 1980] Golumbic, M. Ch. (1980). *Algorithmic Graph Theory and Perfect Graphs*. Academic Press.
- [Harary, 1974] Harary, F. (1974). *Graphentheorie*. Oldenbourg.
- [Kirchhoff, 1847] Kirchhoff, G. (1847). Über die Auflösung von Gleichungen, auf welche man bei der Untersuchung der linearen Verteilung galvanischer Ströme geführt wird. *Poggendorf Ann. Phys.* 72, 497-508.
- [Kuratowski, 1930] Kuratowski, K. (1930). Sur le problème des courbes gauches en topologie. *Fund. Math.* 15, 271-283.
- [Lawler et al., 1985] Lawler, E. L., Lenstra, J. K., Rinnooy Kan A. H. G. und Shmoys, D. B., Hrsg. (1985). *The Traveling Salesman Problem*. John Wiley & Sons.
- [Lovász, 1972] Lovász, L. (1972). Normal hypergraphs and the perfect graph conjecture. *Discrete Math.* 2, 253–267.
- [Menger, 1927] Menger, K. (1927). Zur allgemeinen Kurventheorie. *Fund. Math.* 10, 96–115.

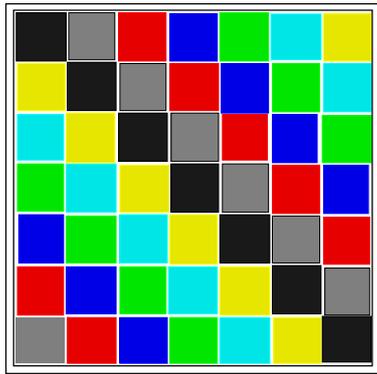
Farbabbildungen



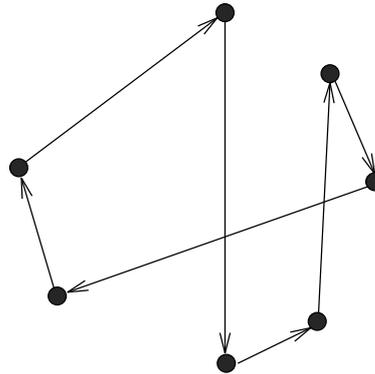
Abbildung 1.19. zu Kapitel 1 von Peter Weibel.
Mario Merz, Ohne Titel.
Foto: Koinegg



Abbildung 1.20. zu Kapitel 1 von Peter Weibel.
Wolfgang Buchner, Studie zu Variationen der Pascalschen Schnecke.
Foto: Koinegg



a)



b)

Abbildung 4.1. zu Kapitel 4 von Gottfried Tinhofer

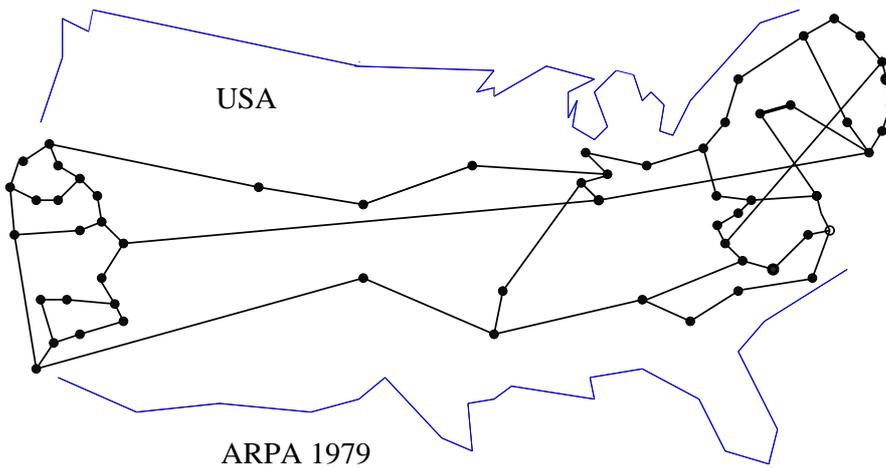


Abbildung 4.11. zu Kapitel 4 von Gottfried Tinhofer

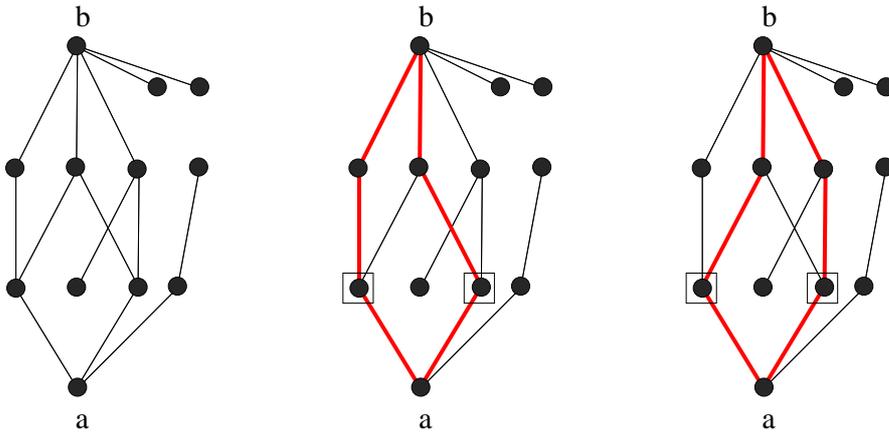


Abbildung 4.12. zu Kapitel 4 von Gottfried Tinhofer

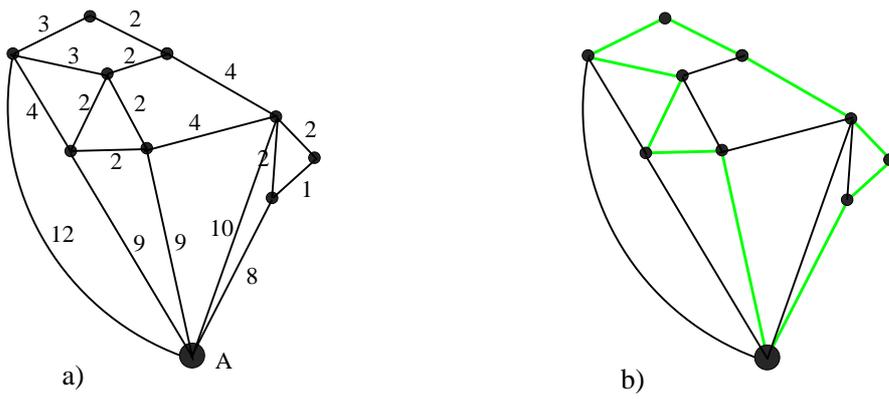


Abbildung 4.13. zu Kapitel 4 von Gottfried Tinhofer

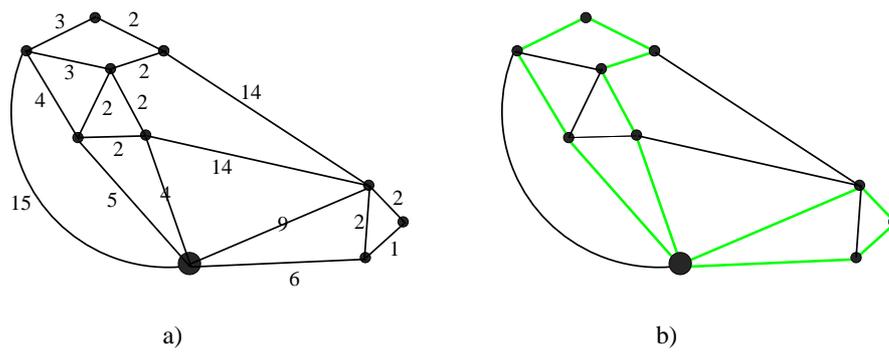
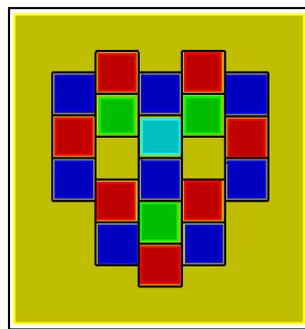
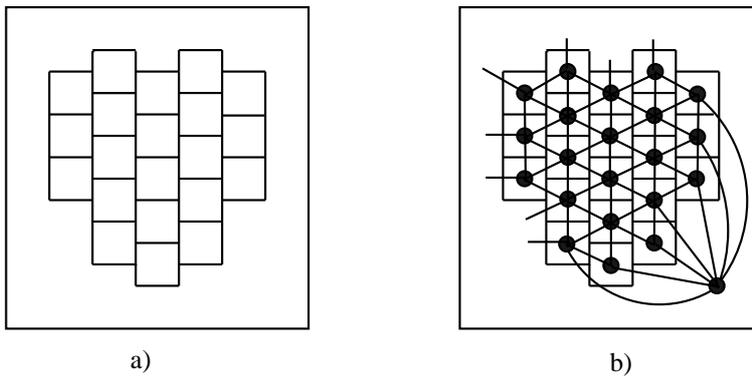


Abbildung 4.14. zu Kapitel 4 von Gottfried Tinhofer



c)

Abbildung 4.15. zu Kapitel 4 von Gottfried Tinhofer

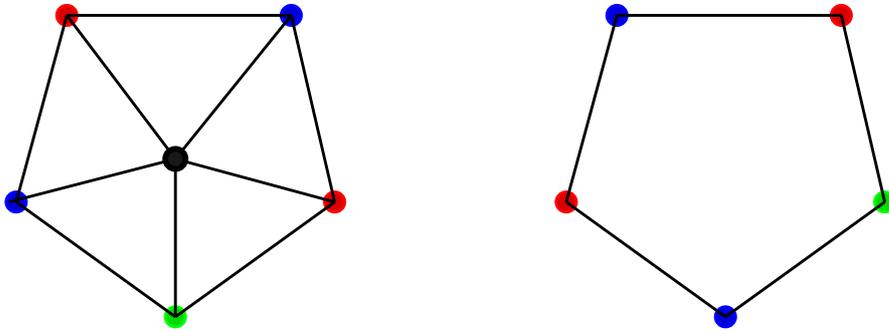


Abbildung 4.16. zu Kapitel 4 von Gottfried Tinhofer

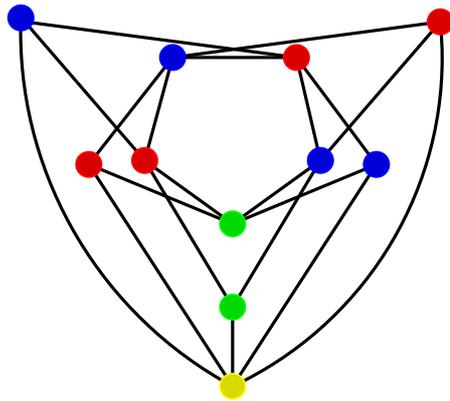
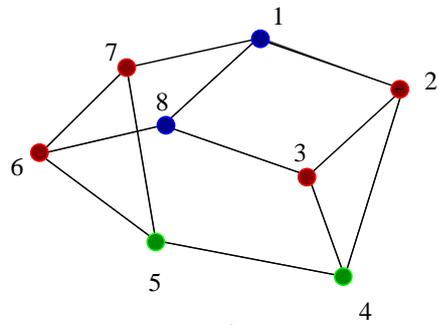
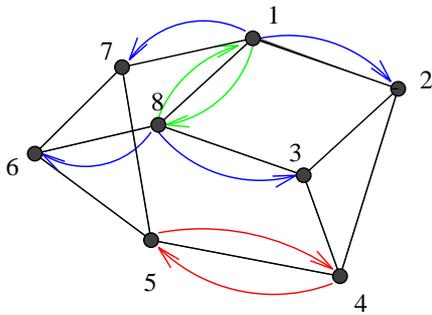


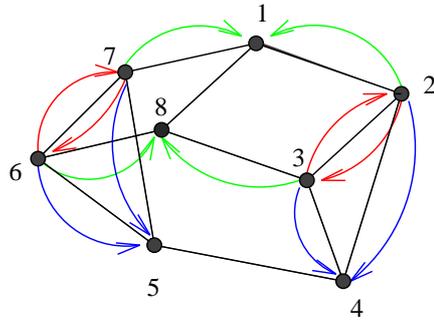
Abbildung 4.17. zu Kapitel 4 von Gottfried Tinhofer



a)

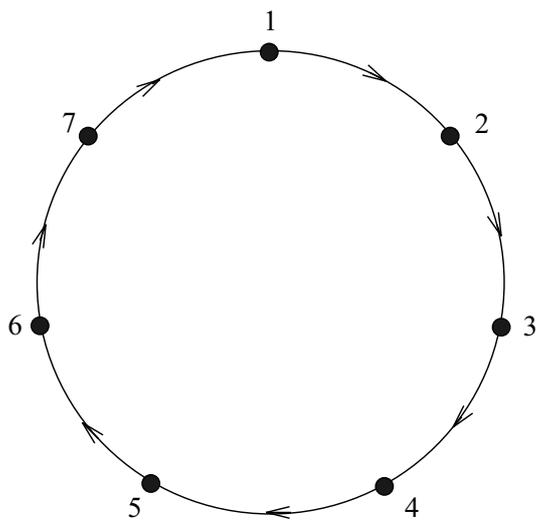


b)

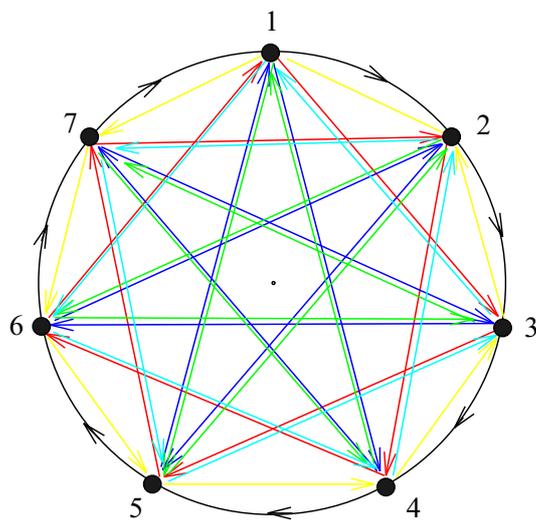


c)

Abbildung 4.23. zu Kapitel 4 von Gottfried Tinhofer



a)



b)

Abbildung 4.25. zu Kapitel 4 von Gottfried Tinhofer

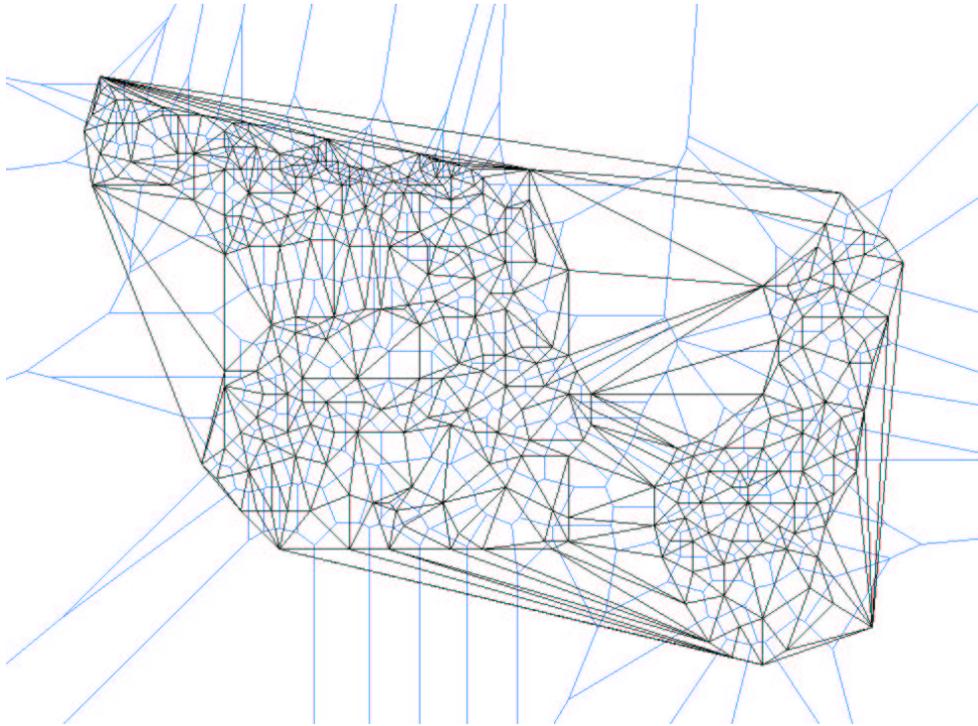


Abbildung 5.4. zu Kapitel 5 von Herbert Edelsbrunner. Die Punkte illustrieren den Baumbestand des Allerton Parkes in Monticello, Illinois. Jedem Punkt oder Baum wird der ihm am nächsten liegende Bereich im ebenen Umfeld zugeordnet.

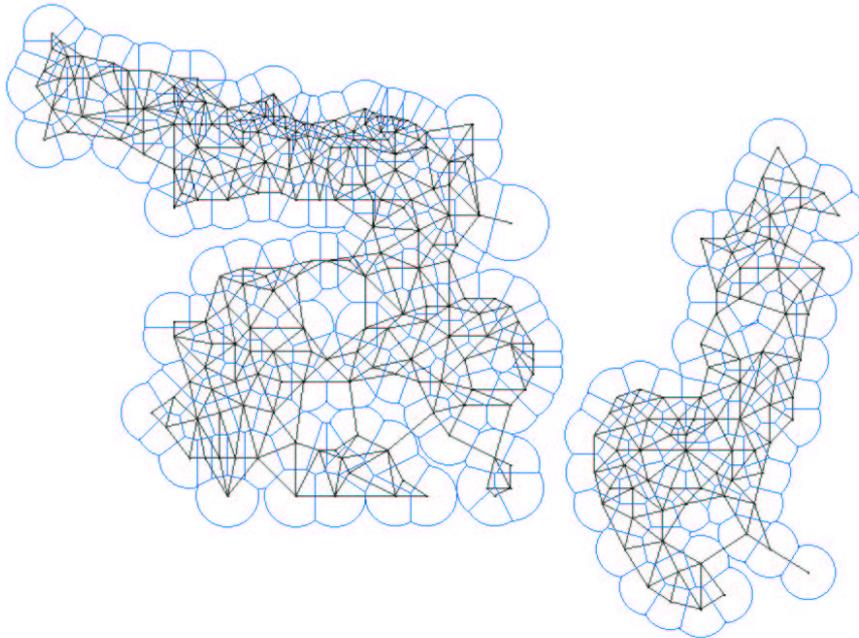


Abbildung 5.5. zu Kapitel 5 von Herbert Edelsbrunner. Jede Voronoi Region wird auf die Einheitskreisscheibe um den Punkt beschränkt und die resultierenden Gebiete erzeugen den dualen Komplex.

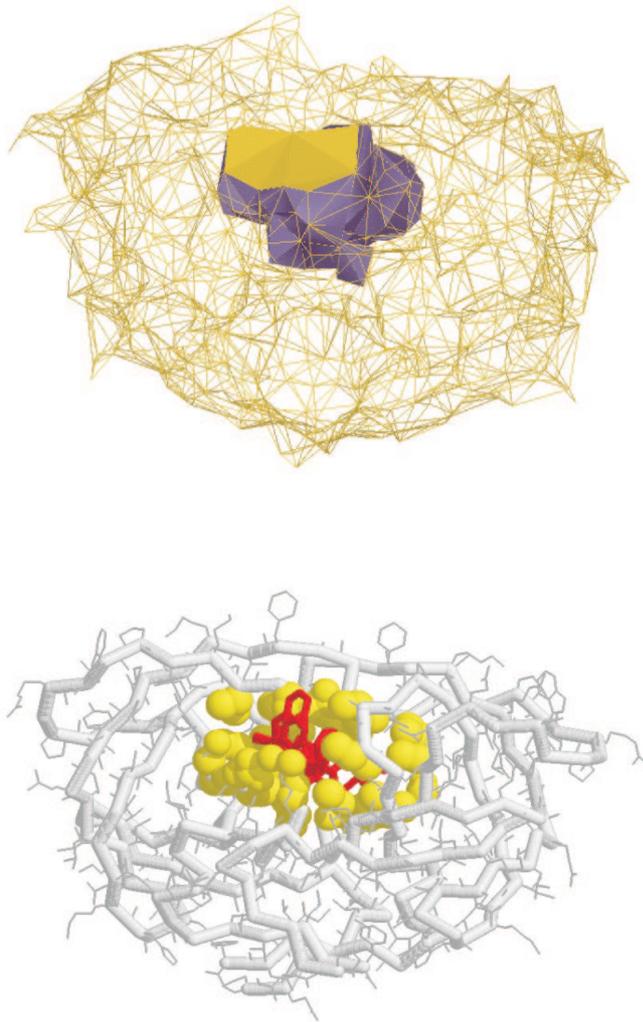


Abbildung 5.7. zu Kapitel 5 von Herbert Edelsbrunner. HIV-1 Protease. Oben beschreiben Tetraeder der Delaunay Triangulierung den dualen Komplex der Tasche. Unten sehen wir die Randkugeln derselben Tasche.

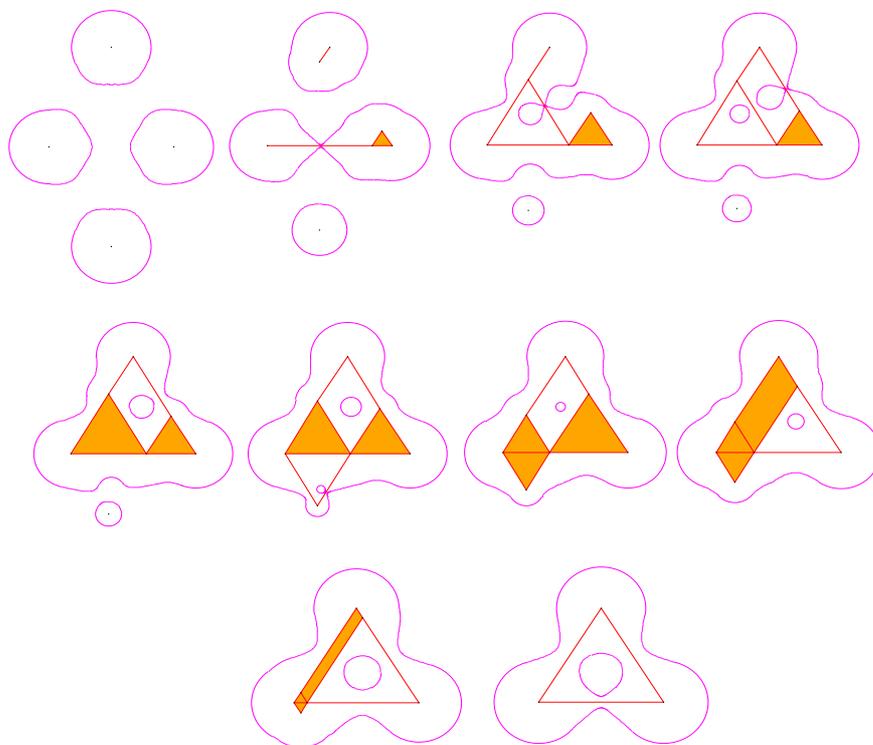


Abbildung 5.9. zu Kapitel 5 von Herbert Edelsbrunner. Zehn Schnappschüsse einer Verformung. Die Form wird durch die glatte Kurve begrenzt und durch den dualen Komplex im Inneren kombinatorisch aufrechterhalten.

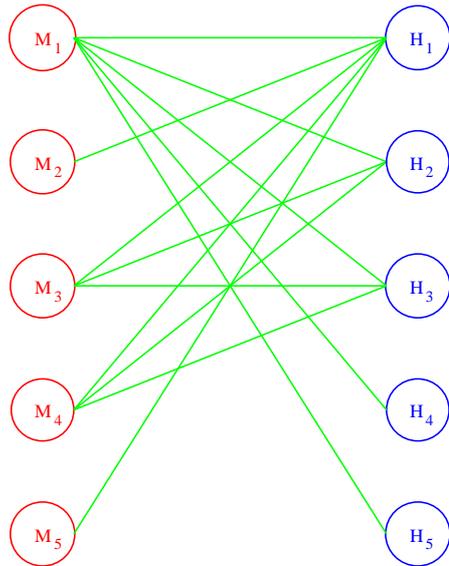


Abbildung 7.1. zu Kapitel 7 von Rainer E. Burkard. Freundschaftssystem (paarer Graph) zwischen 5 Mädchen und 5 jungen Herren. Mädchen M_1 und der junge Herr H_1 sind besonders beliebt, denn sie sind mit allen anderen befreundet.

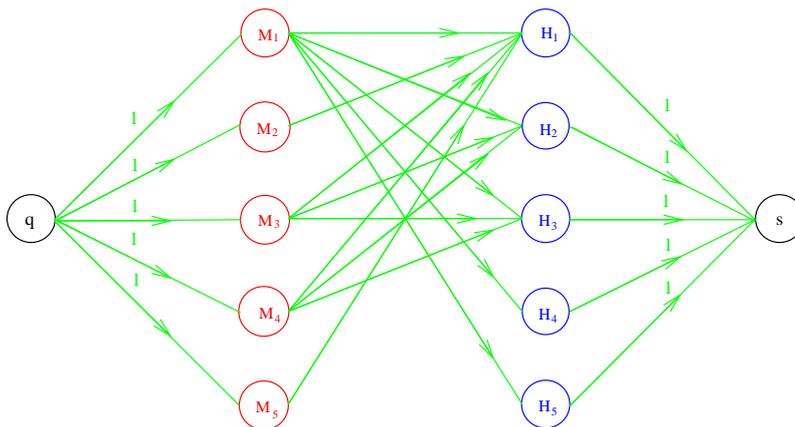


Abbildung 7.2. zu Kapitel 7 von Rainer E. Burkard. Netzwerk, das zum Freundschaftssystem von Abbildung 4.1 gehört. Gesucht wird ein maximaler Fluß von der Quelle q zur Senke s . Die Pfeile von q nach M_i und von H_j nach s haben die Kapazität 1.

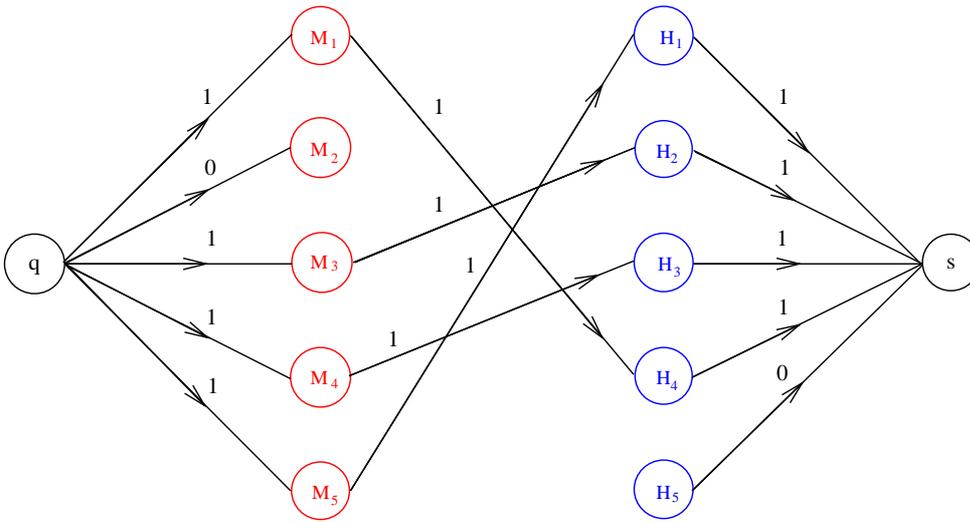


Abbildung 7.3. zu Kapitel 7 von Rainer E. Burkard. Optimale Lösung zum maximalen Flußproblem aus Abbildung 7.2. Die Pfeile tragen nun die entsprechenden Flußwerte. Diese Lösung entspricht der folgenden Paarung: M_1 und H_4 , M_3 und H_2 , M_4 und H_3 , M_5 und H_1 .

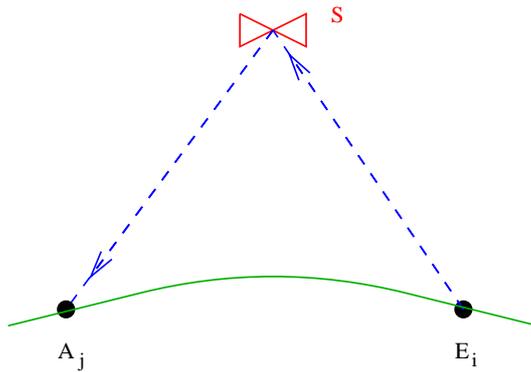


Abbildung 7.4. zu Kapitel 7 von Rainer E. Burkard. Übermittlung von Nachrichten der Sendestation E_i an die Empfangsstation A_j über den Satelliten S .

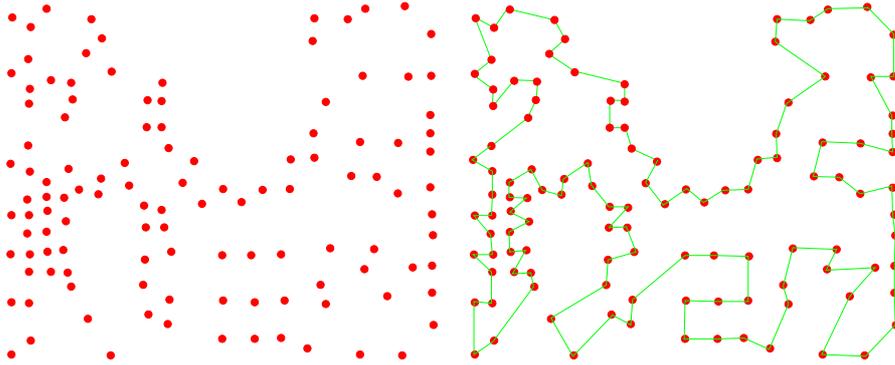


Abbildung 7.5. zu Kapitel 7 von Rainer E. Burkard. Die linke Abbildung zeigt die Positionen von Bohrlöchern auf einer Leiterplatte (Computer Chip). Die rechte Abbildung zeigt die kürzeste Tour, die die Bohrlöcher miteinander verbindet. Diese Reihenfolge erhält man als Lösung des zugehörigen Rundreiseproblems. Sie gibt an, wie ein Bohrer in optimaler Weise vorgehen sollte, wenn sehr viele gleiche Leiterplatten hergestellt werden.

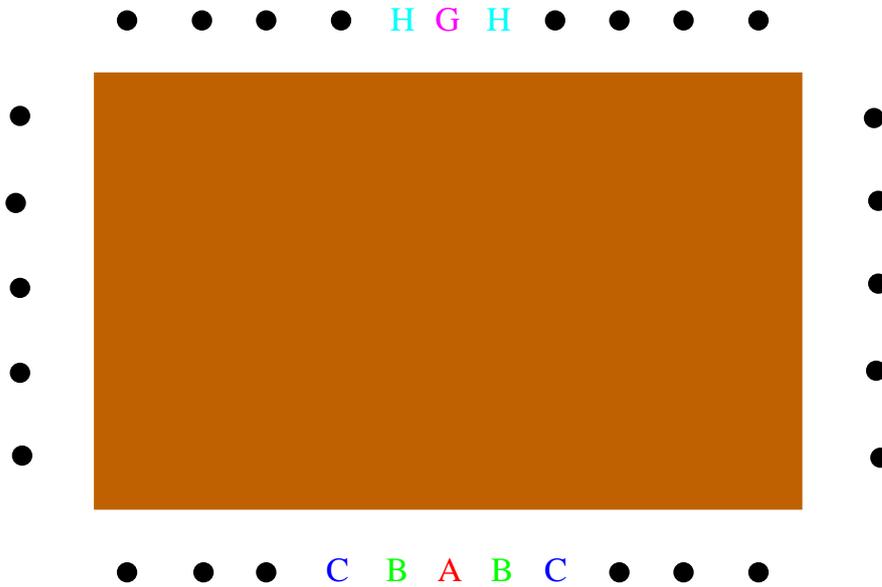


Abbildung 7.6. zu Kapitel 7 von Rainer E. Burkard. Die Person auf Platz A kann gut mit den Personen auf den Plätzen B, weniger gut mit den Personen auf den Plätzen C und G und noch weniger gut mit den Personen auf den Plätzen H kommunizieren.

Spielereien mit Kreisen und Kugeln. Zum Thema Form und Verformung Herbert Edelsbrunner

Zusammenfassung

Allzulange wurde die spielhafte Beschäftigung als Gegensatz zu ernsthafte Arbeit gesehen. Dieser Artikel propagiert die spielerische Untersuchung von Kreis- und Kugelmengen. Gleichzeitig belegt er die nutzbare Anwendung von elementaren Einsichten in der Molekularbiologie und allgemeiner in der Beschreibung von Form und Verformung.

Stichworte. Kreise, Kugeln, Fläche, Volumen, Komplexe, Form, Biologie, Wachstum, Verformung, Formenraum.

Einleitung

Vor mehr als 350 Jahren proklamierte Galileo Galilei, daß das grandiose Buch des Universums in der Sprache der Mathematik geschrieben sei und daß Kreise, Dreiecke und andere geometrische Figuren die Buchstaben und Zeichen dieser Sprache ausmachten. Ohne Kenntnis dieser Zeichen sei es uns unmöglich, ein einziges Wort dieses Buches zu verstehen und seien wir zu Irrfahrten in einem Labyrinth des Dunkels verurteilt [Galilei, 1960].

Von Kreis und Dreieck zu Form

Wir nehmen den Rat von Galilei wörtlich und suchen den Weg zum Verständnis von natürlichen Zusammenhängen im Studium von Kreisen und Dreiecken. Wir verwenden diese als Grundbestandteile der Beschreibung von ebenen geometrischen Formen. Unser Ziel ist eine Theorie der Beschreibung, Analyse, und Manipulation von komplizierten Formen mittels des Computers. Um die Theorie auf den alltäglichen 3-dimensionalen Raum auszuweiten, untersuchen wir Kugeln und Tetraeder, die wir als Verwandte von Kreisen und Dreiecken auffassen. Im allgemeinen leitet uns das Prinzip der Allgemeinheit: Augenmerk wird auf Tatsachen konzentriert, die in beliebigen und nicht nur in zwei oder drei Dimensionen gelten.

Relationen, die auf geometrischer Form beruhen, sind in unserer Welt so

häufig wie die sprichwörtlichen Bäume, die uns den Wald nicht erkennen lassen. Schlüssel, die Türen öffnen, Formen, die uns schön erscheinen, chemische Substanzen, die miteinander reagieren, und so fort. Die vielleicht interessanteste aber auch außerordentlich schwierige Frage ist auf die Beziehung *zwischen* Formen gerichtet. Wir leben und erleben ohne Unterlaß den Unterschied und nicht das Absolute. Sogar das Sehvermögen, das uns konstant und untrüglich erscheint, beruht auf dem Prinzip des Unterschieds und des Komplements. Beim jetzigen Stand der Theorie werden wir der zentralen Frage der Beziehung zwischen Formen nicht gerecht, wir verwenden sie aber als Ziel in der Ferne, auf das sich unsere Schritte richten.

Mathematik und Informatik

Vielleicht ist dem Leser nicht klar was die Diskussion von Form und Relation mit Mathematik zu tun hat. Der Autor sieht die Mathematik als Instrument des Erkennens, als abstraktes Auge. Die Mathematik hat mit der Kunst gemein, daß sie Ordnung in einem Chaos von unendlichen Phänomenen sucht. Vielleicht hat die Mathematik den Vorteil, daß sie auf dem unbestechlichen Prinzip von falsch und richtig basiert. Das macht ihren Fortschritt mühsam erzeugt dafür aber Einsichten, die von Wetter und Zeit nicht unmittelbar abhängen.

Die außerordentliche Mühe mit der sich Fortschritt und Anwendung in der Mathematik verbinden ist wohl der größte Hemmschuh für unser abstraktes Auge. Die Informatik zeigt einen Weg die Mathematik nutzbar zu machen, ohne unseren eigenen Denkapparat dabei zu investieren. Sie ist ein Vehikel das der Mathematik die Leichtigkeit verleiht, die notwendig ist, sie allgemeiner als zusätzlichen Sinn zielsicher einzusetzen. Umgekehrt bietet die Mathematik den notwendigen Inhalt, um der Informatik den Hauch des Lebens zu verleihen. Ein großer Schritt in die Zukunft besteht in der Entwicklung der symbiotischen Beziehung zwischen Mathematik und Informatik, die nach der Meinung des Autors teilweise hinter aber hauptsächlich noch vor uns liegt.

Übersicht

Im Abschnitt *Mehr oder weniger Fläche* beschäftigen wir uns mit einer elementaren aber bis heute unbeantworteten Frage über Einheitskreisscheiben. Im Abschnitt *Teilkomplexe* verwenden wir alternierende Summen, um Fläche und Volumen zu berechnen. Hier erklären wir auch die duale Verbindung

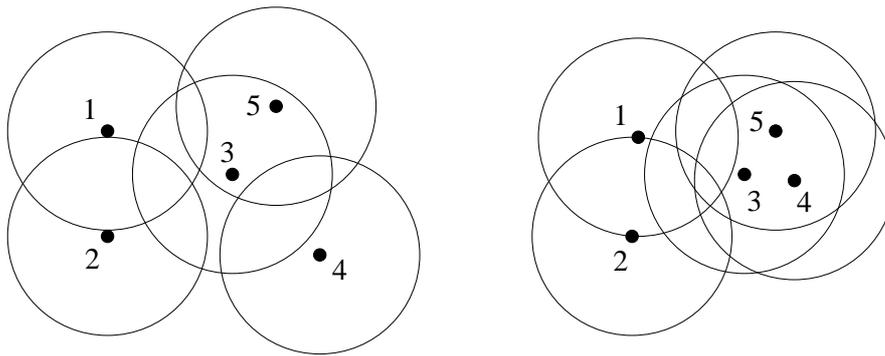


Abbildung 5.1. Zwei Mengen von je fünf Kreisscheiben mit Einheitsradius. Jedes Paar von Kreisscheiben in der rechten Menge überlappt in einem größeren Bereich als das entsprechende Paar der linken Menge.

zwischen Kreis und Dreieck, auf der unser Denken über geometrische Form basiert. Im Abschnitt *Bausteine des Lebens* geben wir einen Einblick in die Anwendung unserer Ideen in der Molekularbiologie. Im Abschnitt *Form und Formenraum* stellen wir allgemeine Formen mit glatter Oberfläche als Einhüllende von unendlichen Kugelmengen dar. Diese Idee führt zu geometrischen Theorien von Verformung und Formraum.

Mehr oder weniger Fläche

Die folgende Frage über Einheitskreisscheiben reicht mehr als 40 Jahre in die Vergangenheit und ist erstaunlicherweise bis heute unbeantwortet: nimmt die abgedeckte Fläche notwendigerweise ab, oder genauer gesagt nicht zu, wenn die Mittelpunkte näher aneinander rücken? Auf den ersten Blick scheint die Antwort ein klares ja zu sein, und es ist bis heute niemandem gelungen, diese Antwort zu widerlegen. Allerdings gibt es bis heute auch keinen Beweis, der die Möglichkeit eines Gegenbeispiels ausschließt.

Problemformulierung

Sei \mathcal{A} eine Menge von n Einheitskreisscheiben A_i mit Mittelpunkten a_i in der Ebene. Analog sei \mathcal{B} eine Menge von n Einheitskreisscheiben B_i mit Mittelpunkten b_i . Wir bezeichnen mit $\|a - b\|$ die euklidische Distanz zwischen zwei Punkten a und b .

Vermutung. Wenn $\|a_i - a_j\| \geq \|b_i - b_j\|$, für alle $1 \leq i, j \leq n$, dann decken

die Kreisscheiben A_i nicht weniger Fläche ab als die Kreisscheiben B_i :

$$\text{Fläche}(\bigcup A) \geq \text{Fläche}(\bigcup B).$$

Abbildung 5.1 illustriert die Fragestellung und die Vermutung für zwei Mengen von je $n = 5$ Einheitskreisscheiben. Wir sehen, daß Kreismittelpunkte in der rechten Menge näher beieinander liegen, weil jeder paarweise Schnitt von links nach rechts zunimmt.

Kurze Geschichte

Das Kreisscheibenproblem wurde unabhängig voneinander von Kneser und von Thue Poulsen in den fünfziger Jahren des zwanzigsten Jahrhunderts aufgeworfen [Kneser, 1955, Thue Poulsen, 1954]. Klee and Wagon beschreiben es als Problem 3 in ihrer Sammlung von offenen Problemen in der Geometrie und Zahlentheorie [Klee und Wagon, 1991]. Eine einfachere Version des Problems nimmt an die Kreisscheiben können so kontinuierlich von Konfiguration A zu Konfiguration B bewegt werden, daß die Distanz zwischen je zwei Mittelpunkten in keinem noch so kleinen Zeitintervall zunimmt. In diesem Fall existiert eine *kontinuierliche Kontrahierung* von A nach B . Diese Version des Problems wurde 1968 von Bollobás positiv beantwortet [Bollobás, 1968]. Vielleicht ist das Überraschende bei dem Kneserschen Kreisscheibenproblem, daß es in der Tat Konfigurationen A and B gibt, die die Voraussetzung der Vermutung erfüllen, für die aber keine kontinuierliche Kontrahierung von A nach B führt. Die Existenz solcher Konfigurationen macht die Schwierigkeit des Problems aus.

Kontinuierliche Kontrahierung

Wir skizzieren einen alternativen und bislang unveröffentlichten Beweis für die positive Beantwortung der Vermutung unter der Annahme einer kontinuierlichen Kontrahierung. Der Beweis gibt uns die Gelegenheit, einige grundlegende kombinatorische Denkweisen auf spielerische Art einzuführen.

Statt der Fläche betrachten wir den Umfang der Vereinigung von Kreisscheiben. Da die Fläche als Integral des Umfangs für wachsenden Radius verstanden werden kann, impliziert die Abnahme des Umfanges die Abnahme der Fläche. Abbildung 5.2 zeigt den Umfang der zwei Mengen von Kreisscheiben in Abbildung 5.1, beide mit positiver Orientierung. Außerdem zeigt das Bild einen dualen Polygonzug von Kreisstücken. Die-

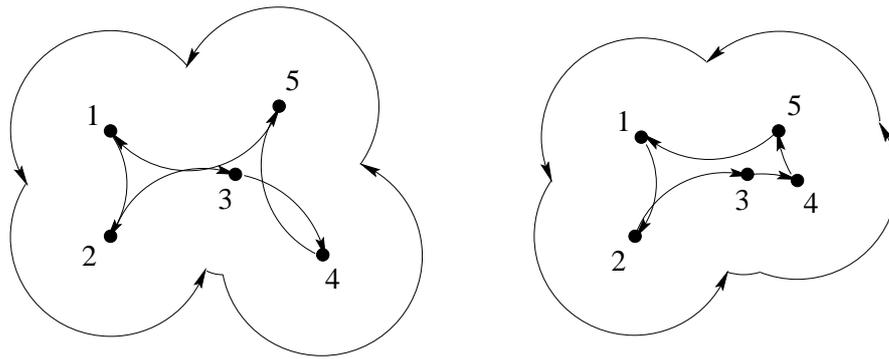


Abbildung 5.2. Orientierter Umfang und dualer Polygonzug von Kreisstücken.

ser wird erzeugt, indem wir den Zirkel mit Einheitsradius an den Eckpunkten des Umfangs ansetzen und Kreismittelpunkte ebenfalls in positiver Orientierung verbinden. Wenn wir die Kreisstücke des Umfangs und des dualen Polygonzuges abwechselnd hintereinanderfügen, erzeugen wir einen neuen Polygonzug ohne scharfen Endpunkt: die Tangente ist an jedem Punkt des Zuges eindeutig gegeben. Der neue Polygonzug ist geschlossen und der Anteil des Umfangs ist genau um einen Einheitskreis größer als der Anteil des dualen Polygonzuges. Wenn wir nun die Konfiguration A kontinuierlich nach B kontrahieren, kann sich jedes Stück des dualen Zuges nur verkürzen. Der Grund ist einfach, daß sich Kreismittelpunkte nur aufeinander zu bewegen. Die Gesamtlänge des dualen Polygonzuges kann sich also nur verkürzen, und da der Längenunterschied konstant ist, gilt dasselbe für den Umfang.

Jetzt sehen wir auch warum der Beweis nicht direkt geführt werden kann: individuelle Kreisstücke des Umfangs können von links nach rechts wachsen, und ein Beispiel dafür ist der Beitrag von Kreis 3 zum Umfang in Abbildung 5.2. Obwohl individuelle Kreisstücke zunehmen haben wir belegen können, daß der Gesamtumfang nicht zunimmt. Der Vollständigkeit wegen erwähnen wir die Möglichkeit von unzusammenhängenden Umfängen und der Änderung der kombinatorischen Beschreibung des Umfangs. Ohne weiter ins Detail zu gehen sagen wir nur, daß das obige Argument auf diese allgemeineren Umstände leicht erweitert werden kann.

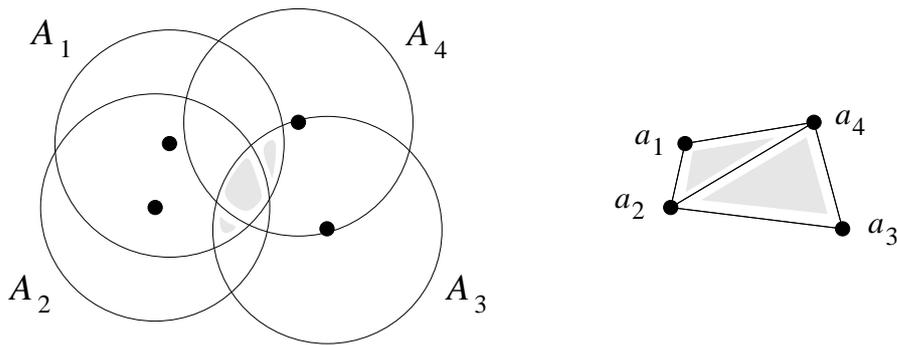


Abbildung 5.3. Die schattierten Teilflächen der Vereinigung von Kreisscheiben gehören zu den redundanten Termen der Flächengleichung. Die Punkte, Kanten und Dreiecke der Zeichnung rechts entsprechen den Termen einer nicht-redundanten Gleichung.

Teilkomplexe

Die erste Reaktion auf das Knesersche Kreisscheibenproblem ist üblicherweise Unglauben, daß die Frage wirklich schwierig ist: "warum können wir es nicht einfach mit dem Prinzip der Inklusion-Exklusion beantworten?". Obwohl es sich herausstellt, daß die Sache nicht so einfach liegt, steht es dafür sich mit diesem mathematischen Prinzip kurz auseinanderzusetzen.

Vier Kreisscheiben

Das Prinzip von Inklusion-Exklusion sagt aus, daß die Fläche der Vereinigung von Kreisscheiben folgendermaßen berechnet werden kann: addiere die Einzelflächen, subtrahiere die Flächen der paarweisen Schnitte, addiere die Flächen der dreifachen Schnitte, und so weiter. Um dieses Resultat formal hinzuschreiben, notieren wir die Fläche von A_i als F_i , die Fläche des Durchschnittes $A_i \cap A_j$ als F_{ij} , und so weiter. Für die vier Kreisscheiben in Abbildung 5.3 ergibt das

$$\begin{aligned}
 F &= F_1 + F_2 + F_3 + F_4 \\
 &\quad - F_{12} - F_{13} - F_{14} - F_{23} - F_{24} - F_{34} \\
 &\quad + F_{123} + F_{124} + F_{134} + F_{234} \\
 &\quad - F_{1234}.
 \end{aligned}$$

Von Abbildung 5.3 lesen wir $F_{13} + F_{1234} = F_{123} + F_{134}$ ab. In Worten, die vier Summanden in der Gleichung für F sind redundant. Jeder nicht-redundante Term der verbleibenden Gleichung entspricht einem Element des dualen Komplexes, der rechts von den vier Kreisen in Abbildung 5.3 gezeichnet ist. Die vier Punkte entsprechen den Kreisscheiben, die fünf Kanten entsprechen den nicht-redundanten paarweisen Schnitten, und die zwei Dreiecke entsprechen den nicht-redundanten dreifachen Schnitten. Unser Ziel ist nun, den Begriff des dualen Komplexes auf möglichst ökonomische and natürliche Art und Weise so zu erweitern, daß seine Elemente eine Flächengleichung ohne redundante Terme beschreiben. Wir müssen ein wenig ausholen.

Voronoi und Delaunay Komplexe

Wir bestimmen für jeden Punkt x der Ebene den nächsten Mittelpunkt a_i eines Kreises in \mathcal{A} . Die *Voronoi Region* von a_i ist die Menge aller Punkte x für die a_i der nächste Mittelpunkt ist. Die Menge der Voronoi Regionen zerlegen die Ebene wie in Abbildung 5.4 (Farbabbildungen: S. 146). Wenn zwei Regionen überlappen, dann tun sie das entlang einer gemeinsamen Kante. Wenn drei Regionen überlappen, dann tun sie das in einem gemeinsamen Punkt. Im allgemeinen Fall treten keine anderen Überlappungen auf. Die *Delaunay Triangulierung* ergibt sich als duales Diagramm. Die a_i sind Knotenpunkte der Triangulierung. Zwei Punkte werden durch eine Kante verbunden, wenn sich die zwei entsprechenden Voronoi Regionen überlappen. Drei Punkte werden durch ein Dreieck verbunden, wenn sich die drei entsprechenden Voronoi Regionen überlappen. Abbildung 5.4 (Farbabbildungen: S. 146) zeigt die Delaunay Triangulierung über die Voronoi Regionen gezeichnet.

Duale Komplexe

Der *duale Komplex* von \mathcal{A} ergibt sich durch Selektion von Punkten, Kanten, und Dreiecken der Delaunay Triangulierung. Eine Kante gehört zum dualen Komplex, wenn die Überlappung der zwei entsprechenden Voronoi Regionen in die Vereinigung der Kreisscheiben hineinreicht. Ein Dreieck gehört zum dualen Komplex, wenn der Punkt, in dem sich die drei entsprechenden Voronoi Regionen überlappen, in der Vereinigung der Kreisscheiben liegt. Abbildung 5.5 (Farbabbildungen: S. 147) zeigt, wie die Kreisscheiben einen Teilkomplex der Delaunay Triangulierung selektieren.

Wir stellen uns vor, die Kreisscheiben, die den dualen Komplex definieren, wachsen kontinuierlich und gleichzeitig. Jede Scheibe deckt einen wachsenden Teil des unveränderten Voronoi Gebietes ab. Somit ergeben sich immer mehr Berührungspunkte und damit immer mehr Verbindungen im dualen Komplex. Anders ausgedrückt macht der duale Komplex einen zunehmenden Teilkomplex der Delaunay Triangulierung aus. Im Endstadium des Wachstums nimmt der duale Komplex die ganze Delaunay Triangulierung ein.

Inklusion-Exklusion Formeln

Mit der Einführung des dualen Komplexes ist die Berechnung der Fläche, die eine endliche Menge von Einheitskreisscheiben abdeckt, denkbar einfach: verwende den dualen Komplex als Indikator in der Auswahl und addiere Einzelflächen, subtrahiere paarweise Schnittflächen und addiere dreifache Schnittflächen. Wir behaupten diese Berechnung ergibt die Fläche der Vereinigung genau und nicht nur annähernd. Warum das so ist, kann in [Edelsbrunner, 1995] nachgelesen werden.

Wichtiger als der Beweis der Behauptung ist für uns die Einsicht, daß das angegebene Rezept zur Flächenberechnung wesentlich allgemeiner gilt als nur für Einheitskreisscheiben in der Ebene. Sei im weiteren \mathcal{A} eine endliche Menge von Kugeln im d -dimensional Raum. Die Radien sind beliebig und können alle verschieden sein. Wir verallgemeinern Voronoi Regionen, indem wir die euklidische Distanz durch die Potenz ersetzen: die *Potenz* eines Punktes x zu einer Kugel A_i mit Mittelpunkt a_i and Radius r_i ist $\pi_i(x) = \|x - a_i\|^2 - r_i^2$. Optisch verändern sich dadurch die Voronoi Regionen kaum, außer daß zwei Regionen nicht mehr genau in der Mitte zwischen den Kugelmittelpunkten getrennt werden. Wenn wir Abbildung 5.5 (Farbabbildungen: S. 147) genauer untersuchen, sehen wir in der Tat verschieden große Kreise und damit verallgemeinerte Voronoi Regionen in Abbildung 5.4 (Farbabbildungen: S. 146). Die Delaunay Triangulierung ergibt sich als duales Diagramm und der duale Komplex durch die Einschränkung der dualen Korrespondenz auf die Vereinigung der Scheiben.

Für Kugeln in allgemeiner Lage im d -dimensional Raum ist der duale Komplex eine Menge von Simplexen, deren Dimension zwischen 0 und d variiert. Punkte sind 0-dimensional Simplexe, Kanten sind 1-dimensional Simplexe, und so weiter. Kugelmengen in spezieller Lage können infinitesimal in allgemeine Lage perturbiert werden. Geometrisch gesehen ist ein Simplex der Dimension k die konvexe Hülle von $k + 1$ Kugelmittelpunk-

ten. Kombinatorisch gesehen ist derselbe Simplex die Menge der $k + 1$ entsprechenden Kugeln. Um die allgemeine Volumensformel angeben zu können schreiben wir $K = K(\mathcal{A})$ für den dualen Komplex, σ für einen allgemeinen Simplex in K , $\dim \sigma$ für die Dimension von σ , und $\bigcap \sigma$ für den Durchschnitt der $\dim \sigma + 1$ Kugeln in σ .

Satz: Das d -dimensionale Volumen der Vereinigung einer Menge \mathcal{A} von Kugeln im d -dimensionalen Raum ist

$$\text{Volumen}(\bigcup \mathcal{A}) = \sum_{\sigma \in K} (-1)^{\dim \sigma} \cdot \text{Volumen}(\bigcap \sigma).$$

Kurze Geschichte

Voronoi Regionen sind nach dem ukrainischen Mathematiker Gheorghis Voronoi benannt [Voronoi, 1907]. Die Idee zur Aufteilung der Ebene durch solche Regionen läßt sich allerdings wesentlich weiter in die Vergangenheit zurückverfolgen. Die vielleicht ältesten Belege sind Zeichnungen von René Descartes aus dem siebzehnten Jahrhundert. Die dualen Diagramme sind nach dem russischen Mathematiker Boris Delone (auch Delaunay) benannt. Dieser hat 1934 einen Artikel zu diesem Thema veröffentlicht und Gheorghis Voronoi gewidmet [Delaunay, 1934]. Die Geschichte der dualen Komplexe beginnt 1983 mit einer Zusammenarbeit des Autors mit David Kirkpatrick und Raimund Seidel [Edelsbrunner et al., 1983].

Die Idee, daß die Fläche des Durchschnitts von Kreisscheiben durch kurze Inklusion-Exklusions Formeln ausgedrückt werden kann, finden wir erstmals bei Kratky [Kratky, 1978], der allerdings keine genaueren Angaben zum Aussehen dieser Formeln gibt. Naiman und Wynn konkretisieren Kratkys Idee und zeigen 1992, daß die Simplexe der Delaunay Triangulierung eine exakte Formel für die Vereinigung von Einheitskreisscheiben liefern [Naiman und Wynn, 1992]. Die Verallgemeinerung auf beliebige Radien und die Verbesserung von Delaunay Triangulierungen zu dualen Komplexen geht auf die Arbeit [Edelsbrunner, 1995] des Autors dieses Artikels zurück.

Bausteine des Lebens

Die letzten Jahrzehnte zeigen überraschenderweise, daß die geometrische Form eine Schlüsselrolle in der Molekularbiologie spielt. Frei nach Rose

hängt die Rolle eines Proteins in der grandiosen Oper des Lebens auf molekularer Ebene von genau einem ab, nämlich ihrer Form [Rose, 1996]. Für ein Proteinmolekül folgt die Funktion der Form.

Atomkugel Modelle

Geometrische Modelle von Makromolekülen sind anfangs der siebziger Jahre des zwanzigsten Jahrhunderts von Frederick Richards und seinen Studenten eingeführt worden [Lee und Richards, 1971, Richards, 1977]. Der Grundidee folgend wird jedem Atom ein geometrischer Bereich zugeordnet und das Molekül als Gesamtheit dieser Bereiche aufgefaßt. In der einfachsten Version ist jedes Atom eine Kugel und das Molekül die Vereinigung dieser Kugeln. Die verbleibende Frage nach den Radien wird durch Rückrechnung aus experimentellen Betrachtungen der van der Waals Kräfte beantwortet. Die Berechnungen eines eindeutigen van der Waals Radius pro Atomtyp ist nicht unproblematisch, schon aus dem Grund, daß jede Kraftauswirkung die Funktion von zumindest zwei Atomen ist. Trotzdem wird dem van der Waals Modell von Molekülen beträchtliche Aufmerksamkeit gewidmet und ist in jedem einschlägigen Einführungswerk zu finden [Creighton, 1994].

Eine einfache Änderung der Radien erlaubt die Untersuchung des Verhaltens eines Moleküls innerhalb einer Lösung. Für Wasserlösungen nähern wir ein Wasserteilchen mit einer Kugel von Radius R_0 an. Der Ort der Mittelpunkte von Wasserteilchen, die ein gegebenes Molekül überlappen, ist dann die Vereinigung von Atomkugeln deren Radien um R_0 vergrößert wurden. Zum Beispiel ist das ursprüngliche Molekül in der Lage, Wasserteilchen einzufangen, wenn immer ihr um R_0 vergrößertes Modell Höhlen im Inneren besitzt.

Löcher

Eine interessante Frage von praktischer Relevanz ist die Charakterisierung und Berechnung des Zusammenhangs von Molekülen. Topologisch unterscheiden wir bei drei-dimensionalen Formen drei Arten von Löchern:

Höhlen im Inneren der Form, die vollständig abgeschlossen und dadurch von außen nicht zugänglich sind,

Tunnels, die wie Flugbahnen durch die Form hindurch führen, und

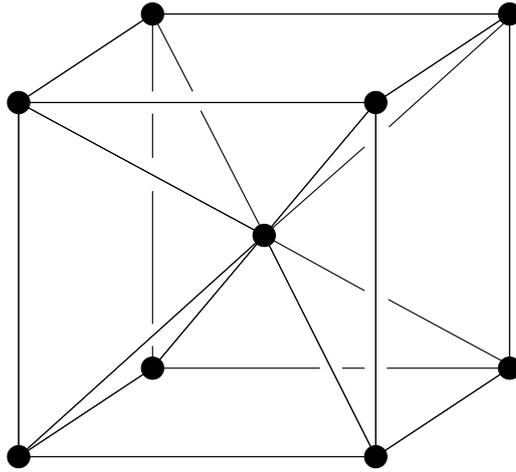


Abbildung 5.6. Ein Skelett von 20 Kanten das dem linken Gebilde in Abbildung 5.8 entspricht.

Brüche, die die Form in zwei oder mehr Zusammenhangskomponenten zerlegen.

Alle drei Arten von Löchern können mathematisch als Ränge von Homologie Gruppen eindeutig abgezählt werden. Wir verweisen auf das Buch von Boltjanskij und Efremovič als eine relativ leicht lesbare Einführung in topologische Konzepte [Boltjanskij und Efremovič, 1986]. Bei gegebener Form im drei-dimensionalen Raum sind Höhlen und Zusammenhangskomponenten eindeutig geometrisch lokalisierbar und damit auch leicht abzählbar. Bei Tunnels ist der Sachverhalt anders, und es ist nicht ganz einfach einzusehen, daß es eine eindeutig bestimmte Anzahl von Tunnels gibt. Wie viele Tunnels hat zum Beispiel das linke Gebilde in Abbildung 5.8? Die abstrakte Version dieses Gebildes ist ein Skelett von Kanten die Eckpunkte und Mittelpunkt eines Würfels wie in Abbildung 5.6 verbinden. Wir können die Tunnels durch inkrementelles Hinzufügen von Kanten abzählen. Die neun Punkte werden bereits mit acht der Kanten zu einer einzigen Komponente verbunden. Jede weitere Kante erzeugt einen Tunnel, also zählen wir insgesamt 12. Das Verwirrende an Tunnels ist die Möglichkeit, sie in verschiedenster Reihenfolge zu durchfahren. In der Tat betrachten Homologie Gruppen jeden möglichen Tunnel als die Kombination von Basistunnels. Verschiedene Basen sind möglich, aber alle minimalen Basen sind gleich groß.

Taschen

Die algorithmische Berechnung von Löchern in Molekülen verbindet sich für den Autor mit einer interessanten Episode. Wie man sich vorstellen kann, ist die Implementierung nicht ohne Aufwand, vor allem wenn Korrektheit und Schnelligkeit notwendige Kriterien sind. Wir erweiterten auch die Inklusion-Exklusion Formeln, um Löcher zu messen. Nach all der Arbeit und investierten Zeit konfrontierten uns die Kollegen von der Biologie mit der Tatsache, daß Löcher in Molekülen, die zählen, gar keine Löcher im topologischen Sinn des Wortes sind. Vielmehr sind es die Vertiefungen in der Oberfläche, die eine Rolle in der Interaktion zwischen Molekülen spielen. Wir bezeichnen solche Löcher im weiteren als *Taschen*. Auf den eigenartigen Gegensatz zwischen dem topologischen und den allgemeinsprachlichen Gebrauch des Wortes Loch haben schon Casati und Varzi in ihrer philosophischen Abhandlung [Casati und Varzi, 1994] hingewiesen. Sie führen einen taschenähnlichen Begriff ein, der aber mathematisch und algorithmisch nicht weiter brauchbar ist.

Als Reaktion auf die Kritik der Biologen haben wir uns entschieden einen kleinen Anteil von Geometrie in die Definition einer neuen Art von Löchern einfließen zu lassen. Dazu betrachten wir den Wachstum des Moleküls, der sich ergibt, wenn wir alle Kugelradien kontinuierlich und gleichzeitig vergrößern. Eine *Tasche* verengt sich am Eingang und verwandelt sich in eine Höhle bevor diese verschwindet und von Kugeln völlig überdeckt wird. Das zeitweilige Erscheinen von Höhlen unterscheidet Taschen von weniger ausgeprägten Vertiefungen. Abbildung 5.7 (Farbabbildungen: S. 148) veranschaulicht den Begriff durch die Darstellung des entsprechenden dualen Komplexes oben und der berandenden Kugeln unten. Die technischen Details und die algorithmische Berechnung von Taschen kann in [Edelsbrunner et al., 1998] nachgelesen werden.

Anwendungen

Um eine Definition zu rechtfertigen, müssen wir ihren theoretischen oder praktischen Nutzen nachweisen. In diesem Fall suchen wir praktische Rechtfertigung. Die computerunterstützte Untersuchung von Makromolekülen und vor allem von Proteinen ist glücklicherweise schon so weit fortgeschritten, daß eine gute Anzahl von Konformationen öffentlich zugänglich ist. Eine Konformation in diesem Zusammenhang ist eine detaillierte geometrische Beschreibung: jedem Atom ist eine Lokalität durch

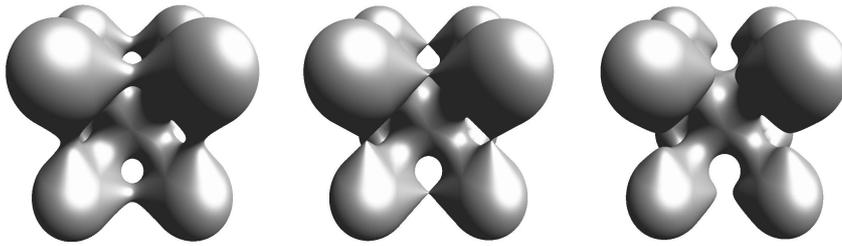


Abbildung 5.8. Jedes der drei Gebilde ist durch neun Kugeln im 3-dimensionalen Raum vollständig definiert. Die Kugelmittelpunkte beschreiben jeweils die acht Ecken und den Mittelpunkt eines Würfels. Die Kugelradien nehmen von links nach rechts ab.

Ort des Zentrums und Radius im drei-dimensionalen Raum zugeschrieben. Wir konnten also unsere Software dazu verwenden, diese Konformationen neu algorithmisch zu studieren. Die Ergebnisse der Untersuchung von 51 Proteinen sind in [Liang et al., 1998] veröffentlicht. Es seien hier nur einige wenige Resultate erwähnt.

Das wichtigste Ergebnis ist, daß in ungefähr 80 Prozent der Fälle das Aktivgebiet des Proteins mit der Tasche des größten Volumens übereinstimmt. Damit haben wir ein rein geometrisches Kriterium gefunden, das mit guter Wahrscheinlichkeit und automatisch die wichtigste Stelle eines Proteins identifiziert. Es stellt sich heraus, daß in vielen Fällen das Aktivgebiet noch von kleineren Taschen in der nahen Umgebung unterstützt wird. Diese können, müssen aber nicht bei einer Molekülanlagerung beteiligt sein. Es ist auch interessant, daß größere Proteine statistisch gesehen keine größeren aber dafür mehr Taschen aufweisen. Die in Abbildung 5.7 (Farbabbildungen: S. 148) dargestellte HIV-1 Protease hat die unübliche Eigenschaft, daß die Konformation sich stark verformen und dadurch eine Tasche von sehr variabler Größe erzeugen kann.

Form und Formenraum

Die Untersuchung von Molekülen deutet schon darauf hin, daß es nicht reicht, nur statische Form zu untersuchen. Wir müssen die Variabilität, die sich mit der Möglichkeit der Verformung ergibt, mit ins Kalkül ziehen.

Glatte Oberflächen

Bevor wir uns in die Behandlung von sich kontinuierlich verändernder Form stürzen, betrachten wir die Modellierung von Formen mit glatter Oberfläche. Wir verwenden wiederum Kugeln, erzeugen aber automatisch Blendflächen, die scharfe Kreisränder an Schnittstellen zwischen Kugeln überbrücken, siehe Abbildung 5.8. Wir verwenden ausschließlich einschalige und zweischalige rotationssymmetrische Hyperboloide sowie konkave Kugelflächen. Zusammen mit den konvexen Kugelflächen der erzeugenden Kugeln macht das unser Repertoire von elementaren Teilflächen aus. Wir überspringen die mathematischen Details der Konstruktion, die in [Edelsbrunner, 1999] nachgelesen werden können. Die wichtigsten Eigenschaften sind wie folgt.

- (1) Die Oberfläche hat keinen Rand und begrenzt damit immer einen dreidimensionalen Körper.
- (2) Die Tangentialebene ist überall definiert und variiert kontinuierlich.
- (3) Die Krümmung ist überall definiert und die lokal maximale eindimensionale Krümmung variiert kontinuierlich.
- (4) Die Fläche ist symmetrisch bezüglich innen und außen, was bedeutet, daß die genau selbe Fläche von zwei verschiedenen Mengen von Kugeln erzeugt wird.

Eigenschaft (4) sollte sich in der Modellierung von Interaktionen zwischen Molekülen als relevant herausstellen. Eine Ausnahme zu Eigenschaften (2) und (3) wird durch das mittlere Gebilde in Abbildung 5.8 aufgezeigt. Im Grenzfall zwischen einschaligem und zweischaligem Hyperboloid ergibt sich ein Doppelkegel an dessen Spitze Tangentialebene und Krümmung nicht definiert sind. Wie wir bald sehen werden sind solche Spitzen notwendig, um die Topologie der Form ohne ruckartige Bewegung zu verändern.

Topologische Verwandlung

Wir verwenden das glatte Oberflächenmodell für geometrische Formen und untersuchen die Art und Weise wie eine Form ihren topologischen Zusammenhang durch kontinuierliche Verformung verändert. Wir beschränken uns zunächst auf Verformung durch Wachstum. Wir erkennen vier topologieverändernde Operationen, von denen eine in Abbildung 5.8 sichtbar ist.

Erscheinung. Eine neue Komponente der Form erscheint aus dem Nichts, zunächst als Punkt, dann als wachsende Kugel, und schlußendlich als möglicherweise kompliziertes Gebilde.

Verbindung. Es bildet sich eine Brücke zwischen zwei Komponenten oder zwischen vorher getrennten Teilen derselben Komponente. Wenn wir Abbildung 5.8 von links nach rechts lesen sehen wir die simultane Bildung von zwölf Brücken, alle zwischen verschiedenen Teilen derselben Komponente.

Zuschüttung. Ein Tunnel verengt sich zu einem Punkt und verschwindet. Geometrisch ist der Tunnel durch ein einschaliges Hyperboloid modelliert. Im Augenblick der Verengung zu einem Punkt mutiert dieses zu einem Doppelkegel und im weiteren zu einem zweischaligen Hyperboloid.

Auffüllung. Eine Höhle im Inneren der Form verkleinert sich und verschwindet. Kurz vor dem Verschwinden sehen wir eine leere Kugel, die sich zu einem Punkt zusammenzieht.

Die vier Operationen entsprechen vier weiteren inversen Operation in der umgekehrten Zeitrichtung. Wir erkennen grundlegende Symmetrien indem wir auch den Ort umkehren. Genauer gesagt tauschen wir eine Form mit ihrem Komplement (dem Außen) und lesen die Geschichte verkehrt. Durch die Umdrehung der Zeitrichtung kehrt sich das Wachstum in sein Gegenteil, aber durch Umdrehung des Ortsinnes ergibt sich wiederum Wachstum. Allerdings kehrt sich Erscheinung in Auffüllung und umgekehrt. Verbindung kehrt sich in Zuschüttung und umgekehrt. Damit verstehen wir alle acht Operationen als symmetrische Varianten von nur zwei Grundtypen. Dem mathematisch gebildeten Leser wird der Morse-theoretische Hintergrund [Milnor, 1963] in der Beschreibung der Operationen nicht verborgen bleiben.

Kanonische Verformung

Obwohl Bewegung wesentlich allgemeiner ist als reines Wachstum, ergeben sich keine neuen topologieverändernden Operationen. Eine allgemeine Verformung kann durch kontinuierliche Verschiebung und Größenänderung der definierenden Kugeln erzeugt werden.

Abbildung 5.9 (Farbabbildungen: S. 149) illustriert eine Verformung in der Ebene. Die Ausgangsform ist durch vier Kreise definiert und verändert

sich in die durch drei Kreise definierte Endform. Die Punkte, Kanten, Dreiecke und Vierecke beschreiben den dualen Komplex der ebenfalls eine Sequenz von Veränderungen durchläuft. Wir sehen, daß der duale Komplex im Inneren der Form zu liegen kommt und zu jedem Zeitpunkt denselben topologischen Zusammenhang aufweist. Im Unterschied zur kontinuierlichen Verformung der glatten Form durchläuft der duale Komplex eine diskrete Sequenz von kombinatorisch verschiedenen Zuständen. Die Erscheinung korrespondiert mit der Erzeugung eines neuen Knotenpunktes im dualen Komplex. In ähnlicher Weise korrespondieren Verbindung, Zuschüttung, und Auffüllung mit der Einfügung einer neuen Kante, eines neuen Dreiecks, eines neuen Tetraeders in den dualen Komplex. Vom algorithmischen Gesichtspunkt ist diese Korrespondenz zwischen glattem Oberflächenmodell und dualem Komplex von erstrangiger Bedeutung. Erst durch sie werden Änderungen in der Topologie vorhersagbar und können rechtzeitig verarbeitet werden. Weitere Details der Berechnung von Verformungen im zwei-dimensionalen Fall sind in [Cheng S.-W. et al., 1998] beschrieben.

Formenraum

Die in Abbildung 5.9 (Farbabbildungen: S. 149) illustrierte Verformung ist das Ergebnis einer kanonischen Bewegung von Ausgangs- zu Endform. Wir ersparen dem Leser die Details erwähnen aber, daß die Verformung nichts mit physikalischen Kräften oder Ähnlichem zu tun hat. Sie beruht einzig und allein auf der geometrischen Interpretation von allgemeiner drei-dimensionaler Form als die Projektion einer Sektion eines vier-dimensionalen konvexen Körpers. Die Verformung kann dann durch konvexe Verbindung von zwei solchen Körpern, die in parallelen Hyperebenen im fünf-dimensionalen Raum eingebettet sind, erzeugt werden. Wie auch immer. Einzelheiten können in [Cheng H.-L. et al., 1998] nachgelesen werden. Wichtig ist einzusehen, daß eine kanonische Verformung einen Formenraum erzeugt.

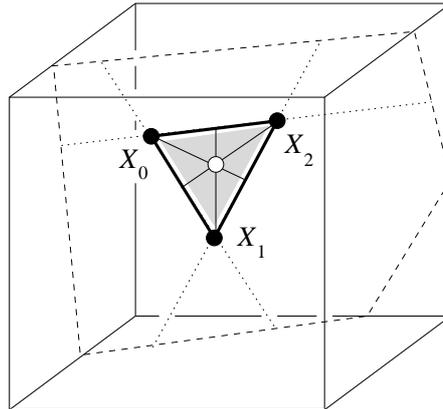


Abbildung 5.10. Eine schematische Darstellung eines zwei-dimensionalen Formenraumes aufgespannt von drei Grundformen. Der Würfel symbolisiert den unendlich-dimensionalen Raum in dem der Formenraum eingebettet ist.

Die Verformung zwischen zwei Formen erzeugt einen ein-dimensionalen Formenraum, nämlich das Zeitintervall der Verformung und mit jedem Zeitpunkt die zugeordnete Form. Wir weiten die paarweise kanonische Verformung auf $k + 1 > 2$ Basisformen aus. Der Fall $k + 1 = 3$ ist in Abbildung 5.10 schematisch dargestellt.

Die Punkte X_0, X_1, X_2 repräsentieren die drei Basisformen. Die drei Kanten des Dreiecks repräsentieren die paarweise Verformung zwischen X_0 und X_1 , X_1 und X_2 , und X_2 und X_0 . Ein Punkt im Inneren des Dreiecks kann durch die Verbindung eines Eckpunktes mit einem Punkt der gegenüberliegenden Kante erreicht werden. In anderen Worten, wir halten die Verformung von X_1 zu X_2 in der Mitte an und verformen diese gemischte Form zu X_0 , unterbrechen aber wiederum die Verformung frühzeitig und zwar genau am Zielpunkt im Inneren des Dreiecks.

Derselbe Punkt kann auf zumindest drei verschiedenen Wegen erreicht werden. Die Verformung ist nun so definiert, daß die erzeugte gemischte Form immer dieselbe ist. Somit erhalten wir in der Tat einen zwei-dimensionalen Raum von Formen. Bei $k + 1$ Basisformen erhalten wir einen k -dimensionalen Teil des unendlich-dimensionalen Formenraumes der bereits bei Riemann eine Rolle in der Einführung von abstrakten Mannigfaltigkeiten gespielt hat [Riemann, 1868].

Literatur

- [Bollobás, 1968] Bollobás, B. (1968) Area of the union of disks. *Elem. Math.* 23, 60–61.
- [Boltjanskij und Efremovič, 1986] Boltjanskij, V. G. und Efremovič, V. A. (1986). *Anschauliche kombinatorische Topologie*. Vieweg, Braunschweig/Wiesbaden, Deutschland.
- [Casati und Varzi, 1994] Casati R. und Varzi, A. C. (1994). *Holes and Other Superficialities*. MIT Press, Cambridge, Massachusetts.
- [Cheng H.-L. et al., 1998] Cheng, H.-L., Edelsbrunner H. und Fu P. (1998). Shape space from deformation. In *Proc. 6th Pacific Conf. Comput. Graphics Appl.*, 1998, 104–113.
- [Cheng S.-W. et al., 1998] Cheng, S.-W., Edelsbrunner, H., Fu P. und Lam, K.-P. (1998). Design and analysis of planar shape deformation. In *Proc. 14th Ann. Sympos. Comput. Geom.*, 1998, 29–38.
- [Creighton, 1994] Creighton, T. E. (1994). *Proteins. Structures and Molecular Principles*. Freeman, New York.
- [Delaunay, 1934] Delaunay, B. (1934). Sur la sphère vide. *Izv. Akad. Nauk SSSR, Otdelenie Matematicheskii i Estestvennyka Nauk* 7, 793–800.
- [Edelsbrunner, 1995] Edelsbrunner, H. (1995). The union of balls and its dual shape. *Discrete Comput. Geom.* 13, 415–440.
- [Edelsbrunner, 1999] Edelsbrunner, H. (1999). Deformable smooth surface design. *Discrete Comput. Geom.* 21, 87–115.
- [Edelsbrunner et al., 1998] Edelsbrunner, H., Facello M. A. und Liang, J. (1998). On the definition and the construction of pockets in macromolecules. *Discrete Appl. Math.* 88, 83–102.
- [Edelsbrunner et al., 1983] Edelsbrunner, H., Kirkpatrick, D. G. und Seidel, R. (1983). On the shape of a set of points in the plane. *IEEE Trans. Inform. Theory* 29, 551–559.
- [Galilei, 1960] Galilei (1960). The assayer. In *The Controversy on the Comets of 1618*, Drake, S. und O'Malley, C. D., Hrsg., Univ. Pennsylv. Press, Philadelphia, 151–336.

- [Klee und Wagon, 1991] Klee, V. und Wagon, S. (1991). *Old and New Unsolved Problems in Plane Geometry and Number Theory*. Math. Assoc. Amer., Dolcini Math. Exp. 11.
- [Kneser, 1955] Kneser M. (1955). Einige Bemerkungen über das Minkowskische Flächenmass. *Arch. Math.* 6, 382–390.
- [Kratky, 1978] Kratky, K. W. (1978). The area of intersection of n equal circular disks. *J. Phys. A: Math. Gen.* 11, 1017–1024.
- [Lee und Richards, 1971] Lee, B. und Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* 55, 379–400.
- [Lewandowski et al., 1983] Lewandowski, J. L., Liu J. W. S. und Liu, C. L. SS/TDMA time slot assignment with restricted switching modes. *IEEE Transactions on Communications*, COM-23, 149–154.
- [Liang et al., 1998] Liang, J., Edelsbrunner, H. und Woodward, C. (1998). Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Science* 7, 1884–1897.
- [Milnor, 1963] Milnor, J. (1963). *Morse Theory*. Annals Math. Studies, Princeton Univ. Press.
- [Naiman und Wynn, 1992] Naiman, D. Q. und Wynn, H. P. (1992). Inclusion-exclusion-Bonferroni identities and inequalities for discrete tube-like problems via Euler characteristics. *Ann. Statist.* 20, 43–76.
- [Richards, 1977] Richards, F. M. (1977). Areas, volumes, packing, and protein structures. *Ann. Rev. Biophys. Bioeng.* 6, 151–176.
- [Riemann, 1868] Riemann, B. (1868). *Über die Hypothesen, welche der Geometrie zugrunde liegen*. Abh. Königl. Ges. Wiss. Göttingen 13.
- [Rose, 1996] Rose, G. D. (1996). No assembly required. *The Sciences* 36, 26–31.
- [Thue Poulsen, 1954] Thue Poulsen, E. (1954). Problem 10. *Math. Scand.* 2, 346.
- [Voronoi, 1907] Voronoi, G. (1907, 1908). Nouvelles applications des paramètres continus à la théorie des formes quadratiques. *J. Reine Angew. Math.* 133 (1907), 97–178, und 134 (1908), 198–287.

Irrfahrten

Wolfgang Woess

Einleitung

Wie im Einladungstext zum Symposium erklärt wird, hat die Kunstwelt seit jeher ein starkes Interesse an den formal-ästhetischen Aspekten der Mathematik. (Hier ist anzumerken, daß der Mathematiker dabei mitunter den Eindruck hat, daß hier vom Künstler Dinge "mystifiziert" werden, die für ihn selbst ganz und gar nicht geheimnisvoll sind.)

Umgekehrt ist in der Mentalität des Mathematikers der ästhetische Aspekt aus seiner Forschung nicht wegzudenken. Natürlich wird die persönliche Mischung von Ästhetik und Zweckdenken für jeden von uns verschieden sein.

Im Bild, das sich die *Öffentlichkeit* von der Mathematik macht, hat aber der künstlerische Aspekt kaum Platz. Ich erinnere mich, wie eine befreundete Künstlerin (Musikerin) die Existenz eines solchen künstlerischen Aspektes strikt abgelehnt hat: die Mathematik sei zu rigide, um spontane Intuition - wesentlicher Bestandteil der künstlerischen Kreativität - zuzulassen.

Natürlich ist das falsch. Aber hier handelt es sich um ein intrinsisches Problem der Mathematik: es ist unheimlich schwer, mit Nichtmathematikern über die Inhalte der eigenen Forschung zu kommunizieren, den meisten bleibt es unverständlich, was ein Mathematiker forschen kann und daß es überhaupt noch etwas zu forschen gibt.

Viele haben nur die negative Erinnerung aus der Schule behalten, wo sie die Mathematik als mehr oder weniger unverständliche Sammlung von seit eh und je bekannten Formeln erlebt haben, oder als Disziplinierungsmittel unbotmäßiger Schüler.

Es handelt sich hier also um eine Kunst, die nur ein begrenztes Publikum hat. Es haben auch andere moderne Kunstformen, z.B. moderne Musik oder auch Medienkunst, nur ein begrenztes Publikum, und hier wie da gibt man sich letzten Endes mit der Anerkennung der Fachwelt als Maßstab des Erfolges zufrieden.

Hierzu kommt, daß es heute dem Mathematiker selbst eigentlich nicht er-

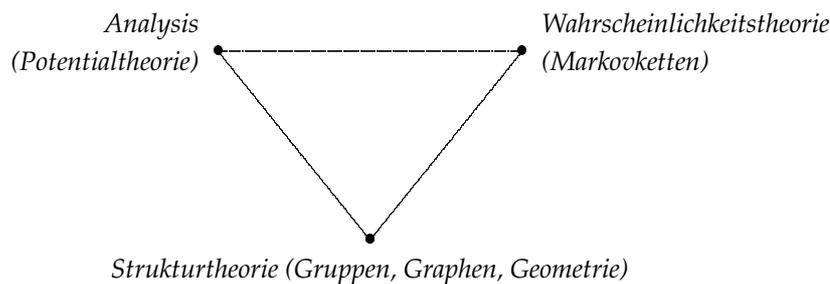
laubt ist, den künstlerischen Aspekt seiner Arbeit offen zuzulassen. Die Universitäten und die Forschung sind fast ausschließlich auf Aspekte der unmittelbaren Wirtschaftlichkeit reduziert. Wobei man den Eindruck hat, daß die "Mäzene" der Wissenschaft - also Regierung und Politiker - in Zeitspannen von der Dauer einer Legislaturperiode denken und den materiellen Profit der Forschung entsprechend rasch erwarten.

Die Grundlagenfächer geraten da in die Defensive, und nur selten werden Mathematiker sich ein öffentliches Bekenntnis zum formal-ästhetischen Aspekt ihrer Forschung erlauben.

Thema dieser Veranstaltung ist aber sicher nicht die industrielle Verwertbarkeit meines Faches. Ich möchte in meinen Ausführungen aus meinem Arbeitsgebiet erzählen; dabei werde ich nicht vorrangig auf die neuesten Resultate meiner eigenen Forschung eingehen, sondern versuchen, ausgehend von einfachen Beispielen einen Faden zu spinnen, um neben rein mathematischen Inhalten ansatzweise auch Denkweise und Rezeption zu erläutern. Ganz bewußt werde ich in meiner Darstellung die genannte "industrielle Verwertbarkeit" komplett ausklammern.

Irrfahrten in endlichen und unendlichen Straßen

Bei den *Irrfahrten* denken manche wahrscheinlich zuerst an Homer und Odysseus. Hier geht es aber um eine gewisse Art von Zufallsprozessen; mein Arbeitsgebiet bewegt sich in einem Dreieck:



Beispiel 1: Irrfahrt des Betrunkenen. Ein Betrunkener will aus einem Gasthaus heimkehren. Das Gasthaus liegt an einer geraden Straße; an einem Ende befindet sich ein See, am anderen das Wohnhaus des Betrunkenen. Der Betrunkene erinnert sich nicht an die richtige Richtung. Er macht einen Schritt nach links oder rechts mit gleicher Wahrscheinlichkeit ($1/2$), dann hält er inne. Ohne sich zu erinnern, woher er beim letzten Schritt gekommen war,

irrt er in dieser Weise weiter: einen Schritt nach rechts oder links, mit gleicher Wahrscheinlichkeit $1/2$. Erreicht er den See, so fällt er hinein und ertrinkt; kommt er zu Hause an, so bleibt er dort und schläft seinen Rausch aus. Frage: mit welcher Wahrscheinlichkeit ertrinkt der Betrunkene, und mit welcher Wahrscheinlichkeit erreicht er sein Haus?

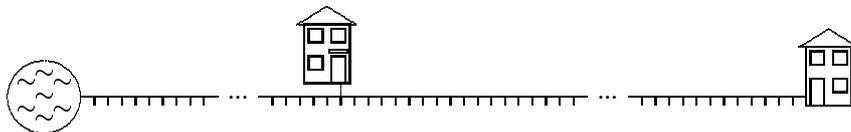


Abbildung 6.1.

Zur Berechnung formalisieren wir: wir nehmen an, daß jeder Schritt des Betrunkenen gleich lang ist. Sein Haus ist vom See n Schritte entfernt, das Gasthaus k Schritte. Die möglichen Positionen des Betrunkenen entsprechen den Zahlen $0, 1, \dots, n - 1, n$ (der Abstand vom See, in Schritten). Befindet sich der Betrunkene in Position i (zwischen 1 und $n - 1$), so führt sein nächster Schritt mit Wahrscheinlichkeit $1/2$ nach $i - 1$, und mit Wahrscheinlichkeit $1/2$ nach $i + 1$. (Man kann sich vorstellen, daß er vor jedem Schritt eine Münze wirft, um über die Richtung zu entscheiden.) Die Positionen 0 (See) und n (Haus) sind sogenannte absorbierende Zustände: einmal dort angelangt, ist die Wahrscheinlichkeit dort zu bleiben gleich 1 .

Wir schreiben $w(i)$ für die Wahrscheinlichkeit, daß der Betrunkene, ausgehend von Position i , jemals nach Hause kommt. Dann ist $w(0) = 0$, da 0 für den absorbierenden Zustand "See" steht, von dem aus es unmöglich ist (Wahrscheinlichkeit 0), das Haus zu erreichen. Weiters ist $w(n) = 1$ (die 1 steht für "mit Sicherheit"), da der Betrunkene dann ja schon zu Hause ist.

Für i zwischen 1 und $n - 1$ führt der erste Schritt jeweils mit Wahrscheinlichkeit $1/2$ nach $i - 1$ oder $i + 1$, von wo aus das Haus dann irgendwann erreicht werden soll. Dies drückt sich durch die Beziehung

$$w(i) = \frac{1}{2}w(i - 1) + \frac{1}{2}w(i + 1).$$

Diese Gleichung nennt man eine lineare Rekursion; mit den Anfangswerten $w(0) = 0$ und $w(n) = 1$ ist sie leicht zu lösen, und man findet $w(i) = i/n$. Das Gasthaus steht an der Position k , also ist

$$W[\text{nach hause}] = \frac{k}{n}, \quad \text{und analog} \quad W[\text{ertrinken}] = 1 - \frac{k}{n}.$$

Beispiel 2: Irrfahrt entlang einer unendlich langen Straße. Wir lassen nun den See und das Haus unendlich weit in die Ferne rücken, d.h., verschwinden, sodaß wir es mit einer unendlich langen Strasse zu tun haben. Die Positionen werden nun mit den (negativen und positiven) ganzen Zahlen beschrieben, der Betrunkene startet in einem Punkt (z.B. in der Position 0). Die "Spielregel" ist dieselbe wie vorhin: wo immer der Betrunkene gerade ist, bleibt es dem Zufall überlassen, ob er den nächsten Schritt nach links oder rechts macht (jeweils mit Wahrscheinlichkeit $1/2$).

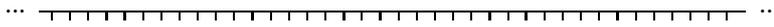


Abbildung 6.2.

Bevor ich näher darauf eingehe, hier eine Anmerkung zum *Unendlichen*. Vor längerer Zeit habe ich mich mit einem Freund (angehender Arzt) über meine Arbeit unterhalten. Ich erklärte ihm, daß ich mich mit Irrfahrten auf unendlichen Strukturen beschäftige. Er war zunächst sehr fasziniert, dann aber auch enttäuscht, weil er sich vorgestellt hatte, daß meine Arbeit in allerhand philosophischen Spekulationen über das Unendliche bestehe. Solche Begriffe sind aber für uns nicht Gegenstand philosophischer Spekulation, sondern durch eine sinnvolle Axiomatik gegeben, werden in der Arbeit vorausgesetzt und sind sozusagen tägliches Brot des Mathematikers.

Ähnlich erinnere ich mich, wie ich als 12-jähriger von der Frage nach der Existenz einer - physisch existierenden - vierten Dimension fasziniert war. Für den Mathematiker ist das Rechnen, das Arbeiten in 4 oder 5 Dimensionen genauso natürlich wie in 2 oder 3. Es besteht keine Notwendigkeit zur Spekulation hierüber, im Denkgebäude haben höhere Dimensionen ihren selbstverständlichen Platz. Aber natürlich handelt es sich dabei nicht um eine 4. Dimension im Sinne einer physikalisch zu erforschenden Realität.

Zurück zur Irrfahrt entlang der unendlichen Straße. Dies ist ein Zufallsprozeß: wir haben es mit einer zeitlichen Abfolge von zufälligen Ereignissen (den einzelnen Schritten) zu tun, die durch ein gewisses "Gesetz" beschrieben werden (jeweils mit Wahrscheinlichkeit $1/2$ nach links oder rechts; wir könnten uns auch ein anderes Gesetz vorstellen, z.B. mit Wahrscheinlichkeit $1/3$ nach rechts und $2/3$ nach links). Wir schreiben Z_n für die zufällige Position zum Zeitpunkt n ; die Zeit ist hier diskret, also in ganzen Einheiten, eine Zeiteinheit ist ein Schritt des Irrfahrers.

Im Vergleich mit Beispiel 1 werden in der unendlichen Straße die Frage-

stellungen ganz anders! See und Haus sind verschwunden, die Umgebung ist sozusagen monotoner. Die erste typische Frage betrifft die Rückkehr zum Ausgangspunkt. Man nennt die Irrfahrt *rekurrent*, wenn es sicher ist, daß der Betrunkene irgendeinmal zum Ausgangspunkt zurückkehrt. "Sicher" bedeutet, daß die Wahrscheinlichkeit zurückzukehren gleich 1 ist. Andernfalls nennt man den Zufallsprozeß *transient*.

Bevor wir beantworten, ob die Irrfahrt entlang der unendlichen Strasse rekurrent oder transient ist, wollen wir das Problem noch auf andere Weise formulieren.

Wenn es sicher ist, daß der Irrfahrer einmal zum Ausgangspunkt zurückkehrt, so ist es auch sicher, daß er ein zweites, drittes, ... Mal wiederkehrt. Wir sind also sicher, unendlich oft zum Ausgangspunkt zurückzukehren:

$$W[Z_n = 0 \text{ unendlich oft}] = 1.$$

Damit gleichwertig ist auch, daß im Verlauf der Zeit die mittlere Anzahl der Besuche im Ausgangspunkt unendlich groß wird. Das heißt, wir betrachten die Zufallsgröße N , deren Wert die Anzahl jener Zeitpunkte n ist, für die $Z_n = 0$ (die Ausgangsposition). N ist eine ganze Zahl oder $+\infty$. Der Mittelwert oder Erwartungswert $E(N)$ ist die Summe über alle Ausdrücke $n \cdot W[N = n]$. Er stimmt überein mit dem Ausdruck

$$\sum_{n=0}^{\infty} W[Z_n = 0].$$

Im transienten Fall ist hingegen die Wahrscheinlichkeit, (mindestens) einmal zurückzukehren eine Zahl $p < 1$, die Wahrscheinlichkeit, (mindestens) zweimal zurückzukehren $\leq p^2$, die Wahrscheinlichkeit, (mindestens) n mal zurückzukehren $\leq p^n$, also im Grenzwert

$$W[Z_n = 0 \text{ unendlich oft}] = 0;$$

weitere ist im transienten Fall der Erwartungswert $E(N)$ eine endliche Zahl.

Dies können wir nun zur Beantwortung der Frage nach Rekurrenz oder Transienz beantworten. Es ist eine einfache kombinatorische Aufgabe nachzuweisen, daß

$$W[Z_{2n} = 0] = \frac{1}{4^n} \binom{2n}{n} \sim \frac{1}{\sqrt{\pi n}}.$$

(Zu einem ungeraden Zeitpunkt kann man nicht zum Ausgangspunkt zurückkehren.) Das Symbol \sim bedeutet hier "hat die Größenordnung von"

(der Quotient zwischen linker und rechter Seite strebt gegen 1, wenn $n \rightarrow \infty$.) Die Reihe $\sum_n \frac{1}{\sqrt{\pi n}}$ divergiert ($= \infty$), also ist die Irrfahrt rekurrent.

Beispiel 3: zweidimensionales Straßennetz. Die Beispiele 1 und 2 wurden schon von *Lord Rayleigh* im letzten Jahrhundert studiert. Wir spinnen nun die Gedanken weiter, so wie es der ungarische Mathematiker *Georg Pólya* in einer Arbeit tat, die 1921 in den *Mathematischen Annalen* unter dem schönen Titel "Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Straßennetz" erschien.

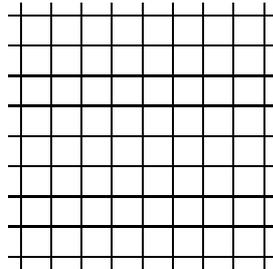


Abbildung 6.3.

Wir stellen uns ein unendliches, rechtwinkeliges, zweidimensionales Straßennetz (Gitter) vor. In jedem Kreuzungspunkt sind die Übergangswahrscheinlichkeiten $1/4$ in jede der vier Richtungen (Nord, Süd, Ost, West). Entsprechend der zufällig gewählten Richtung geht der Irrfahrer bis zur nächsten Kreuzung, wo er dann wieder über den nächsten, zufälligen Schritt entscheidet (z.B. indem er einen Tetraeder wirft, oder zwei Münzen).

Man findet leicht

$$W[Z_{2n} = 0] \sim \frac{1}{\pi n}.$$

Die Reihe $\sum_{n=1}^{\infty} \frac{1}{\pi n}$ divergiert. Die Irrfahrt ist rekurrent.

Wie sieht es in höheren Dimensionen aus?

Beispiel 4: d-dimensionales Straßennetz. Im dreidimensionalen Straßennetz hat jeder Kreuzungspunkt 6 Nachbarpunkte (Nord, Süd, Ost, West, Oben, Unten); im d -dimensionalen Straßennetz (oder Gitter) hat jeder Punkt $2d$ Nachbarn. In jedem Kreuzungspunkt wählt der Irrfahrer mit Wahrscheinlichkeit $1/2d$ einen der Nachbarpunkte aus; dorthin führt der nächste Schritt, usw. Wir fragen wieder nach der Rückkehrwahrscheinlichkeit. Die Antwort gibt

der folgende Satz [Pólya, 1921]:

Im d -dimensionalen Straßennetz gilt

$$W[Z_{2n} = 0] \sim \frac{C_d}{\sqrt{n^d}}$$

Dabei ist C_d eine von der Dimension d abhängige Konstante. Für $d \geq 3$ ist daher

$$\sum_{n=0}^{\infty} W[Z_{2n} = 0] < \infty,$$

die Irrfahrt ist *transient*! Wie wir wissen, ist dies damit gleichbedeutend, daß

$$W[\text{keine Rückkehr}] > 0 \quad \text{und} \quad W[Z_n = 0 \text{ unendlich oft}] = 0.$$

Weiters damit gleichwertig: für jede endliche Menge A von Kreuzungspunkten im Straßennetz (d -dim. Zahlengitter) ist

$$W[Z_n \text{ besucht } A \text{ nur endlich oft}] = 1$$

Also in verliert sich der Zufallspunkt Z_n für zunehmendes n in gewissem Sinne in den unendlichen Weiten des Raumes.

Wir sehen einen drastischen Unterschied zwischen Dimension 2 und Dimension 3: im Zweidimensionalen ergeht es dem Irrfahrer ähnlich wie dem Verdurstenden in der Wüste, der nach einigem Herumirren auf Fußspuren stößt und hofft, die Rettung zu finden; in Wirklichkeit hat er aber seine eigenen Spuren wiedergefunden und irrt im Kreis. Im Dreidimensionalen hingegen verliert sich der Irrfahrer zu den "Grenzen" des Alls.

Damit ist für den Mathematiker zwar rein rechnerisch ein Problem gelöst, aber damit allein gibt er sich nicht zufrieden. Warum ist das so? Wie kann man ein besseres Verständnis des Phänomens erlangen? – Man *verallgemeinert*, sucht, das Problem in einen größeren Zusammenhang zu stellen, entwirft ein weiterreichendes Modell, in dem das zwei- und dreidimensionale Gitter als Spezialfälle vorkommen. Man will sozusagen zurücktreten und die Frage aus einer weiteren Perspektive betrachten!

Irrfahrten in unendlichen Graphen

Wir wollen uns nun kompliziertere unendliche Straßennetze vorstellen. Ein Straßennetz beschreiben wir als einen *Graphen*, also eine Menge X von Punkten oder *Knoten* (die "Kreuzungen") und Verbindungsstücken zwischen Punkten, den Kanten des Graphen. Jede Kante hat zwei Endpunkte,

die dann als *Nachbarn* bezeichnet werden. Der Graph soll *zusammenhängend* sein: für je zwei Punkte gibt es einen Weg entlang von Kanten, der die beiden verbindet. Natürlich kann es auch mehr als einen Weg geben. Außerdem wird hier nur von *lokalendlichen* Graphen die Rede sein: von jedem Punkt x gehen nur endlich viele Kanten aus; deren Anzahl $\text{deg}(x)$ ist der *Grad* von x .

Der Zufallsprozeß, den wir untersuchen wollen, ist wieder die sogenannte *einfache Irrfahrt* auf (oder in) dem Graphen X . Die Übergangswahrscheinlichkeiten sind

$$p(x, y) = 1/\text{deg}(x), \text{ falls } y \text{ zu } x \text{ benachbart ist, andernfalls } p(x, y) = 0.$$

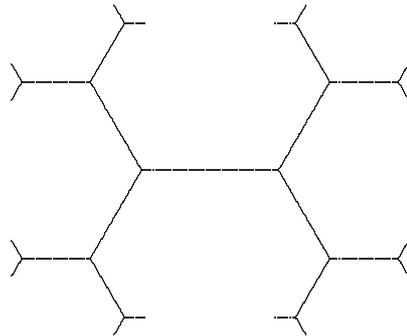


Abbildung 6.4

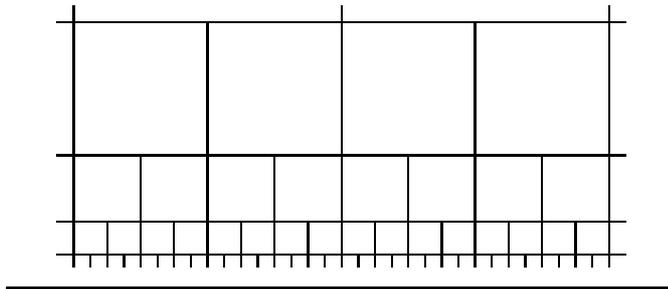


Abbildung 6.5

Befindet sich also der Irrfahrer zu irgendeinem Zeitpunkt im Knoten x , so wählt er nach dem Zufallsprinzip (Gleichverteilung) einen Nachbarn von x aus (mit Wahrscheinlichkeit $1/\text{deg}(x)$), zu dem dann der nächste Schritt führt.

Wir bezeichnen wieder mit Z_n die zufällige Position zum Zeitpunkt n . Die generelle Fragestellung ist dann: was kann man über den Zusammenhang zwischen der Struktur des Graphen X und dem zeitlichen Verhalten der Irrfahrt Z_n aussagen?

Abbildungen 6.4 und 6.5 sind zwei Beispiele von unendlichen Graphen, die eine andere Struktur als die d -dimensionalen Gitter haben.

Irrfahrten, elektrische Netzwerke und Rohrsysteme

Wir schreiben $E(X)$ für die Kantenmenge unseres Graphen X . Wir interpretieren nun jede Kante e in $E(X)$ nicht als Straßenstück, sondern als Leiterkabel mit elektrischem Widerstand 1 Ohm. Wenn wir wollen, können wir uns die Kante auch als 1 Meter langes Rohr vorstellen, dessen Querschnitt "genormt" ist (z.B. 10cm). An den Knoten von X sind die Kabel, bzw. Rohre verbunden. Wir wollen nun Strom durch das elektrische Netzwerk, beziehungsweise Wasser (oder besser eine "ideale" nicht komprimierbare Flüssigkeit) durch das Rohrsystem schicken.

Um einen *Fluß* im Netzwerk zu formalisieren, legen wir zunächst für jede Kante e eine Orientierung fest, sodaß wir von einem Anfangsknoten e^- und einem Endknoten e^+ in X sprechen können. Ein Fluß ist dann eine reellwertige Funktion $e \mapsto f(e)$ auf $E(X)$; unsere Interpretation ist die folgende: Das Rohrsystem ist mit Wasser gefüllt. $|f(e)|$ Liter pro Sekunde fließen von e^- nach e^+ , wenn $f(e) > 0$, bzw. von e^+ nach e^- , wenn $f(e) < 0$.

Wir sprechen von einem Fluß von $a \in X$ nach $b \in X$, wenn mit einer konstanten Rate von R Litern pro Sekunde ($R > 0$) beim Knoten a Wasser in das (schon vorher gefüllte) Rohrsystem hineingepumpt und beim Knoten b wieder abgezapft wird (in der elektrischen Interpretation wird der Pluspol einer Batterie in a und der Minuspol in b angelegt). Um wirklich von einem Fluß sprechen zu können, muß in jedem Knoten soviel Wasser ankommen wie wegfließt; dabei ist das Rohrsystem überall geschlossen, mit Ausnahme der Knoten a (Eingang) und b (Ausgang). Formal heißt das, daß für jeden Knoten x des Rohrsystems

$$\sum_{e:e^+=x} f(e) - \sum_{e:e^-=x} f(e) = \begin{cases} -R, & \text{wenn } x = a \\ 0, & \text{wenn } x \neq a, b \\ R, & \text{wenn } x = b \end{cases}$$

gilt: dies ist das *erste Kirchhoffsche Gesetz*. Die *Energie* des Flusses f ist die Zahl

$$En(f) = \sum_{e \in E(X)} f(e)^2$$

Etwas schwieriger Platz in unserer Vorstellungswelt (soferne wir das Unendliche schon akzeptiert haben) hat ein Fluß von $a \in X$ nach Unendlich, bei dem mit konstanter Rate von R Litern pro Sekunde ($R > 0$) Wasser beim Knoten a in das Rohrsystem hineingepumpt wird, aber an keinem anderen Knoten abgezapft wird (also "nach ∞ fließt"). Formal bedeutet das

$$\sum_{e: e^+ = x} f(e) - \sum_{e: e^- = x} f(e) = \begin{cases} -R, & \text{wenn } x = a \\ 0, & \text{wenn } x \neq a. \end{cases}$$

Kann ein Fluß von a nach ∞ mit endlicher Energie existieren? In einem endlichen Rohrsystem bedarf es natürlich einer unendlichen Anstrengung, um Flüssigkeit in die bereits gefüllten Rohre hineinzupumpen, ohne sie anderswo abzuzapfen. Bei unendlichen Rohrsystemen ist die (mathematische) Möglichkeit, dies zu tun, gerade ein Kriterium für die Transienz der Irrfahrt.

Satz [Lyons, 1983]: Die Irrfahrt im Graphen X ist transient genau dann wenn ein Fluß von $a \in X$ nach Unendlich mit positiver Rate und endlicher Energie existiert.

Dieser Satz gibt eine Interpretation des Problems Transienz / Rekurrenz für Irrfahrten in Graphen, die für den Mathematiker *schön* ist. Das "physikalische" Kriterium (in der Tat sind Irrfahrten in der theoretischen Physik ein aktuelles Forschungsthema) spricht die Vorstellung an, läßt sich oft zur Lösung des Problems verwenden, beleuchtet das Pólyasche Phänomen von anderer Seite und stellt es in eine weitere Perspektive.

Transienz bedeutet also, das man in das gefüllte Rohrsystem immer weiter Wasser hineinpumpen kann, und zwar mit endlichem Energieaufwand pro Zeiteinheit. Der Graph ist so stark verzweigt, daß sich das Wasser fortwährend im Unendlichen verlieren kann.

Wir wollen nun sehen, wie man das Flußkriterium anwenden kann, um Rekurrenz, bzw. Transienz der Irrfahrt in verschiedenen Graphen nachzuweisen. Will man Rekurrenz zeigen, so muß man nachprüfen, daß jeder Fluß von einem ausgewählten Knoten nach Unendlich auch unendliche Energie haben muß; wir können dabei annehmen, daß die Rate $R = 1$ ist. Um Transienz zu zeigen, genügt es, ein einziges Beispiel für einen Fluß von

einem geeigneten Knoten nach Unendlich zu "erfinden", der endliche Energie hat.

Beispiel 5: Irrfahrt entlang der unendlichen Straße. Hier kennen wir schon die Antwort. Als Ausgangspunkt wählen wir die Position 0, und die Kanten werden so orientiert, daß sie von 0 wegzeigen. Sei f ein Fluß von 0 nach ∞ mit Rate $R = 1$. Wenn t der Wert des Flusses von 0 zur Position 1 ist, so muß der Fluß von 0 nach -1 den Wert $1 - t$ haben. Da es keine Verzweigungen gibt, muß entlang den Kanten von 1 nach 2, 2 nach 3, usw., die Quantität t weiterfließen. Genauso fließt von -1 nach -2 , -2 nach -3 , usw., die Quantität $(1 - t)$. Die Energie von f ist also

$$t^2 + (1 - t)^2 + t^2 + (1 - t)^2 + t^2 + (1 - t)^2 + \dots = \infty,$$

da nicht gleichzeitig $t = 0$ und $1 - t = 0$ sein kann. Jeder Fluß von 0 nach Unendlich hat also unendliche Energie, und die Irrfahrt ist rekurrent.

Beispiel 6: zweidimensionales Straßennetz. Das ist natürlich etwas komplizierter als der eindimensionale Fall des letzten Beispiels, und bedarf einiger Vorbereitungen. In einem zusammenhängenden Graphen X kann man eine Distanz definieren: sind x und y zwei Knoten, so können wir für jeden Weg von x nach y die Kanten zählen, dies ist die Weglänge. Die Distanz von x zu y ist die kleinste unter diesen Weglängen.

Im zweidimensionalen Gitter wählen wir nun als Ausgangsknoten den Nullpunkt. Mit S_n bezeichnen wir die Menge der Punkte in Distanz n vom Nullpunkt, und E_n bezeichne die Kanten zwischen S_n und S_{n+1} . Wir orientieren diese Kanten "nach außen" (vom Nullpunkt wegweisend). Die Anzahl der Elemente in E_n ist $|E_n| = 8n + 4$. Nun sei f ein Fluß von 0 nach Unendlich mit Rate $R = 1$. Wegen dem ersten Kirchhoffschen Gesetz, das ja besagt, daß das hineingepumpte Wasser an keinem Knoten des Rohrsystems verloren gehen kann, muß auch der Gesamtfluß zwischen S_n und S_{n+1} den Wert 1 haben, also $\sum_{e \in E_n} f(e) = 1$.

Wir verwenden nun die Ungleichung von Cauchy-Schwarz: für beliebige reelle Zahlen a_1, \dots, a_k ist

$$\left(\sum_{i=1}^k a_i \right)^2 \leq k \cdot \sum_{i=1}^k a_i^2.$$

Für unseren Fluß f erhalten wir

$$\begin{aligned} En(f) &\geq \sum_{n=0}^{\infty} \sum_{e \in E_n} f(e)^2 \\ &\geq \sum_{n=0}^{\infty} \frac{1}{|E_n|} \left(\sum_{e \in E_n} f(e) \right)^2 \\ &= \sum_{n=0}^{\infty} \frac{1}{8n+4} \cdot 1 = \infty \end{aligned}$$

Dieses Argument läßt sich auf allgemeinere Graphen als das zweidimensionale Gitter anwenden: S_n und E_n können genauso definiert werden. Das Wachstum des Graphen spielt eine wichtige Rolle: wenn $|S_n| \leq C \cdot n$ für eine Konstante C und alle $n = 1, 2, \dots$, so ist die Irrfahrt rekurrent.

Beispiel 7: dreidimensionales Straßennetz. Das Argument von vorhin läßt sich nicht anwenden, da die Anzahl der Elemente von S_n hier die Größenordnung n^2 hat und die Reihe $\sum 1/n^2$ konvergiert.

Der folgende "Baum" kann in das dreidimensionale Straßennetz eingebettet werden [Gerl, 1984], [Doyle und Snell, 1984] (das ist allerdings nicht ganz einfach):

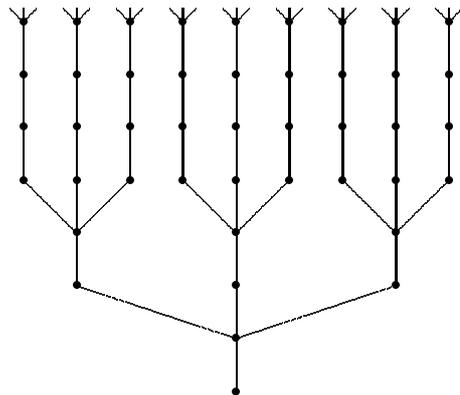


Abbildung 6.6

Wir orientieren die Kanten vom Wurzelknoten weg (in Abbildung 6.6 aufwärts) und wählen den möglichst einfachen Fluß: in jedem Verzweigungspunkt wird die von unten hereinkommende Wassermenge zu gleichen Teilen in die herausgehenden Rohre (Kanten) weitergeleitet. Die Energie dieses

Flusses vom Wurzelknoten nach Unendlich ist

$$En(f) = 1 + 6 \cdot \frac{1}{3^2} + 36 \cdot \frac{1}{9^2} + \dots = \sum_{n=0}^{\infty} \left(\frac{2}{3}\right)^n < \infty.$$

Dieser Fluß kann auch als Fluß im dreidimensionalen Gitter aufgefaßt werden: entlang der nicht vom Baum erfaßten Kanten fließt gar nichts, also die Quantität 0. Daher ist die Irrfahrt im dreidimensionalen Gitter transient.

Wir haben hier eine Folgerung aus dem Flußkriterium verwendet: ist die Irrfahrt in einem Graphen rekurrent, so ist auch die Irrfahrt in jedem Teilgraphen (einen Teilgraphen erhält man durch Weglassen von Knoten und / oder Kanten) rekurrent. Das ist zwar von Anfang an plausibel, aber nicht unmittelbar klar: die Übergangswahrscheinlichkeiten sind im Teilgraphen anders. Umgekehrt folgt aus der Transienz der Irrfahrt in einem Teilgraphen die Transienz der Irrfahrt im ursprünglichen Graphen.

Da das dreidimensionale Gitter ein Teilgraph aller höherdimensionalen Gitter ist, können wir auf diese Weise die Transienz der Irrfahrten in den letzteren folgern.

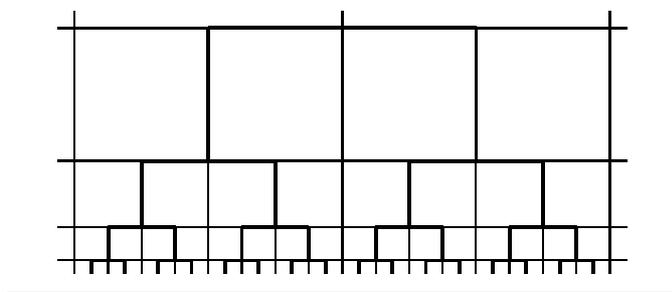


Abbildung 6.7

Beispiel 8: der reguläre Baum. Ein Baum ist ein Graph, in dem es keine Kreise gibt. Der reguläre Baum vom Grad q ist der Baum, in dem alle Knoten die gleiche Anzahl q von Nachbarn haben. Die Abbildung 6.4 weiter oben zeigt den homogenen Baum vom Grad 3.

Es ist ganz einfach zu sehen, daß die Irrfahrt im regulären Baum transient ist, wenn $q \geq 3$. Wir wählen einen Wurzelknoten a und orientieren alle Kanten von a weg. Den Fluß f definieren wir auf die einfachste Art: $1/q$ für die Kanten, die von a ausgehen, $1/q(q-1)$ für die von ihren Endpunkten

ausgehenden Kanten, usw., also $f(e) = 1/q(q-1)^{n-1}$ für Kanten e in E_n mit $n \geq 1$. Dann ist

$$En(f) = q \cdot \frac{1}{q^2} + q(q-1) \cdot \frac{1}{[q(q-1)]^2} + \dots + q(q-1)^{n-1} \cdot \frac{1}{[q(q-1)^{n-1}]^2} + \dots < \infty.$$

Auf ganz ähnliche Weise können wir die Transienz der Irrfahrt im Graphen der Abbildung 6.5 überlegen. Stellen wir zur Sicherheit klar, wie dieser Graph aufgebaut ist. Er füllt die obere Halbebene aus. Wir beginnen mit einem Quadrat der Seitenlänge 1, dessen linker unterer Eckpunkt die Koordinaten $(0, 1)$ hat. Links und rechts an dieses Quadrat schließen wir eine Abfolge unendlich vieler Quadrate der gleichen Form an. An diesen Streifen von Quadraten schließen wir oben und unten weitere Streifen von Quadraten an, und zwar oben mit der doppelten und unten mit der halben Seitenlänge. Wir fahren so fort, sodaß nach oben die Quadrate immer größer und nach unten immer kleiner werden.

Übrigens ist dieser Graph ein diskretes Modell der hyperbolischen (oder Poinaréschen) Ebene.

Es ist nun recht einfach, in diesen Graphen einen Baum einzubetten, für den man die Transienz der Irrfahrt ganz ähnlich wie für den regulären Baum sehen kann, siehe Abbildung 6.7.

Das Flußkriterium hat ein analytisches Analogon in der Differentialgeometrie, das schon länger bekannt ist (Kelvin - Nevanlinna - Royden); [Royden, 1952]. Davon inspiriert, wurde es für Irrfahrten (reversible Markovketten) 1983 von T. Lyons bewiesen, der mit dieser Arbeit (unter anderen) bekannt wurde.

Davor wurde es aber schon [Yamasaki, 1979] publiziert (nur für lokalendliche Graphen, wie hier), aber erstens gab Yamasaki keine probabilistische Interpretation und beschränkte sich auf eine sehr trockene, formale Präsentation, und zweitens publizierte er es in einer eher sekundären japanischer Fachzeitschrift. Dort blieb seine Arbeit bis 1988 unbemerkt. Ich schreibe mir den "Verdienst" zu, 1988 diese und andere Arbeiten von Yamasaki rein zufällig "entdeckt" und anderen Kollegen weitergegeben zu haben...

Schon viel früher hatte ein anderer englischer Mathematiker [Nash-Williams, 1959] den Zusammenhang zwischen Irrfahrten und elektrischen Netzwerken aufgezeigt und zum Studium des Polyaschen Phänomens verwendet. Lyons greift darauf zurück.

Die Arbeit von Nash-Williams blieb aber für lange Jahre mehr oder weniger unbeachtet. Die Gründe sind sein komplizierter Stil und die schwer

verständliche Notation, und das damals geringere Interesse der Fachwelt an dieser Art von Fragestellung. ("Die Zeit war noch nicht reif.")

Hier wäre ein längerer Diskurs möglich über das Thema, wie Mathematik von anderen Mathematikern rezipiert wird. Wie man sieht, spielen Zeitpunkt, Fachzeitschrift, Stil der Arbeit, und auch Zufall eine große Rolle! Zurück zu den Irrfahrten. Nun haben wir eine allgemeinere Theorie, und damit eine andere / bessere / neue / ... Optik, mit der wir das Phänomen der Rekurrenz / Transienz im 2-, bzw. 3-dimensionalen Raum sehen können. Daraus entstehen aber eher mehr neue Fragen als weniger. Zum Beispiel

1. Hat jeder transiente Graph einen transienten Teilbaum? (Antwort: nein, es gibt ein Gegenbeispiel [McGuinness, 1988])
2. Im d -dimensionalen Gitter haben wir die Asymptotik (Größenordnung, wenn $n \rightarrow \infty$) von $p^{(n)}(0,0) = W[Z_n = 0 \mid Z_0 = 0]$ verwendet, also der Wahrscheinlichkeit, nach n Schritten wieder am Ausgangspunkt 0 zu sein. Eine allgemeinere Frage, über die viel gearbeitet wurde und wird, ist: Wie hängt die Struktur des Graphen X mit der Asymptotik von $p^{(n)}(x,x)$ zusammen, wobei x ein Knoten von X ist?
3. Im transienten Fall strebt die Irrfahrt Z_n im Raum "nach Unendlich". Kann man mögliche "Grenzpunkte" im Unendlichen besser unterscheiden? Hier geht es um das Studium von Kompaktifizierungen, sogenannten harmonischen Funktionen, usw.
4. Wie kann man insbesondere einem Graphen in der Ebene (*planaren* Graphen) ansehen, ob die Irrfahrt in ihm rekurrent / transient ist?
usw.

Isoperimetrische Ungleichungen

Das Studium von *isoperimetrischen Ungleichungen* ermöglicht die Antwort auf einige der vorhin aufgeworfenen Fragen. Für einen Körper im d -dimensionalen Raum vergleicht eine solche Ungleichung sein Volumen mit seiner Oberfläche; die Kugel vom Radius r in d Dimensionen hat Volumen und Oberfläche von der Größenordnung r^d und r^{d-1} , also $\text{Volumen}^{1-1/d} \simeq \text{Oberfläche}$.

Für eine Menge A von Knoten eines Graphen X muß man erst die Begriffe "Volumen" und "Oberfläche" definieren. Das *Volumen* von A ist die

Summe über die Knotengrade in A : $\text{Vol}(A) = \sum_{x \in A} \text{deg}(x)$. Der Rand ∂A von A ist die Menge aller Kanten des Graphen, die aus A herausführen (also einen Endpunkt in A und den anderen außerhalb haben), und die Oberfläche von A ist die Anzahl dieser Kanten: $\text{Area}(\partial A) = |\partial A|$. Für eine Zahl d , $1 \leq d \leq \infty$, sagt man, daß der Graph X eine d -dimensionale isoperimetrische Ungleichung, in Symbolen (IS_d) , erfüllt, wenn eine Konstante $c > 0$ existiert, sodaß für jede endliche Menge $A \subset X$

$$\text{Area}(\partial A) \geq c \cdot \text{Vol}(A)^{1-1/d}.$$

Wenn speziell $d = \infty$ ($1/d = 0$), so spricht man von einer "starken isoperimetrischen Ungleichung" (IS): Volumen und Oberfläche sind von der gleichen Größenordnung.

Beispiele: Das d -dimensionale Gitter erfüllt (IS_d) , aber nicht $(IS_{d'})$ für $d' > d$. Der reguläre Baum (Abbildung 6.4) erfüllt (IS) . Hier sind einige Resultate, die den Zusammenhang zwischen isoperimetrischen Ungleichungen und dem Verhalten von Irrfahrten in Graphen erhellen:

Satz [Thomassen, 1992]: Falls X (IS_d) mit $d > 2$ erfüllt, so enthält X einen transienten Baum.

Unter einer (in gewissem Sinne kleinen) Einschränkung gibt es also doch eine positive Antwort auf die Frage (1) von oben.

Satz [Varopoulos, 1984]: Wenn X die Ungleichung (IS_d) erfüllt (wobei $1 \leq d < \infty$), so gilt für alle x, y in X

$$p^{(n)}(x, y) \leq C \cdot \text{deg}(y) \cdot \frac{1}{\sqrt{n^d}}.$$

Wir erinnern, daß $p^{(n)}(x, y)$ die Wahrscheinlichkeit ist, daß sich der Irrfahrer n Schritte nach dem Start in x im Knoten y befindet. Dieser sehr wichtige und schwierig zu beweisende Satz gibt eine Antwort auf die Frage (2).

Satz [Dodziuk, 1984], [Gerl, 1988]: X erfüllt (IS) genau dann wenn $\rho < 1$ existiert, sodaß für alle Knoten x in X

$$p^{(n)}(x, x) \leq \rho^n.$$

Dies impliziert natürlich Transienz der Irrfahrt, da dann $\sum_n p^{(n)}(x, x) < \infty$, also der Erwartungswert $E(N)$ endlich ist, wobei – wir erinnern – die Zufallsgröße N die Anzahl der Besuche in x bei Start in x ist.

Starke isoperimetrische Ungleichung für Pflasterungen in der Ebene

Sei \mathcal{O} offenes, einfach zusammenhängende Gebiet in der Ebene (also zusammenhängend und ohne Löcher), z.B. die ganze Ebene, die obere Halbebene oder eine offene Kreisscheibe.

Eine *Pflasterung* oder *Kachelung* von \mathcal{O} besteht aus einer Familie \mathcal{T} von abgeschlossenen topologischen Kreisscheiben (also Mengen, die stetig zu einer Kreisscheibe deformiert werden können; diese sind die "Pflastersteine", bzw. "Kacheln") mit paarweise fremdem Inneren, die zusammen das Gebiet \mathcal{O} bedecken. Wir nehmen zusätzlich an, daß jede beschränkte Menge in \mathcal{O} nur endlich viele Kacheln $T \in \mathcal{T}$ trifft.

Wir ordnen der Pflasterung den *Kantengraph* $X = X(\mathcal{T})$ zu. Das ist einfach der Graph, den wir sehen, wenn wir die Pflasterung anschauen: die Knoten sind die Punkte, wo sich drei oder mehr Kacheln treffen, die Kanten sind die Randstücke von Kacheln, die zwischen den Knoten liegen. Der Graph von Abbildung 6.5 ist der Kantengraph einer Kachelung der oberen Halbebene; die Kacheln (Quadrate) sind eigentlich Fünfecke, da auf jedem von ihnen 5 Knoten liegen.

Wie kann man dem Kantengraph X einer Pflasterung ansehen, ob die Irrfahrt transient ist? Man kann über (IS) entscheiden und erhält so eine hinreichende Bedingung.

Definition. Die charakteristische Zahl einer Kante $e \in E(X)$, bzw. eines Knotens $x \in X$, beziehungsweise einer Kachel $T \in \mathcal{T}$ ist

$$\begin{aligned}\phi(e) &= 1 - \sum_{x \in e} \frac{1}{\deg(x)} - \sum_{T: e \in E(T)} \frac{1}{|T|}, \\ \psi(x) &= \frac{\deg(x)}{2} - 1 - \sum_{T: x \in T} \frac{1}{|T|}, \\ \chi(T) &= \frac{|T|}{2} - 1 - \sum_{x: x \in T} \frac{1}{\deg(x)}.\end{aligned}$$

Dabei ist $|T|$ die Anzahl der Knoten von X , die auf T liegen. Weiters steht $E(T)$ für jene Kanten, die am Rand der Kachel T liegen, $\sum_{T: e \in E(T)}$ bedeutet z.B., daß über alle Kacheln summiert wird, auf denen die Kante e liegt, und $|T|$ steht für die Anzahl der Knoten von X die auf T liegen.

Satz [Woess, 1998]: Falls $\phi(e) > 0$ für alle Kanten $e \in E(X)$, oder $\psi(x) > 0$ für alle Punkte $x \in X$,

oder $\chi(T) > 0$ für alle Pflastersteine $T \in \mathcal{T}$,
so erfüllt $X(\mathcal{T})$ die starke isoperimetrische Ungleichung.

Der Beweis dieses Satzes verwendet die Eulersche Formel für endliche Graphen in der Ebene.

Beispiel 9. Für den Graphen aus Abbildung 6.5 ist $|T| = 5$ für jede Kachel T , und auf jedem T liegen 3 Knoten mit Grad 4 und 2 Knoten mit Grad 3. Daher ist $\chi(T) = \frac{5}{2} - 1 - \frac{3}{4} - \frac{2}{3} = \frac{1}{12}$ stets positiv. Auch haben wir $\psi(e) = \frac{1}{60}$ für jede Kante. Der Graph erfüllt die starke isoperimetrische Ungleichung, und die Irrfahrt ist transient, wie wir schon wissen. (Allerdings ist $\psi(x) = -\frac{1}{10}$ für die Knoten mit Grad 3.)

Beispiel 10. Das zweidimensionale Straßennetz (Gitter) ist der Kantengraph einer Pflasterung der ganzen Ebene mit Vierecken. In jedem Knoten treffen sich 4 Vierecke. Wir wissen, daß die Irrfahrt rekurrent ist, und in der Tat prüft man leicht nach, daß keine der drei Bedingungen des Satzes erfüllt ist.

Versuchen Sie nun, sich einen Graphen vorzustellen, wo sich in jedem Knoten nicht 4 Vierecke, sondern 8 Achtecke treffen. Diese müssen nicht unbedingt gleichmäßig sein, und die Kanten müssen nicht unbedingt gerade sein. Trotzdem ist es nicht ganz leicht, so diesen Graphen zu zeichnen. Man kann ihn als Kachelung einer offenen Kreisscheibe erhalten, die Kanten sind dann Kreisbögen, die zum Rand der Kreisscheibe rechtwinkelig sind. In der Nähe des Randes werden die Achtecke immer kleiner. Für jede Kachel (Achteck) T ist $\chi(T) = 2$. Also erfüllt der Graph die starke isoperimetrische Ungleichung. Es ist dies eine typische Kachelung der hyperbolischen Ebene, also der Kreisscheibe mit der Poincaréschen Metrik; die Kacheln sind in der hyperbolischen Geometrie alle kongruent.

Schlußbemerkungen

Die Kunstwelt und ihre "Körpersprache" sind mir nicht unvertraut. In den 80er Jahren hatte ich eine bescheidene "Zweitkarriere" in der künstlerischen Fotografie (Autorenfotografie), war Mitbegründer einer Fotogalerie in Salzburg und habe später einige Fotoausstellungen in Leoben organisiert. Bis kurz vor meiner Habilitation war ich stolz darauf, gleich viele Ausstellungen meiner Fotoarbeiten wie mathematische Publikationen vorweisen zu können. Erst mit dem Umzug nach Italien hat sich das Gewicht endgültig zugunsten der Mathematik verlagert.

Einer der Kunsttheoretiker hat in seinem Vortrag beim Symposium an-

gemerkt, wie überrascht er gewesen war, zu erfahren, daß sich Naturwissenschaftler beim Austausch ihrer Forschungsergebnisse bemühen, eine möglichst einfache Sprache zu verwenden. In diesem Sinne beabsichtige ich mit meinem Text nicht, rhetorische Brillanz zu versprühen und durch komplizierte Wort- und Satzwahl zu beeindrucken.

Ich habe vielmehr in diesen Ausführungen versucht, auf möglichst einfache Weise einige Aspekte eines aktuellen Forschungsgebietes zu *erklären*, ohne dabei auf schwierigere Beweise einzugehen. Beweise sind das Um und Auf der sogenannten "reinen" Mathematik, und haben ihre eigene Ästhetik. *Schönheit* bezieht sich dabei auf die Architektur des Gedankengebäudes.

Literatur

- [Dodziuk, 1984] Dodziuk, J. (1984). Difference equations, isoperimetric inequality, and transience of certain random walks. *Transactions of the American Mathematical Society*, 284, 787–794.
- [Gerl, 1984] Gerl, P. (1984). *Rekurrente und transiente Bäume*, Séminaire Lotharingien de Combinatoire (IRMA Strasbourg), 10, 80–87.
- [Gerl, 1988] Gerl, P. (1988). Random walks on graphs with a strong isoperimetric inequality. *Journal of Theoretical Probability*, 1, 171–187.
- [McGuinness, 1988] McGuinness, S. (1988). *Random Walks on Graphs and Digraphs*. Dissertation, University of Waterloo, Ontario.
- [Lyons, 1983] Lyons, T. (1983). A simple criterion for transience of a reversible Markov chain. *Annals of Probability*, 11, 393–402.
- [Pólya, 1921] Pólya, G. (1921). Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Straßennetz. *Mathematische Annalen*, 84, 149–160.
- [Royden, 1952] Royden, H. L. (1952). Harmonic functions on open Riemann surfaces. *Transactions of the American Mathematical Society*, 73, 40–94.
- [Thomassen, 1992] Thomassen, C. (1992). Trees, ends, and transience. In *Harmonic Analysis and Discrete Potential Theory* (Proceedings, Frascati 1991; Picardello, M. A., Hrsg.), Plenum, New York, 259–266.
- [Varopoulos, 1984] Varopoulos, N. Th. (1984). Isoperimetric inequalities and Markov chains. *Journal of Functional Analysis*, 63, 215–239.

- [Woess, 1998] Woess, W. (1998). A note on tilings and strong isoperimetric inequality. *Mathematical Proceedings of the Cambridge Philosophical Society*, 124, 385–393.
- [Woess, 1999] Woess W. (1999). *Random Walks on Infinite Graphs and Groups*. Cambridge Tracts in Mathematics, Cambridge University Press, in Druck.
- [Yamasaki, 1979] Yamasaki, M. (1979). Discrete potentials on an infinite network. *Memoirs of the Faculty of Science, Shimane University*, 13, 31–44.

Zuordnungsprobleme - ein Streifzug durch die kombinatorische Optimierung

Rainer E. Burkard

Das Heiratsproblem

In der ersten Hälfte unseres Jahrhunderts wurde folgendes Problem studiert: Wir betrachten eine Gruppe von n Mädchen und n jungen Herren. Jedes Mädchen hat zumindest einen jungen Mann zum Freund. Können n Paare heiraten, so daß jedes der Mädchen einen ihrer Freunde heiratet?

Dieses Problem läßt sich graphisch auf folgende Weise darstellen: Jedem Mädchen ordnen wir einen Punkt M_i , $i = 1, \dots, n$, zu. Jedem jungen Herren ordnen wir einen Punkt H_j , $j = 1, \dots, n$, zu. Ist H_j ein Freund von M_i , dann verbinden wir M_i und H_j durch eine Linie (vgl. Abbildung 7.1, Farbabbildungen: S. 150). Mathematisch gesehen handelt es sich in Abbildung 7.1 um einen *paaren Graphen* $(M, H; E)$ mit den *Knotenmengen* M und H und der *Kantenmenge* E . Im folgenden wollen wir immer annehmen, daß die Anzahl $|M|$ der Knoten in M gleich der Anzahl $|H|$ der Knoten in H ist. Jede Kante $e \in E$ verbindet einen Knoten $v \in M$ mit einem Knoten $w \in H$, in unserem Falle ein Mädchen mit einem jungen Herren. Wenn die Kante e die Knoten v und w verbindet, sagen wir, die Kante e *inzidiert* mit den Knoten v und w , bzw. *trifft* diese Knoten. Eine *Heirat* oder *Zuordnung* ist nun eine Teilmenge $Z \subseteq E$, so daß jeder Knoten in M und in H mit einer Kante in Z inzidiert, aber kein Knoten zwei oder mehr Kanten in Z trifft.

Im Jahr 1935 zeigte Philipp Hall [Hall, 1935], daß eine Heirat genau dann existiert, wenn jede Teilmenge von k Mädchen, $k = 1, 2, \dots, n$, zusammen mindestens k Freunde hat (*Hall-Bedingung*). In unserem Beispiel von Abbildung 7.1 (Farbabbildungen: S. 150) haben die beiden Mädchen M_2 und M_5 zusammen nur den einen Freund H_1 , den sie nicht beide gleichzeitig heiraten können. Daher existiert in diesem Beispiel keine Heirat. Würde jedoch M_2 auch mit H_5 befreundet sein, so könnten die folgenden Paare heiraten: M_1 heiratet H_4 , M_2 heiratet H_5 , M_3 heiratet H_2 , M_4 heiratet H_3 , und M_5 heiratet H_1 . Aus dieser Beschreibung einer Heirat erkennt man, daß sich Heiraten auch durch sogenannte *Permutationen* darstellen lassen. Statt M_1 heiratet H_4 schreiben wir kurz $1 \rightarrow 4$. In dieser Schreibweise erhält man weiters

$2 \rightarrow 5, 3 \rightarrow 2, 4 \rightarrow 3$ und $5 \rightarrow 1$. Wie man sieht, wird jedem Index i eines Mädchens M_i der Index j eines jungen Mannes H_j zugeordnet, so daß alle Indizes vorkommen, aber keiner doppelt. Eine solche Zuordnung nennt man *Permutation* der Menge $\{1, 2, \dots, 5\}$. Die obige Permutation läßt sich nun als $\varphi(1, 2, 3, 4, 5) = (4, 5, 2, 3, 1)$ schreiben. Jede Heirat entspricht einer Permutation und umgekehrt. Im Falle, daß alle jungen Leute miteinander befreundet sind, gibt es $n!$ (sprich: n Fakultät) verschiedene Heiraten (Permutationen). Dabei ist $1! = 1, 2! = 2, 3! = 2! \cdot 3 = 6, 4! = 3! \cdot 4 = 24$ und allgemein $n! = (n - 1)! \cdot n$.

So einfach die Hall-Bedingung auch formuliert werden kann, so schwierig ist sie in konkret gegebenen Freundschaftssystemen direkt zu überprüfen. Es müssen dazu alle Teilmengen der Knotenmenge M überprüft werden. Wenn M n Elemente hat, müssen also $2^n - 1$ viele Teilmengen überprüft werden. Man sagt, daß man einen exponentiellen Aufwand dafür treiben muß.

Ein interessantes, schnelles Verfahren, das mit großer Wahrscheinlichkeit feststellt, ob in einem gegebenen Freundschaftssystem eine Heirat existiert, basiert auf einem berühmten Ergebnis des Graphentheoretikers W. T. Tutte [Tutte, 1947]. Gibt es $m = |E|$ Freundschaftsbeziehungen (Kanten im zugehörigen paaren Graphen), so werden für jede dieser Freundschaftsbeziehungen zufällig zwei Zahlen zwischen 1 und $4m^2$ bestimmt. Aus diesen zufällig bestimmten Zahlen wird dann auf schnelle Weise der numerische Wert einer Größe (der Determinante der Tutte-Matrix) bestimmt. Diese Größe ist ungleich Null, wenn das Freundschaftssystem eine Heirat besitzt. Sollte sich bei der Rechnung herausstellen, daß die Größe den Wert 0 hat, wiederholt man die Berechnung mit anderen, zufällig erzeugten Zahlen. Wird auf diese Weise bei r Wiederholungen der Rechnung mit zufällig erzeugten Zahlen immer der Wert 0 ermittelt, so gibt es mit großer Wahrscheinlichkeit keine Heirat. Die Fehlerwahrscheinlichkeit p , daß es doch eine Heirat gibt, ist kleiner als $1/m^r$, wobei m die Anzahl der Kanten (Freundschaftsbeziehungen) ist und r die Anzahl angibt, wie oft die Berechnung der Determinante der Tutte-Matrix wiederholt wurde. Dieses Verfahren liefert nur die Aussage, ob es eine Heirat gibt, oder nicht. Es liefert jedoch keine Hinweise dafür, wer eventuell wen heiraten soll. Im nächsten Abschnitt werden wir ein Verfahren beschreiben, das es uns ermöglicht, auf rasche Weise möglichst viele Paare miteinander zu verheiraten.

Möglichst große Paarungen

Wir wollen in diesem Abschnitt zunächst der folgenden Frage nachgehen: Gegeben sei ein paarer Graph, also ein Freundschaftssystem. Was ist die *maximale* Anzahl von Paaren, die miteinander verheiratet werden können (möglichst große Paarung)? Im Beispiel der Abbildung 7.1 (Farbabbildungen: S. 150) war diese Anzahl 4. Es wird sich herausstellen, daß diese Aufgabe *effizient* lösbar ist. Dies heißt, daß die Anzahl der Rechenschritte nicht exponentiell mit der Anzahl der Mädchen und Herren wächst. Dazu formulieren wir unser Heiratsproblem in ein sogenanntes *maximales Flußproblem* um. Zu diesem Zweck erweitern wir den paarer Graphen durch zwei zusätzliche Knoten, eine *Quelle* q und eine *Senke* s . Nun verbinden wir q durch gerichtete Kanten (*Pfeile*) mit den Knoten in M . Ferner orientieren wir alle Kanten zwischen M und H so, daß die Pfeile von den Knoten in M zu den Knoten in H zeigen. Schließlich verbinden wir alle Knoten von H durch Pfeile mit der Senke s , wobei alle diese Pfeile zur Senke s zeigen (vgl. Abbildung 7.2, Farbabbildungen: S. 150).

Auf diese Weise haben wir ein *Netzwerk* gewonnen. In diesem Netzwerk werden wir einen *Fluß* einführen, der von der Quelle ausgeht und in die Senke einmündet. Dieser Fluß wird dadurch beschrieben, daß den Pfeilen Zahlen zugeordnet werden. In den Knoten von M und H müssen dann die folgenden *Flußerhaltungsgleichungen* erfüllt sein:

Die Summe der Kantenflüsse, die in einen Knoten hineinführen ist gleich der Summe der Kantenflüsse, die aus diesem Knoten herausführen.

Die Bedingung, daß eine Person nur *eine* andere Person heiraten kann, kann nun durch *Kapazitäten* auf den Pfeilen modelliert werden: In unserem Beispiel bekommt jeder Pfeil, der von der Quelle q zu einem Knoten $v \in M$ führt, die Kapazität $c(q, v) = 1$. Analog erhält jeder Pfeil, der von einem Knoten $w \in H$ zur Senke s führt, die Kapazität $c(w, s) = 1$. Für unseren Fluß auf den Pfeilen fordern wir, daß die einzelnen Werte auf den Pfeilen nicht die vorgegebenen Kantenkapazitäten überschreiten (Kapazitätsrestriktionen). Gesucht wird nun ein Fluß, sodaß die gesamte von der Quelle zur Senke gesandte Menge möglichst groß wird. Dieses Problem spielt in Anwendungen eine wichtige Rolle und wird als *maximales Flußproblem* bezeichnet. Um 1956 entwickelten L. R. Ford und D. R. Fulkerson [Ford und Fulkerson, 1956] den ersten Algorithmus zu seiner Lösung. Besonders in den letzten Jahren hat sich die Forschung intensiv mit diesem

Problem auseinandergesetzt und immer schnellere Lösungsverfahren dafür gefunden.

Wie hängt nun ein maximaler Fluß, d.h. eine Lösung des maximalen Flußproblems, mit einer größtmöglichen Paarung zusammen? Zunächst zeigt sich, daß man im betrachteten speziellen Netzwerk immer einen maximalen Fluß finden kann, der auf den Pfeilen entweder den Wert 0 oder 1 und keine Werte dazwischen annimmt. Um eine Paarung zu finden, braucht man nur die Pfeile betrachten, die von einem Knoten in M zu einem Knoten in H führen und den Flußwert 1 haben. Genau diese zeigen an, welches Mädchen in einer größtmöglichen Paarung welchen Herrn heiraten soll. Die Anzahl der möglichen Paare wird durch die Anzahl der Kanten angegeben, die von der Quelle ausgehen und den Flußwert 1 haben, siehe Abbildung 7.3 (Farbabbildungen: S. 151).

Hopcroft und Karp [Hopcroft und Karp, 1973] haben gezeigt, daß auf diese Weise eine größtmögliche Paarung in $K \cdot m \sqrt{n}$ Rechenschritten gefunden werden kann, wenn das Freundschaftssystem m Kanten hat. Dabei ist K eine fest vorgegebene Konstante, die nur vom gewählten Rechenverfahren, nicht aber von den Problemgrößen m und n abhängt. Natürlich kann man dieses Verfahren auch verwenden um festzustellen, ob es eine Heirat gibt. Man hat dazu ein maximales Flußproblem zu lösen, bei dem der maximale Flußwert, d.h. die Summe der Flüsse auf den Pfeilen, die aus der Quelle herausführen, gleich n ist. Nachdem diese Frage in $K \cdot m \sqrt{n}$ Schritten beantwortet werden kann, hat man auf diese Weise ein recht schnelles Verfahren zur Lösung des Heiratsproblems zur Hand. Man erhält dabei nicht nur die Antwort, ob es eine Heirat gibt, sondern auch den Vorschlag für eine tatsächliche solche Heirat. Wir werden nun im nächsten Abschnitt zeigen, daß derartige Heiratsprobleme auch interessante andere Anwendungen haben, etwa in der Kommunikation über Satelliten.

Eine Anwendung bei der Kommunikation über Satelliten

Bei der Nachrichtenübermittlung über Satelliten hat sich inzwischen auch die digitale Technik durchgesetzt. Bei der Übertragung eines Telefongesprächs von Europa nach Amerika werden zunächst die Daten binär kodiert und in der sendenden Erdfunkstelle zwischengespeichert. Dann werden sie in ganz kurzen Datenstößen an den Nachrichtensatelliten übermittelt. Dort werden sie von einem sogenannten Transponder empfangen, verstärkt und auf einer anderen Frequenz zur empfangenden Erd-

funkstelle in Amerika weitergeleitet (vgl. Abbildung 7.4, Farbabbildungen: S. 151).

An einem solchen System nehmen mehrere Erdfunkstellen gleichzeitig teil. Dafür stehen im Satelliten mehrere Transponder zur Verfügung. Ein Transponder kann aber in einem gewissen Zeitintervall jeweils nur ein Paar von Erdfunkstellen miteinander verbinden. Dabei entsteht folgendes Problem: Die n Erdfunkstellen E_1, E_2, \dots, E_n in Europa sollen Nachrichten an n Erdfunkstellen A_1, A_2, \dots, A_n in Amerika übermitteln. Für die Übermittlung der Daten von E_i an A_j werden genau t_{ij} Zeiteinheiten benötigt. Im Satelliten stehen n Transponder zur Verfügung. Auf welche Weise kann die Übermittlung der gesamten Daten in möglichst kurzer Zeit erfolgen? Die Schwierigkeit, die mit dieser Vorgehensweise verbunden ist, besteht darin, daß eine Erdfunkstelle in Europa nicht gleichzeitig Daten an verschiedene Erdfunkstellen in Amerika übermitteln kann. Zu jedem Zeitpunkt sind durch die Transponder an Bord des Satelliten die Erdfunkstellen in Europa mit festen Empfangsstationen in Amerika verbunden. Aber diese Zuordnung einzelner Stationen zueinander kann sich nach einer kurzen Zeit ändern. Jede solche Zuordnung entspricht einer Heirat aus dem ersten Abschnitt, die jedoch nur eine kurze Zeit andauert und dann durch eine neue Zuordnung (Heirat) ersetzt wird. Das *Zeitschlitz-Zuordnungsproblem* in der Nachrichtentechnik kann also folgenderweise formuliert werden: Gegeben sind Übermittlungszeiten t_{ij} von $E_i, i = 1, 2, \dots, n$, nach $A_j, j = 1, 2, \dots, n$. Gesucht werden Zuordnungen φ_k zwischen den Stationen E_i und A_j sowie Zeiten λ_k , die angeben, wie lange die Zuordnung φ_k andauern soll, daß insgesamt alle Daten in der kürzestmöglichen Zeit übermittelt werden können.

Um dieses Problem zu lösen, können wir folgenderweise vorgehen. Zunächst sei t^* die längste Zeit, die von einer einzelnen Station zum Senden oder Empfangen der Daten benötigt wird. Das Verfahren, das wir nun angeben, wird eine Lösung ermitteln, die garantiert, daß in t^* Zeiteinheiten die Daten *aller* Stationen übermittelt werden. Falls erforderlich, erhöhen wir einige Übermittlungszeiten, daß jede teilnehmende Station in Summe die Zeit t^* zum Senden oder Empfangen ihrer Daten benötigt. Wir bilden nun einen paaren Graphen mit den Knoten $E_i, i = 1, 2, \dots, n$, und $A_j, j = 1, 2, \dots, n$. Zwei Knoten E_i und A_j sind miteinander verbunden, wenn t_{ij} positiv ist. Aufgrund eines Satzes von Birkhoff [Birkhoff, 1946] weiß man nun, daß dieser Graph sicher eine Heirat besitzt. So eine Heirat, die einer Permutation φ_1 entspricht, kann nach Abschnitt *Möglichst große Paarungen* in $K \cdot m \sqrt{n}$ Schritten gefunden werden, wobei m die Anzahl der positiven Zahlen t_{ij}

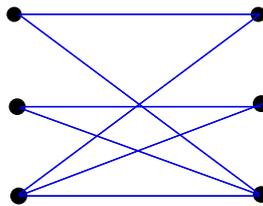
ist. Wie lange soll nun diese Zuordnung andauern? Dazu bestimmen wir

$$\lambda_1 = \min_{1 \leq i \leq n} t_{i, \varphi_1(i)}.$$

Dies heißt, daß λ_1 die kleinste der Zahlen $t_{1, \varphi_1(1)}, t_{2, \varphi_1(2)}, \dots, t_{n, \varphi_1(n)}$ ist. Dadurch wird gewährleistet, daß während λ_1 Zeiteinheiten die Stationen E_i an die Stationen $A_{\varphi_1(i)}$, $i = 1, 2, \dots, n$, Nachrichten übermitteln und keine der Stationen nicht arbeitet. Wir lassen nun diese Stationen ihre Daten in dieser Zeit übermitteln. Danach haben alle Stationen den Zeitbedarf von $t^* - \lambda_1$ Zeiteinheiten für die Übermittlung bzw. den Empfang der verbliebenen Daten. Nun aber können wir das gleiche Verfahren wie oben anwenden: Wir stellen wieder den zugehörigen paaren Graphen auf, der nun mindestens eine Kante weniger besitzt, da im ersten Zeitintervall ein Paar von Sendern und Empfängern bereits alle Daten austauschen konnte. In diesem paaren Graphen wird nun eine neue Zuordnung φ_2 bestimmt und für diese Zuordnung die Zeitdauer λ_2 analog wie oben ermittelt. Da in jedem Schritt die Anzahl der Kanten des paaren Graphen um mindestens 1 abnimmt, bricht das Verfahren nach endlich vielen Schritten ab. Am Ende des Verfahrens sind alle Daten übermittelt. Die Konstruktion im oben beschriebenen Rechenverfahren garantiert, daß $\lambda_1 + \lambda_2 + \dots = t^*$ gilt.

Beispiel.

$$(t_{ij}) = \begin{pmatrix} 3 & 0 & 3 \\ 0 & 4 & 2 \\ 3 & 2 & 1 \end{pmatrix} \quad \text{mit} \quad t^* = 6.$$

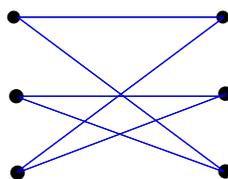


Dabei steht das Element t_{ij} in der i -ten Zeile und j -ten Spalte des obigen Schemas, z.B. $t_{32} = 2$. Im ersten Schritt des Verfahrens wählen wir die Zuordnung $\varphi_1(1, 2, 3) = (1, 2, 3)$ und erhalten

$$\lambda_1 = \min\{t_{11}, t_{22}, t_{33}\} = \min\{3, 4, 1\} = 1.$$

Nachdem diese Werte von den Anfangswerten abgezogen wurden, erhält man

$$(t_{ij}) = \begin{pmatrix} 2 & 0 & 3 \\ 0 & 3 & 2 \\ 3 & 2 & 0 \end{pmatrix} \quad \text{mit} \quad t^* = 6.$$

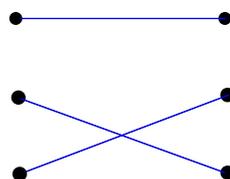


Da $t_{33} = 0$ ist, enthält der neue paare Graph nicht mehr die Kante von E_3 nach A_3 . Wir können als neue Zuordnung φ_2 die Zuordnung $\varphi_2(1, 2, 3) = (3, 2, 1)$ wählen und erhalten

$$\lambda_2 = \min\{t_{13}, t_{22}, t_{3,1}\} = \min\{3, 3, 3\} = 3.$$

Damit erhalten wir als neue Daten

$$(t_{ij}) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 2 & 0 \end{pmatrix} \quad \text{mit} \quad t^* = 12.$$



Daher lautet die letzte Zuordnung $\varphi_3(1, 2, 3) = (1, 3, 2)$ und $\lambda_3 = 2$. Damit haben wir einen Weg gefunden, alle Daten innerhalb eines Zeitintervalles von $\lambda_1 + \lambda_2 + \lambda_3 = 6$ Zeiteinheiten zu übermitteln.

Das obige Rechenverfahren liefert im allgemeinen sehr viele verschiedene Zuordnungen φ_k , die nur sehr kurze Zeiten λ_k verwendet werden. Dies macht dieses Verfahrens in der Praxis unbrauchbar. Möchte man eine optimale Lösung erhalten, die nur n Zuordnungen verwendet, so erhält man ein sehr schweres Problem, das bis jetzt grundsätzlich nur durch systematisches Ausprobieren aller Möglichkeiten lösbar ist. Es gibt nämlich in der diskreten algorithmischen Mathematik zwei große Problemklassen \mathcal{P} und \mathcal{NP} -vollständig. Die Probleme der Komplexitätsklasse \mathcal{P} zeichnen sich dadurch aus, daß es für sie Lösungsverfahren gibt, die nur polynomiell viele Rechenschritte in den (binär kodierten) Eingangsdaten verwenden. Während sich also die Probleme in der Klasse \mathcal{P} durch effiziente Verfahren lösen lassen, kennt man kein solches Verfahren für die Probleme in der Klasse \mathcal{NP} -vollständig. Würde man ein Problem in dieser Klasse

\mathcal{NP} -vollständig effizient lösen können, so könnte man *alle* Probleme in dieser Klasse effizient lösen. Nach dem heutigen Kenntnisstand ist es aber ziemlich unwahrscheinlich, daß es schnelle Lösungsverfahren für die Probleme in der Klasse \mathcal{NP} -vollständig gibt. Man ist daher zum Lösen solcher Probleme auf systematisches Durchsuchen von großen Teilmengen aller Lösungen angewiesen, was einen exponentiellen Aufwand, gemessen an den Eingangsdaten, mit sich bringt.

Um nun doch zu einer praktisch brauchbaren Lösung für das Zeitschlitzzuordnungsproblem zu kommen, haben Lewandowski, Liu und Liu [Lewandowski et al., 1983] vorgeschlagen, an Bord des Satelliten $2n$ Zuordnungen fix zu verdrahten und nur diese Zuordnungen zu verwenden. Wie in Burkard [Burkard, 1991] gezeigt wurde, ist dieses Problem dann wieder effizient lösbar, denn es führt auf die Lösung eines *linearen Zuordnungsproblems*. Für nähere Einzelheiten sei auf [Burkard, 1991] verwiesen.

Lineare Zuordnungsprobleme

Kehren wir zu unserem Heiratsproblem zurück. Wir lassen nun jedes Mädchen M_i angeben, wie gerne es den Herrn H_j hat. Dieser Beliebtheitskoeffizient sei c_{ij} , der umso größer ist, je lieber das Mädchen M_i den jungen Herren H_j hat. Sind alle n^2 Koeffizienten c_{ij} , $i, j = 1, 2, \dots, n$, gegeben, kann man nach einer Heirat (Zuordnung) φ suchen, die das Gesamtglück der heiratenden Paare maximiert. Das heißt, es ist eine Zuordnung φ zu finden, die die *Zielfunktion*

$$c_{1,\varphi(1)} + c_{2,\varphi(2)} + \dots + c_{n,\varphi(n)}$$

maximiert. Dieses Problem heißt *lineares Zuordnungsproblem*. Meist wird die Zielfunktion in einem linearen Zuordnungsproblem nicht maximiert, sondern minimiert. Wir können aber leicht ein Maximierungsproblem in ein Minimierungsproblem überführen. Dazu gehen wir so vor: Sei C eine Konstante, die größer als jeder einzelne der Koeffizienten c_{ij} ist. Man definiert $d_{ij} := C - c_{ij}$ als *Verdruß*, der bei einer Verbindung von M_i mit H_j entsteht. Äquivalent zum Maximierungsproblem von oben ist die Minimierung des Gesamtverdrusses bei einer Heirat, nämlich

$$\text{Minimiere } d_{1,\varphi(1)} + d_{2,\varphi(2)} + \dots + d_{n,\varphi(n)}.$$

Zur Lösung linearer Zuordnungsprobleme gibt es mehrere schnelle Methoden, die auf Ideen der linearen Optimierung oder auf Verfahren zur Lösung

von Flußproblemen in Graphen (vgl. Abschnitt *Möglichst große Paarungen*) beruhen. So können diese Probleme etwa mit $K \cdot n^3$ Rechenoperationen gelöst werden, wobei K wieder eine Konstante ist, die nicht von den Eingangsdaten des Problemes, sondern vom verwendeten Lösungsverfahren abhängt.

Auch in der Praxis kommen lineare Zuordnungsprobleme immer wieder vor: Im letzten Abschnitt wurde bereits erwähnt, daß das Zeitschlitz-zuordnungsproblem bei fest verdrahteten Zuordnungen an Bord des Satelliten durch ein lineares Zuordnungsproblem gelöst werden kann. Brogan [Brogan, 1989] beschreibt, wie man n anfliegende Objekte im Raum lokalisieren und verfolgen kann. Dazu wird von 2 Radarstationen aus der Winkel gemessen, unter denen man diese Objekte zum Zeitpunkt t_0 sieht. Diese Winkel bestimmen Geraden im Raum, auf denen sich die Objekte zur Zeit t_0 befinden. Die von der ersten Radarstation ermittelten Geraden seien g_1, g_2, \dots, g_n , die von der zweiten Radarstation ermittelten Geraden seien h_1, h_2, \dots, h_n . Was man nun nicht weiß, ist, welche der Geraden g_i und h_j dasselbe Objekt bestimmen. Um dies herauszufinden, löst man ein lineares Zuordnungsproblem mit den Koeffizienten c_{ij} , die als kürzeste Abstände zwischen den Geraden g_i und h_j definiert werden. (Man beachte, daß aufgrund von Meßfehlern sich die Geraden, die dasselbe Objekt bestimmen, nicht in einem Punkte schneiden müssen). Auf diese Weise erhält man die Paarung der richtigen Geraden, die jeweils das gleiche Objekt bestimmen und damit kann man auch die Koordinaten der Objekte zum Zeitpunkt T_0 bestimmen. Wiederholt man dieses Verfahren für einen Zeitpunkt t_1 kurze Zeit später, so kann man auch die Koordinaten der Objekte zu diesem Zeitpunkt ermitteln. Löst man nun ein weiteres lineares Zuordnungsproblem mit den Koeffizienten d_{ij} , wobei d_{ij} die Entfernung des i -ten Objektes zum Zeitpunkt t_0 zum j -ten Objekt zum Zeitpunkt t_1 angibt, dann kann man die Bahn der Objekte rekonstruieren und die Objekte somit auf ihrer Flugbahn verfolgen.

Wir können aber auch eine Modifikation des linearen Zuordnungsproblems betrachten. Anstelle wie oben die Summe der einzelnen Verdrußkoeffizienten zu minimieren, können wir nach einer Heirat suchen, die den maximalen Einzelverdruß so klein als möglich macht. Dies führt auf sogenannte Engpaß-Zuordnungsprobleme der Form

$$\text{Minimiere } \max(d_{1,\varphi(1)}, d_{2,\varphi(2)}, \dots, d_{n,\varphi(n)}).$$

Diese Engpaß-Zuordnungsprobleme sind schneller lösbar als die analogen Zuordnungsprobleme, in denen eine Summe minimiert werden soll. Auch im obigen Objekt-Lokalisierungsproblem würde es mehr Sinn machen, Engpaß-Zuordnungsprobleme anstelle gewöhnlicher linearer Zuordnungsprobleme zu lösen. Darüber hinaus treten Engpaß-Zuordnungsprobleme etwa in folgendem Zusammenhang auf: Zur Erledigung von n Arbeiten stehen n Leute zur Verfügung, die aber nicht alle gleich gut für die einzelnen Arbeiten geeignet sind. Es sei d_{ij} der Zeitbedarf, den die i -te Person zur Erledigung der j -ten Aufgabe benötigt. Um nun alle Aufgaben in der kürzestmöglichen Zeit zu erledigen, löst man ein Engpaß-Zuordnungsproblem mit den Koeffizienten d_{ij} .

Rundreiseprobleme

Im vorigen Abschnitt wurde erwähnt, daß sich lineare Zuordnungsprobleme effizient lösen lassen. Hier betrachten wir nun ein Zuordnungsproblem mit einer zusätzlichen Bedingung. Wir lassen nicht mehr alle Permutationen als Lösungen zu, sondern nur mehr die sogenannten *zyklischen Permutationen*. Eine Permutation φ heißt zyklisch, wenn man durch $1, \varphi(1), \varphi(\varphi(1)), \varphi(\varphi(\varphi(1))), \dots$ alle Zahlen $1, 2, \dots, n$ erhält. So ist etwa die Permutation $\varphi_1(1, 2, 3, 4) = (2, 3, 4, 1)$ zyklisch, während die Permutation $\varphi_2(1, 2, 3, 4) = (2, 1, 4, 3)$ nicht zyklisch ist. Letztere Permutation zerfällt vielmehr in 2 Zyklen $1 \rightarrow 2 \rightarrow 1$ und $3 \rightarrow 4 \rightarrow 3$.

Zyklische Permutationen spielen bei sogenannten *Rundreiseproblemen* eine tragende Rolle. Gegeben seien n Objekte, z.B. Punkte im Raum oder Städte in Europa. Ferner sei d_{ij} die Entfernung zwischen dem i -ten und j -ten Punkt (bzw. der i -ten und j -ten Stadt). Wir suchen nun eine Tour (Rundreise), die alle Punkte genau einmal besucht und dann zum Ausgangspunkt zurückkehrt. So eine Tour oder Rundreise entspricht einer zyklischen Permutation. Fordern wir außerdem, daß die Tour eine minimale Gesamtlänge hat, so müssen wir das folgende *Rundreiseproblem* lösen:

Man finde eine zyklische Permutation φ , die $d_{1,\varphi(1)} + d_{2,\varphi(2)} + \dots + d_{n,\varphi(n)}$ minimiert.

Auch Rundreiseprobleme haben zahlreiche interessante Anwendungen, sei es die Erstellung von optimalen Touren für die tägliche Zeitungszustellung, das Leeren von Briefkästen oder die optimale Bestimmung einer Produktionsreihenfolge für Aufträge auf einer Maschine, die zwischen der Be-

arbeitung zweier verschiedener Aufträge gereinigt werden muß (vgl. Abbildung 7.5, Farbabbildungen: S. 152).

Obwohl das Rundreiseproblem formal einem linearen Zuordnungsproblem sehr ähnlich sieht, ist es viel schwerer zu lösen. Es gehört nämlich der Klasse der \mathcal{NP} -vollständigen Probleme an, für die nur Verfahren bekannt sind, die eine große Teilmenge der zulässigen Lösungen systematisch absuchen.

Man kann sich aber die Frage stellen: Gibt es Bedingungen für die Koeffizienten d_{ij} eines Rundreiseproblems, so daß das entsprechende Problem effizient lösbar wird? Und tatsächlich kennt man eine ganze Reihe von solchen Bedingungen, siehe etwa den jüngst erschienenen Übersichtsartikel von Burkard et al. [Burkard et al., 1998b]. Eine solche Bedingung ist etwa die *Monge Bedingung*, benannt nach dem großen französischen Geometer Gaspard Monge (1746–1818). Monge stellte bei Schanzarbeiten an Verteidigungswällen fest, daß es günstiger ist, wenn die Karren direkt an den Wall heranfahren, ohne sich zu kreuzen. Überträgt man diese Bedingung auf Rundreiseprobleme, so erhält man

Monge Bedingung

Die Koeffizienten d_{ij} erfüllen die Mongebedingung wenn gilt

$$d_{ij} + d_{kl} \leq d_{il} + d_{kj} \text{ für } 1 \leq i < k \leq n, 1 \leq j < l \leq n.$$

Natürlich erfüllt nicht jedes Rundreiseproblem so eine Bedingung. Aber wenn diese Bedingung erfüllt ist, kann man das Rundreiseproblem in $K \cdot n$ Schritten lösen, wobei die Konstante K unabhängig von den Problemdata ist.

Quadratische Zuordnungsprobleme

Nachdem wir im 1. Abschnitt Heiraten betrachtet haben, wenden wir uns nun der Sitzordnung an der Hochzeitstafel zu. Dazu kann man das folgende mathematische Modell verwenden. Wir nehmen an, daß n Gäste kommen und n Plätze an der Tafel zur Verfügung stehen. Von einem Platz an der Tafel aus kann man sich gut mit den Nachbarplätzen unterhalten und einigermaßen gut auch mit gegenüberliegenden und übernächsten Nachbarplätzen (vgl. Abbildung 7.6, Farbabbildungen: S. 152).

Wir messen die Kommunikationsmöglichkeit durch nichtnegative Kommunikationskoeffizienten a_{ij} , $i, j = 1, 2, \dots, n$. Dabei drückt a_{ij} aus, wie

gut man sich zwischen den Plätzen i und j unterhalten kann. Je besser, umso höher ist dieser Koeffizient. Ferner sollen die Größen b_{kl} , $k, l = 1, 2, \dots, n$, beschreiben, wie gerne sich die k -te Person mit dem l -ten Gast der Hochzeitstafel unterhalten würde. Gesucht wird nun eine Zuordnung φ der Plätze an der Tafel zu den Gästen, daß die Gesamtrunde möglichst fröhlich ist. Dies führt auf das Modell

$$\text{Maximiere } \sum_{i=1}^n \sum_{j=1}^n a_{ij} \cdot b_{\varphi(i), \varphi(j)}$$

Dieses Problem heißt *quadratisches Zuordnungsproblem* und spielt in vielerlei Anwendungen eine sehr wichtige Rolle. Dabei wird es meist als Minimierungsproblem formuliert. Im folgenden werden einige Anwendungen des quadratischen Zuordnungsproblems angeführt:

Entwurf von Schreibmaschinentastaturen ([Burkard und Offermann, 1977])

Es soll eine Schreibmaschinentastatur entwickelt werden, die es gestattet, Texte in bestimmten Sprachen möglichst schnell zu schreiben. n sei die Anzahl von Zeichen, die auf der Tastatur plaziert werden sollen. Die Koeffizienten a_{ij} beschreiben, wie schnell es möglich ist, nach der Taste i die Taste j zu drücken. Die Größen b_{kl} beschreiben die Häufigkeit, wie oft die Zeichenpaare k, l im Text auftreten. Die Zuordnung φ belegt jede Taste mit einem Zeichen, so daß die Gesamtsumme $\sum_{i=1}^n \sum_{j=1}^n a_{ij} \cdot b_{\varphi(i), \varphi(j)}$ minimal wird. Diese Gesamtsumme entspricht aber genau der durchschnittlichen Zeitdauer, die zum Schreiben eines Textes nötig ist.

Campusplanung ([Dickey und Hopkins, 1972])

Auf einem neuen Gelände soll ein neuer Universitätscampus mit n Gebäuden errichtet werden. Dazu stehen am Gelände n Bauplätze zur Verfügung. Jedes der Gebäude hat eine spezielle Funktion: Bibliothek, Institutsgebäude, Mensa, Studentenwohnheim etc. Die Entfernungen zwischen den Bauplätzen werden durch die Größen a_{ij} beschrieben. Die Koeffizienten b_{kl} geben an, wie oft in einer Woche Leute zwischen einem Gebäude der Funktion k und einem Gebäude der Funktion l hin- und hergehen. Gesucht ist eine Zuordnung der einzelnen Gebäude zu den Bauplätzen, so daß die wöchentlich zurückgelegte Gesamtentfernung aller Betroffenen minimiert wird. Man sucht also eine Zuordnung φ der Bauplätze zu den Gebäuden, so daß $\sum_{i=1}^n \sum_{j=1}^n a_{ij} \cdot b_{\varphi(i), \varphi(j)}$ minimiert wird.

Weitere Anwendungen des quadratischen Zuordnungsproblems betreffen die Anordnung von Schaltelementen auf einer Kontrolleinheit, um die Ermüdung der Augen des Bedienungspersonals zu minimieren, den Entwurf von gut geplanten Spitälern, das zeitliche Reihen von archäologischen Funden, die optimale Reihung von Staffelläufern, die optimale Planung paralleler Produktionslinien oder etwa die Analyse organisch-chemischer Verbindungen.

Quadratische Zuordnungsprobleme sind leider sehr schwer optimal zu lösen. Sie gehören nämlich auch der Klasse der \mathcal{NP} -vollständigen Probleme an. Eines der größten quadratischen Zuordnungsprobleme wurde im Jahr 1997 an der ETH in Zürich optimal gelöst, und zwar für $n = 25$. Dazu wurde wochenlang die Rechenkraft paralleler Großcomputer ausgenützt.

Man kann sich nun die Frage stellen, warum gerade quadratische Zuordnungsprobleme so schwer zu lösen sind. Eine Teilantwort darauf gibt eine Theorie, die von Burkard und Fincke [Burkard und Fincke, 1983] entwickelt wurde und die das Verhalten von großen kombinatorischen Optimierungsproblemen beschreibt. Während etwa bei Rundreiseproblemen die Werte für die beste und für die schlechteste Lösung immer weiter auseinandergehen, wenn die Problemgröße zunimmt, hat man bei quadratischen Zuordnungsproblemen gerade das umgekehrte Verhalten. Man kann zeigen, das bei zufällig erzeugten Problemen mit wachsender Problemgröße alle Lösungen fast sicher einen konstanten Wert annehmen. Das bedeutet, daß sich die Zielfunktionswerte der einzelnen Lösungen nur sehr wenig unterscheiden. Daher ist es so schwierig, aus all den Lösungen, die fast den gleichen Wert haben, die wirklich beste herauszufinden. Andererseits kann man dieses seltsame Phänomen aber auch positiv bewerten: Da die Lösungen fast alle den optimalen Wert ergeben, findet man sehr rasch sehr gute Lösungen, ohne jedoch zu wissen, ob sie wirklich optimal sind. Dies ist für die Praxis aber meist gar nicht so wichtig. Viel wichtiger ist es, rasch viele gute Lösungen zu finden und dies ist nach dieser Theorie umso eher möglich, je größer das Problem ist. Tatsächlich stützen die numerischen Ergebnisse von vielen Rechentests diese theoretischen Vorhersagen.

Ausblick

In den vorhergehenden Abschnitten haben wir gesehen, daß es zwei wichtige Klassen von kombinatorischen Optimierungsaufgaben gibt: jene der Komplexitätsklasse \mathcal{P} , die durch effiziente Verfahren gelöst werden können,

und jene der \mathcal{NP} -vollständigen Probleme, die sich bis jetzt einer effizienten Lösung hartnäckig widersetzen. Zu den Forschungsaufgaben zählt bei beiden Klassen die Entwicklung von guten Lösungsverfahren. Bei der zweiten Klasse ist man insbesondere auch an guten Näherungsverfahren interessiert. Weitere Forschungsgesichtspunkte betreffen das Identifizieren von schnell lösbaren Spezialfällen \mathcal{NP} -schwerer Probleme, das Verhalten von Problemen bei wachsender Problemgröße und natürlich vor allem die Anpassung der Modelle an die Gegebenheiten der Praxis.

Für ein tieferes Eindringen in den Themenkreis dieses Artikels sei die interessierte Leserin oder der interessierte Leser auf das Buch von D. Jungnickel: *Graphen, Netzwerke und Algorithmen*, Bibliographisches Institut Wissenschaftsverlag, Mannheim, 1994 hingewiesen. Wer sich für den neuesten Stand der Forschung auf dem Gebiet der Zuordnungsprobleme interessiert, sei auf die beiden Übersichtsartikel [Burkard und Çela, 1999] und [Burkard et al., 1998a] im *Handbook of Combinatorial Optimization* hingewiesen.

Literatur

- [Birkhoff, 1946] Birkhoff, G., (1946). Tres observaciones sobre el algebra lineal. *Rev. univ. nac. Tucum n (A)* 5, 147–151.
- [Brogan, 1989] Brogan, W. L., (1989). Algorithm for ranked assignments with application to multiobject tracking. *Journal of Guidance* 12, 357–364.
- [Burkard, 1991] Burkard, R. E., (1991). Time Division Multiple Access systems and matrix decomposition. In *Proceedings of the Fourth European Conference on Mathematics in Industry*, von H. J. Wacker und W. Zulehner, Hrsg., Teubner, Stuttgart, 35–46.
- [Burkard und Çela, 1999] Burkard, R. E. und Çela, E. (1999). Linear Assignment problems and extensions. Bericht Nr.127, SFB Optimierung und Kontrolle, Technische Universität Graz, Juni 1998. Erscheint im *Handbook of Combinatorial Optimization*, Kluwer, Dordrecht, 1999.
- [Burkard et al., 1998a] Burkard, R. E., Çela, E., Pardalos, P. M. und Pitsoulis, L. S. (1999). The Quadratic Assignment Problem. *Handbook of Combinatorial Optimization*, vol. 3, 241–337, Du D. Z. and Pardalos, P. M., Hrsg. Kluwer, Dordrecht, 1998.

- [Burkard et al., 1998b] Burkard, R. E., Deĭneko, V. G., van Dal, R., van der Veen, J. und Woeginger G. J. (1998). Well solvable special cases of the travelling salesman problem: a survey. *SIAM Review* 40, 1998, 496–546.
- [Burkard und Fincke, 1983] Burkard, R. E. und Fincke, U. (1983). The asymptotic probabilistic behaviour of the quadratic sum assignment problem. *Zeitschrift für Operations Research* 27, 73–81.
- [Burkard und Offermann, 1977] Burkard, R. E. und Offermann, J. (1977). Entwurf von Schreibmaschinentastaturen mittels quadratischer Zuordnungsprobleme. *Zeitschrift für Operations Research* 21, B121–B132.
- [Dickey und Hopkins, 1972] Dickey, J. W. und Hopkins, J. W. (1972). Campus building arrangement using TOPAZ. *Transportation Research* 6, 59–62.
- [Ford und Fulkerson, 1956] Ford, L. R. und Fulkerson, D. R. (1956). Maximal flow through a network. *Canadian J. Math* 8, 399–404.
- [Hall, 1935] Hall, Ph. (1935). On representatives of subsets. *J. London Math. Soc.* 10, 26–30.
- [Hopcroft und Karp, 1973] Hopcroft, J. E. und Karp, R. M. (1973). An $n^{5/2}$ algorithm for maximum matchings in bipartite graphs. *SIAM Journal on Computing* 2, 225–231.
- [Karmarkar et al., 1993] Karmarkar, N., Karp, R., Lipton, R., Lovász, L. und Luby, M. (1993). A Monte-Carlo algorithm for estimating the permanent. *SIAM J. Comput.* 22, 284–293.
- [Lewandowski et al., 1983] Lewandowski, J. L., Liu J. W. S. und Liu, C. L. SS/TDMA time slot assignment with restricted switching modes. *IEEE Transactions on Communications*, COM-23, 149–154.
- [Tutte, 1947] Tutte, W.T. (1947). The factorization of linear graphs. *J. London Math. Soc.* 22, 107–111.

Das menschliche Gehirn – nur ein Rechner?

Wolfgang Maass

Einführung

Die Vorstellung, daß das menschliche Gehirn im Prinzip durch einen künstlichen Rechner ersetzt werden könnte, wirkt nicht sehr plausibel. Wir alle haben ja eine recht gute Vorstellung davon, was ein künstlicher Rechner (= Computer) kann, oder vielmehr nicht kann:

- man muß ihm vorher alles haargenau erklären (d.h.: einprogrammieren) bevor er eine Aufgabe ausführen kann
- er wird nicht aus Erfahrung klug: wenn wir sein Programm nicht ändern, wird er denselben Fehler wieder und wieder machen
- er kann sich nicht selbständig weiterentwickeln, und ist in keiner Weise kreativ
- er hat keine Gefühle oder Bewußtsein
- er ist instabil und anfällig: wir alle wissen wie gerne ein Rechner “abstürzt”, und wie anfällig er ist gegenüber eingeschleusten Computerviren sobald er mit anderen Rechnern in Kontakt tritt.

Obwohl die kommerziell vertriebenen Rechner im Laufe der Zeit immer schneller und kleiner wurden, haben sie sich erstaunlich wenig verbessert in bezug auf die angeführte Mängelliste. Ich werde in diesem Aufsatz argumentieren, daß dies gar nicht so erstaunlich ist, weil diese Mängel auf grundlegenden in der Theorie der Rechner verankerten “Planungsfehlern” beruhen, die wiederum tief in althergebrachten aber zum Teil fragwürdigen Vorstellungen über Maschinen und menschliche Intelligenz verwurzelt sind.

Ich werde diese Kritik anhand des wohlbekannten Turing-Tests präzisieren, den der geniale Mathematiker und Computer-Pionier Alan Turing vor fast genau 50 Jahren veröffentlicht hatte [Turing, 1950]. Er schlug vor,

folgendes Kriterium dafür zu wählen, ob ein künstlicher Rechner "intelligent" sei:

Ein Rechner sollte dann als intelligent bezeichnet werden, wenn wir als Mensch bei einem beliebigen Frage-und-Antwort Spiel, das über eine elektronische Verbindung (zum Beispiel per email) durchgeführt wird, nicht unterscheiden können, ob am anderen Ende der Leitung dieser Rechner oder ein anderer Mensch sitzt.

Dieser Test ist gedanklich sehr attraktiv, und er hat unsere Vorstellung von "maschineller Intelligenz" nachhaltig geprägt. Erst wenn man etwas länger über diesen Test nachdenkt, fällt einem auf, daß er stark beeinflusst ist von einem Intelligenzbegriff, dessen Akzeptanz heute im Schwinden begriffen ist. Der Turing-Test beruht auf der Vorstellung von Intelligenz als der Fähigkeit, bei einer Unterhaltung mit Menschen, die sich selbst für intelligent halten, mitreden zu können, ohne dabei unangenehm aufzufallen. Ein offensichtlicher Defekt dieses Intelligenzbegriffs ist, daß er jemanden, der quasi als Schauspieler typische Sprachspiele von selbsternannten intelligenten Menschen mitspielen kann, ohne notwendigerweise zu *verstehen* worüber er redet, als intelligent bezeichnen würde. Weitere Defekte des im Turing-Test formalisierten Intelligenzbegriffs sind dessen Festlegung auf verbale Kommunikation, sowie die Vernachlässigung von anderen wichtigen Aspekten der Intelligenz, wie zum Beispiel der "praktischen Intelligenz", oder allgemeiner der Fähigkeit "sich zu helfen zu wissen", also der Fähigkeit unvorhergesehene Probleme in kreativer Weise zu lösen.

Bedauerlicherweise hat die dem Turing-Test zugrunde liegende einseitige Vorstellung von Intelligenz nicht nur viele Generationen von Schülern und Studenten, sondern auch unsere traditionellen Rechner geprägt. Der Zusammenhang zwischen der im Turing-Test formalisierten Vorstellung von Intelligenz und den Unzulänglichkeiten unserer gegenwärtigen Rechner liegt auf der Hand: man kann Rechner jahrzehntelang weiterentwickeln und deren Abschneiden im Turing-Test verbessern, ohne daß die so entstehenden Rechner notwendigerweise große Fortschritte hinsichtlich der am Anfang angeführten Mängelliste aufweisen würden. In diesem Sinn habe ich am Beginn dieses Artikels etwas überspitzt argumentiert, daß man die angeführten Mängel unserer gegenwärtigen Rechner als Planungsfehler auffassen kann.¹

Als Gegenpol möchte ich hier einen alternativen Test für die "Intelligenz" eines Rechners vorschlagen:

Die Intelligenz eines Rechners wird daran gemessen wie lange er unter wechselnden und teilweise unvorhersehbaren und widrigen Umständen "überlebt".

Dabei kann man "überleben" eines Rechners in geeigneter Kontextbezogener Weise definieren (zum Beispiel als Aufrechterhalten seiner üblichen Rechnerleistung ohne Absturz.) Je länger die Überlebensdauer eines Rechners ist und umso widriger und stärker schwankend die Umstände waren, denen der Rechner getrotzt hat, umso "intelligenter" würde ich einen solchen Rechner nennen. Dabei kann man bei den "unvorhersehbaren und teilweise widrigen Umständen", denen der Rechner ausgesetzt wird, an das Einspielen von fehlerbehafteten neuen Programmen, Unmutsäußerungen des menschlichen Benutzers oder anderer Maschinen, menschliche Bedienungsfehler, sich einschleichende Computerviren, oder sogar an Stromausfall, Wasserrohrbruch, oder Erdbeben denken.

Gemäß diesem Kriterium sind selbst primitive Lebewesen wie Käfer oder Stubenfliegen in gewisser Weise intelligenter als unsere gegenwärtigen Rechner, weil sie vielen physischen Bedrohungen geschickt ausweichen können. Diese Folgerung ist kein Widerspruch, sondern sie macht die Schmalspurigkeit des im Turing-Test formalisierten Intelligenzbegriffs besonders deutlich: Intelligenz bedeutet ja auch Gefahren rechtzeitig zu erkennen, sich mit beschränkten Mitteln in der Not zu helfen wissen.

Offensichtlich ist das "timing", oder genauer gesagt: die Rechengeschwindigkeit, ein wesentlicher Bestandteil des alternativen Intelligenzbegriffs, denn es nützt nichts, eine perfekte Lösung in einer Minute Rechenzeit zu finden, wenn eine das Überleben ermöglichende Reaktion in "Echtzeit", also zum Beispiel innerhalb 1/10 Sekunde erforderlich ist. Das bedeutet, daß die Gültigkeit der Turing Maschine² als geeignetes mathematisches Modell für intelligente Rechner in Frage gestellt wird. Die Turing Maschine ist ein Modell für einen *sequentiellen* Rechner, in dem in jedem Zeittakt jeweils nur eine Rechenoperation ausgeführt werden kann. Daher können nur einige extrem einfache Rechenaufgaben auf einer Turing Maschine in "Echtzeit", also sagen wir in höchstens 10 Zeittakten, ausgeführt werden.³ Dagegen sind biologische Nervensysteme dadurch ausgezeichnet, daß sie einige spezielle Aufgaben der Informationsverarbeitung, deren schnelle Lösung "überlebenswichtig" ist, in Echtzeit ausführen können. Dies erreichen sie dadurch, daß sie solche Aufgaben an eine große Anzahl von "Prozessoren" verteilen, die "parallel" an der Problemlösung arbeiten. Dabei hilft ihnen die Architektur ihrer Schaltkreise, deren Grundbausteine – die Neuronen –

Inputs von 5000 – 10000 anderen Neuronen bekommen, und ihren Output an eine etwa gleichgroße Zahl anderer Neuronen parallel (d.h. gleichzeitig) übermitteln können. Bei der Konstruktion einer gemäß dem alternativen Intelligenz-Kriterium "intelligenten" Maschine muß man sich entscheiden, welche komplexen aber "überlebenswichtigen" Berechnungen auf ihr unbedingt in Echtzeit durchführbar sein müssen. Da auf einer Turing Maschine *keinerlei* komplexe Berechnungen in Echtzeit durchgeführt werden können, ist sie als mathematisches Modell für derart "intelligente" Maschinen ungeeignet.

Ich werde im zweiten Abschnitt dieses Artikels einige Forschungsansätze vorstellen, die zeigen, daß es durchaus schon Ideen und Bausteine gibt für einen neuartigen Typ von Rechnern, die gemäß dem alternativen Intelligenz-Kriterium als "intelligent" bezeichnet werden könnten. Dies läßt uns erahnen wieviel mehr ein Rechner im Prinzip kann – verglichen mit dem traditionellen Begriff eines Rechners. Gleichzeitig werden wir Indizien für eine möglicherweise überraschende Entwicklung sehen: Während man früher das menschliche Gehirn als eine Variante von Idealisierungen vorhandener Rechner (also zum Beispiel von Turing Maschinen) zu verstehen suchte, hat sich mittlerweile der Spieß umgedreht: viele bahnbrechende neue Ideen in der Informatik haben ihre Wurzel in Erkenntnissen und Vermutungen über die Arbeitsweise biologischer Nervensysteme.

Wenn ich im dritten Abschnitt dieses Artikels zurückkomme auf die Frage "Das menschliche Gehirn – nur ein Rechner?" werden wir sehen, daß das eigentlich Problematische an dieser Vermutung das Wörtchen "nur" ist, weil ein Rechner, der alles wirklich "Rechnermögliche" ausschöpft, ein ungleich anderes Wesen ist als die uns bisher geläufigen Rechner. Gleichzeitig werden wir sehen, daß eine positive Antwort auf die diesem Aufsatz zugrundeliegende Frage nicht aufgefaßt werden könnte als ein Sieg der Natur- und Ingenieurwissenschaften über die Geisteswissenschaften, der Maschine über die belebte Materie. Vielmehr wird sichtbar, daß die Entwicklung wirklich "intelligenter" Maschinen dazu führt, daß die Grenzen zwischen beiden Bereichen verwischt werden.

Vorboten eines neuen Typs intelligenter Rechner

Die Entwicklung von Rechnern, die gemäß dem im ersten Abschnitt vorgestellten alternativen Intelligenz-Kriterium als intelligent zu bezeichnen wären, erfordert ein radikales Umdenken. Ich werde in diesem Abschnitt einige erfolgversprechende Ansätze skizzieren.

Lange Zeit war bei der Entwicklung maschineller Intelligenz das Hauptziel, einen maschinellen Theoretiker zu schaffen, also eine Maschine, die ein theoretisches Modell ihrer Umwelt aufbaut, und die zur Lösung eines konkreten Problems Strategien sucht für die sie logisch herleiten kann, daß sie optimal sind. Bei diesem Ansatz ist man auf zwei Probleme gestoßen:

- Solange der Rechner keine begleitende Intuition für sein theoretisches Welt-Modell besitzt, muß er bei der Berechnung von sachgerechten Problemlösungen mehr oder weniger *blind suchen* bis er *zufällig* auf eine Strategie stößt, die "paßt". Für ihn ist ja jede mögliche Strategie oder "Aktion" nur eine Zeichenreihe, die genauso gut in Chinesisch kodiert sein könnte, weil er ihren Inhalt sowieso nicht versteht. Da der Suchraum für mögliche Problemlösungen in der Regel exponentiell groß ist in der Komplexität n des Problems (wobei man n als die Anzahl der "Freiheitsgrade" oder Teilschritte von möglichen Strategien auffassen kann), führt dies schon bei relativ einfachen Problemen (also zum Beispiel für $n = 100$) zu einer erforderlichen Anzahl von Rechenschritten der Größenordnung 2^{100} . Die allerschnellsten gegenwärtig existierenden Rechner können weniger als 2^{44} Rechenschritte pro Sekunde ausführen. Zur Ausführung von 2^{100} Rechenschritten würde solch ein Rechner 2^{56} Sekunden, also ca. 1 000 000 000 Jahre benötigen.

Ein wesentlicher Durchbruch könnte hier im Prinzip erzielt werden, wenn man beweisen könnten, daß die Komplexitätsklassen P und NP zusammenfallen.⁴ An diesem Problem wird in der Theoretischen Informatik und Mathematik seit 30 Jahren intensiv gearbeitet, ohne daß eine Lösung am Horizont sichtbar wäre. Die gängige Vermutung besagt allerdings, daß die Komplexitätsklassen P und NP *nicht* zusammenfallen, was bedeuten würde, daß es bei wichtigen Problemen des maschinellen logischen Denkens keine wesentlich schnellere Alternative zum blinden Suchen nach passenden Schlußketten gäbe.

- Ein weiteres und möglicherweise noch schwierigeres Problem ist das folgende: Wir als Menschen schaffen es in der Regel halbwegs richtige Entscheidungen im Alltagsleben zu treffen, obwohl sowohl das uns zur Verfügung stehende Hintergrundwissen⁵ unmittelbar für die Entscheidung relevanten Sinneseindrücke in der Regel vieldeutig und

oft sogar widersprüchlich sind. Diese Komplikation wird uns meist gar nicht bewußt solange wir nicht darüber nachdenken. Sie führt aber zur bedauerlichen Tatsache, daß die Kalküle der mathematischen Logik und die bisher vorhandenen Methoden des maschinellen Beweisens nicht geeignet sind um einer Maschine in einer natürlichen Umgebung sachgerechte Aktionen vorzuschlagen.⁶

Ein besonders interessantes Anwendungsgebiet für maschinelle Intelligenz ist die Robotik. Die beiden genannten Probleme haben bewirkt, daß ein mit traditioneller künstlicher Intelligenz, also mit Hintergrundwissen und logischen Schlußweisen, ausgestatteter autonomer mobiler Roboter sich in einer nicht speziell für ihn präparierten Umgebung in der Regel nur peinlich langsam und nicht besonders zufriedenstellend bewegt.

Ein überraschender Neuansatz zur Lösung dieses Problems, der heute oft als "Neue Robotik" bezeichnet wird, wurde der staunenden Fachwelt im Jahr 1986 von Rodney Brooks⁷ vom MIT in Cambridge (USA) vorgestellt. Genaugenommen waren einige dieser revolutionären Ideen schon vorher von einem krassen "Außenseiter" in der Robotik, dem Neurophysiologen Valentino Braitenberg, vorgeschlagen worden.⁸ Als Alternative zu den damals besten Robotern, die große Rechner mit sich herumschleppten und lange darüber "nachdenken" mußten wie sie einem Hindernis ausweichen sollten, stellte Rodney Brooks kleine käferartige Leicht-Roboter vor, die flink durch die Gegend wieselten und Hindernisse geschickt umgingen. Ihre Rechner-Architektur beruht auf einem neuen Prinzip, das technisch als Subsumptionsarchitektur⁹ bezeichnet wird. Hier versucht man nicht mehr, ein immer komplizierter werdendes Modell der Wirklichkeit im Rechner nachzubauen und bei Entscheidungen zu befragen. Stattdessen besteht das "Innenleben" dieser neuen Wesen aus einem geschickt koordinierten Bündel von spezialisierten "Reflexen", zum Beispiel dem Reflex ein unmittelbar vor ihm befindliches Hindernis zu umgehen, oder dem Reflex eine Ladestation aufzusuchen, sobald die Spannung am Akku des Roboters eine Schwelle unterschreitet, oder dem Reflex ein bestimmtes vom Benutzer vorgegebenes Ziel aufzusuchen. Das Prinzip von Brooks' Subsumptionsarchitektur besteht darin, daß zunächst jeder einzelne dieser Reflexe für sich möglichst direkt durch geeignete Verbindungen von den "Sinnesorganen" (d.h. von Lichtsensoren, sonaren Sensoren, Kameras oder Laser-Abstandsmessern) zu den Motoren des Roboters in möglichst einfacher und stabiler Weise implementiert wird (mehr dazu im nächsten Unterabschnitt). Für den Fall, daß verschiedene dieser Reflexe zu gegensätzlichen Motor-

Kommandos führen, werden solche Konflikte intern in einer Weise gelöst, die das langfristige "Überleben" des Roboters optimiert. Zum Beispiel: wenn die Akku-Spannung niedrig ist, aber ein Hindernis auf dem direkten Weg zur Ladestation auftaucht, dann ist es momentan wichtiger diesem Hindernis auszuweichen als zu versuchen, mit "dem Kopf durch die Wand" auf dem direktesten Weg zur Ladestation zu gelangen.

Der geschilderte Neuansatz von Rodney Brooks ist aus der gegenwärtigen Robotik nicht mehr wegzudenken. In Schwierigkeiten gerät dieser Ansatz dort, wo die Vielfalt der zu kontrollierenden Reflexe oder die Kompliziertheit der zu bewältigenden Aufgaben es dem Ingenieur nicht mehr erlauben, die Regeln für ein geeignetes Zusammenspiel der Reflexe mittels seiner eigenen Intuition festzulegen. Ein menschlicher Betrachter dieser Problematik ist vielleicht erinnert an nicht ganz unähnliche Konflikte zwischen kurz- und langfristigen Zielen verschiedener Art, die sich – teilweise unbewußt – im Menschen abspielen.¹⁰

Ein gemäß der Subsumptionsarchitektur konstruierter Roboter ist zwar einem relativ "primitiven" Lebewesen wie einem Käfer sehr viel ähnlicher als einem komplexen Lebewesen, aber im Unterschied zu einem gemäß der traditionellen künstlichen Intelligenz konstruierten Roboter haben einige dieser Reflexe eine unmittelbar erfahrbare *Bedeutung* für ihn: die gegenwärtige Akkuspannung ist zum Beispiel nicht eine Zahl wie jede andere, sondern sie ist ein unmittelbarer Indikator dafür wie lange er noch umherfahren kann bevor er eine Ladestation aufsuchen muß. Ebenso ist ein hoher Meßwert von Infrarot-Sensoren in Fahrtrichtung ein recht sicheres Vorzeichen für eine unmittelbar bevorstehende Kollision. In diesem Sinn haben diese Sensor-Werte eine unmittelbare *Bedeutung* für den Roboter. Sobald aber intern kodierte Informationen eine direkte Bedeutung für die Existenz einer Maschine haben, erfordert es nur einen kleinen zusätzlichen Schritt, daß die Maschine Gefühle zeigt, also zum Beispiel Freude über eine Akkuladung oder Sorge, falls eines seiner Räder nicht mehr die gesendeten Dreh-Befehle ausführt (siehe zum Beispiel die ausführliche Diskussion in Kapitel 2 von [Picard, 1997]).

Zugegebenermaßen sind die angesprochenen praktischen Probleme, wie die Vermeidung von Kollisionen, recht primitiv, verglichen etwa mit den abstrakten Überlegungen eines Logikers oder Mathematikers. Vielleicht haben beide Bereiche aber trotzdem etwas miteinander zu tun: Hat nicht fast jeder Logiker und Mathematiker — bewußt oder unbewußt — räumliche Bilder vor Augen, selbst bei vollkommen abstrakten Gedankengängen, also

auch beim *formalen Denken*? Ist dieser Effekt vielleicht ein Überbleibsel des "Käfers in uns", oder genauer gesagt: der über Millionen von Jahren der Evolution angesammelten Erfahrungen im Umgang mit 3-dimensionalen Objekten? Und liegen die zahlreichen Mißerfolge beim Versuch kreatives formales Denken (zum Beispiel: maschinelles Beweisen) von Rechnern durchführen zu lassen, vielleicht daran, daß man zu geradlinig versucht hat, den Rechner wirklich rein "formal", also ohne begleitende quasi-räumliche Anschauung, denken zu lassen — was selbst uns Menschen kaum gelingt?

Schließlich möchte ich anmerken, daß der Paradigmenwechsel "vom Logiker zum Käfer" in der Robotik weniger perfekt ist als ich es bisher dargestellt habe. Probleme treten schon dann auf, wenn man in einem komplexen Roboter das Zusammenspiel seiner "Reflexe" so gestalten möchte, daß sein "Überleben" optimiert wird. Man ist versucht, dem Roboter hierfür Hintergrundwissen über die relative Wichtigkeit einzelner Reflexe einzuprogrammieren, aus denen er sich sinnvolle Koordinationsregeln *logisch* herleiten kann (zum Beispiel: "Auch wenn die Akku-Spannung niedrig ist, müssen Hindernisse auf dem Weg zur Ladestation umgangen werden.").

Bei Käfern und anderen Lebewesen ist das Zusammenspiel der Reflexe im Laufe der *Evolution* über Millionen von Jahren hinweg soweit optimiert worden, daß sie damit überleben können. Diese Methode hat man auch in der gegenwärtigen Robotik erfolgreich eingesetzt: Indem man eine "Evolution" künstlich simuliert, also Teilstücke der Reflex-Koordination der überlebensfähigsten Exemplare der gegenwärtigen Generation von Robotern in verschiedenen Weisen rekombiniert ("Kreuzung") und lokal zufällig verändert ("Mutation"). Aus der so entstehenden neuen Generation von Robotern wählt man wiederum die überlebensfähigsten aus und iteriert dann das geschilderte Verfahren. Schöne Beispiele von Anwendungen dieser sogenannten "genetischen Algorithmen" in der Robotik werden zum Beispiel in den Arbeiten der Robotik-Gruppe der EPFL Lausanne geschildert.¹¹

Der wesentliche Nachteil dieser Technik besteht darin, daß sie in der Praxis erfordert, die individuell Überlebensfähigkeit von Hunderten und Tausenden verschiedener Roboter-Varianten zu evaluieren. Das nimmt sehr viel Zeit in Anspruch. Daher ist es sinnvoll, diese Evolutionstechnik zu ergänzen durch schnellere Lernmethoden, die schrittweise die Überlebensfähigkeit eines *einzelnen* Roboter-Individuums verbessern können. Solche Techniken werden im nächsten Unterabschnitt skizziert.

Am Rande möchte ich anmerken, daß beide Techniken im Prinzip natürlich auch im Bereich nicht-mobiler Rechner anwendbar sind, also dort wo die "Bedrohungen" – abgesehen von einem Stromausfall – nicht physischer Natur sind, und wo "überleben" für einen Rechner bedeutet, möglichst lange eine hohe Rechenleistung ohne "Absturz" aufrecht zu erhalten.

Maschinen, die aus ihren eigenen Erfahrungen lernen

Ein wesentlicher Bestandteil des in unserer Gegenwartskultur verankerten Begriffs der Maschine ist die Vorstellung, daß das Verhalten einer Maschine insofern "trivial" ist, als sie nur das ausführen kann, was ihr von ihrem menschlichen Erbauer vorher einprogrammiert wurde. Diese Vorstellung ist veraltet, weil es in den Forschungslabors¹², in der Industrie, und sogar schon in der Unterhaltungselektronik¹³ eine Reihe von Programmen gibt, die es Rechnern ermöglichen, aus ihrer eigenen Erfahrung zu lernen. Insbesondere können solche lernenden Maschinen eine durch ihre vorhergehenden Erfahrungen geformte *Individualität* hervorbringen, die selbst von ihrem menschlichen Erbauer nicht vollkommen vorhersehbar ist. Das einzige, was der menschliche Erbauer einer solchen Maschine kennt, ist deren *Lernalgorithmus*, also das von ihm einprogrammierte Verfahren, mit dem die Maschine neue Verhaltensmuster aufgrund von vorhergehenden Erfahrungen entwickeln kann. Eine Vielfalt solcher Lernalgorithmen ist inzwischen bekannt [Weiss and Kulikowski, 1991, Mitchell, 1997]. Viele der in den Entwicklungslabors bisher erfolgreichsten Lernalgorithmen sind den Lernmechanismen von lebendigen Organismen abgeschaut worden.¹⁴ Als Beispiele erwähne ich das *Reinforcement Lernen*¹⁵ [Sutton and Barto, 1998], sowie das Lernen mittels *künstlicher Neuronaler Netzwerke* [Arbib, 1995]. Das Reinforcement Lernen ist von besonderem Interesse im Kontext des im ersten Abschnitt geschilderten alternativen Intelligenzkriteriums, weil es direkt auf dem Prinzip aufbaut, gegenwärtige Aktionen jeweil so zu wählen, daß die langfristige Belohnung (zum Beispiel: langes Überleben) maximiert wird. Ein weiterer interessanter Aspekt des Reinforcement Lernen ist die Tatsache, daß es dem Rechner, bzw. dem davon gesteuerten Roboter, beibringt, ständig abzuschätzen, wie weit er von einem vorgegebenen Arbeitsziel noch entfernt ist. Wenn man möchte, kann man das auf diese Weise erzielte Wissen der Maschine über ihre eigene Situation¹⁶ als rudimentäre Vorform eines maschinellen Bewußtseins bezeichnen.

Von der Biologie inspirierte neue Rechnerstrukturen

Die traditionelle Struktur unserer Rechner ist gekennzeichnet durch folgende Merkmale:

- Speicher und Rechnen sind getrennt
- der Rechner kann nur ausführen, was im Programm "vorhergedacht" wurde
- hierarchisch geordnet, sequentielle Arbeitsweise
- Rhythmus durch zentralen Taktgeber vorgegeben
- Kommunikation intern und extern mittels Bits
- Die Prozessoren bestehen aus Gattern, die eine geringe Zahl binärer Inputs verarbeiten.

Im Gegensatz dazu besitzt die Informationsverarbeitung in biologischen Nervensystemen die folgenden charakteristischen Strukturmerkmale:

- Speicher und Rechnen sind kombiniert in den informationsverarbeitenden Grundbausteinen (Neuronen und Synapsen)
- das Nervensystem kann aus Erfahrung lernen
- es ist überwiegend "demokratisch" strukturiert, d.h. es gibt keine zentrale Befehlsausgabe und die Neuronen und Synapsen arbeiten weitgehend autonom
- es gibt keinen zentralen Taktgeber, die zeitliche Abfolge der internen "Rechenschritte" hängt vom konkreten Input ab
- interne Kommunikation und Output mittels spike trains, also zeitlichen Mustern von Pulsen
- Neuronen erhalten durch direkte Verbindungen Signale von 5000 – 10000 anderen Neuronen.

In der Zwitterwelt zwischen diesen beiden Rechnerstrukturen sind die *künstlichen Neuronalen Netze* angesiedelt. Diese folgen in ihrer Architektur eher den Prinzipien biologischer Systeme, können aber in jedem üblichen digitalen Rechner simuliert, oder direkt in neuartiger elektronischer hardware implementiert werden.¹⁷ Ihre Merkmale sind die folgenden:

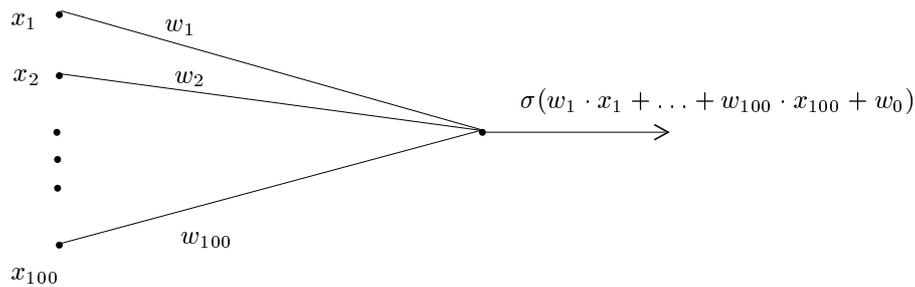


Abbildung 8.1. Schematische Arbeitsweise eines künstlichen Neurons. Der Input besteht in diesem Fall aus 100 Input Zahlen x_1, \dots, x_{100} (wobei die Zahl 100 willkürlich gewählt wurde; ein biologisches Neuron erhält bis zu 10 000 Input Zahlen). Der Output besteht aus einer einzigen Zahl zwischen 0 und 1, die in der Zeichnung mit $\sigma(w_1 \cdot x_1 + \dots + w_{100} \cdot x_{100} + w_0)$ bezeichnet ist. Sie entsteht also dadurch, daß man zunächst die gewichtete Summe $w_1 \cdot x_1 + \dots + w_{100} \cdot x_{100} + w_0$ berechnet, und auf diese Zahl dann die Quetschfunktion σ anwendet.

- Sie bestehen aus künstlichen Neuronen, die mittels geeigneter Computer-Programme oder sogar mittels dafür entwickelter neuer elektronischer Hardware in einem Computer simuliert werden.
- Memory und Rechnen sind kombiniert in jedem künstlichen Neuron
- das Netzwerk kann aus Erfahrung lernen (mittels einem "Lernalgorithmus")
- es ist überwiegend "demokratisch" strukturiert und hat eine parallele Arbeitsweise
- Rhythmus durch zentralen Taktgeber vorgegeben
- Kommunikation mittels Bits (oder Zahlen).

Die Arbeitsweise eines künstlichen Neurons ist recht einfach. Wir betrachten als Beispiel ein künstliches Neuron mit 100 Input-Variablen x_1, \dots, x_{100} , die Bits oder reelle Zahlen repräsentieren können (siehe Abbildung 8.1).

Der Output $\sigma(S)$ eines künstlichen Neurons entsteht dadurch, daß es zunächst eine geeignete gewichtete Summe

$$S = w_1 \cdot x_1 + w_2 \cdot x_2 + \dots + w_{100} \cdot x_{100} + w_0$$

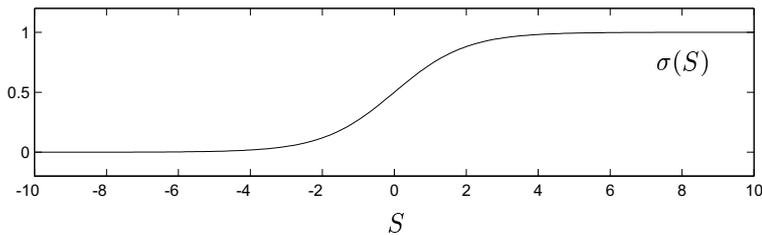


Abbildung 8.2. Quetschungsfunktion σ (die skizzierte Funktion ist die Funktion $\sigma(S) = 1/(1 + e^{-S})$).

der Inputs x_1, \dots, x_{100} ausrechnet, und dann eine sogenannte "Quetschungsfunktion" σ auf diese gewichtete Summe S anwendet (zum Beispiel $\sigma(S) = 1/(1 + e^{-S})$). Die Gewichte w_0, \dots, w_{100} sind Zahlen, die das gelernte Wissen des künstlichen Neurons repräsentieren. Zum Beispiel, w_{17} ist eine große positive Zahl, wenn das Neuron die "Erfahrung gemacht hat, daß es einen relativ großen Output-Wert geben sollte, falls sein Input x_{17} groß ist, und w_{51} ist eine negative Zahl, wenn ein hoher Wert von x_{51} zu einem niedrigen Output-Wert des Neurons führen soll (siehe den folgenden Unterabschnitt). Der konstante Term w_0 legt gewissermaßen die "Reizschwelle" des künstlichen Neurons fest, also er legt fest wie hoch die Summe $w_1 \cdot x_1 + \dots + w_{100} \cdot x_{100}$ werden muß, damit der Output $\sigma(S)$ einen hohen Wert ergibt.

Die nichtlineare Quetschungsfunktion σ transformiert die oft stark positive oder stark negative Zahl S in eine Zahl $\sigma(S)$ zwischen 0 und 1 (Änderung des Maßstabs), siehe Abbildung 8.2. Diese Änderung des Maßstabs ist *nicht-linear*, weil die Randbereiche mehr zusammengestaucht werden als der Mittelbereich.

Wie verbindet man die "Neuronen" in einem künstlichen Neuronalen Netz?

Ein einzelnes künstliches Neuron kann, selbst bei beliebiger Wahl der Gewichte w_1, \dots, w_n und der Reizschwelle w_0 , nur ein beschränktes Repertoire von Funktionen berechnen. Wenn man diesen Berechnungsvorgang aber iteriert, also den Output von einer Gruppe von Neuronen zum Input eines darauffolgenden weiteren Neurons macht (welches in der Regel andere Gewichte verwendet), so kann das so entstandene Netzwerk von künstlichen

Neuronen (siehe Abbildungen 8.3 und 8.4) praktisch beliebig komplizierte Abbildungen von Inputs zu Outputs berechnen, oder vielmehr zu berechnen "lernen", wie wir im folgenden Unterabschnitt sehen.

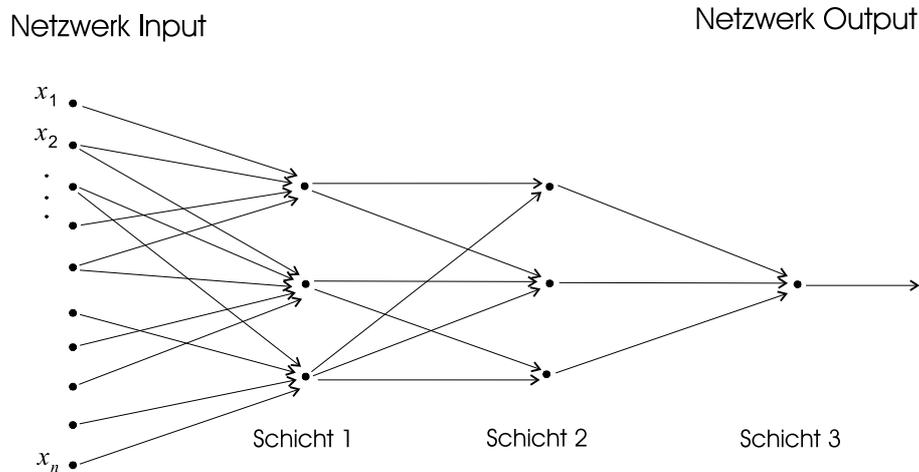


Abbildung 8.3. Verbindungsstruktur eines typischen in einer Richtung ausgerichteten ("feedforward") neuronalen Netzes. Die Richtung der Berechnung in diesem neuronalen Netz geht von links nach rechts. Jeder Punkt auf den Schichten 1,2,3, zusammen mit den in diesen Punkt gerichteten Pfeilen, symbolisiert ein künstliches Neuron wie in Abb. 8.1. Die Inputs der Neuronen auf Schicht 1 bestehen aus Netzwerk Inputs x_1, \dots, x_n . Im 1. Zeitschritt berechnen alle Neuronen auf Schicht 1 gleichzeitig ("parallel") ihre jeweilige Output Zahl. Die drei von den Neuronen auf Schicht 1 berechneten Output Zahlen werden weiter verarbeitet von Neuronen auf Schicht 2, d.h. sie liefern die Inputs für Neuronen auf Schicht 2. Die Neuronen auf Schicht 2 errechnen ihre Output Zahlen während dem 2. Zeitschritt, und aus diesen berechnet dann das Neuron auf Schicht 3 seinen Output, der gleichzeitig den Output des gesamten neuronalen Netzes (bestehend aus 7 Neuronen) bildet.

Woher kommen die "Gewichte" in einem künstlichen neuronalen Netz?

Die Gewichte und Reizschwellen w_i der künstlichen Neuronen in einem neuronalen Netz sind Zahlen, die das "Programm" eines neuronalen Netzes enthalten, also seine Arbeitsweise festlegen. Anstatt die Werte der Gewichte und Reizschwellen fest einzuprogrammieren, überläßt man es dem neuronalen Netz, deren Werte selbst so zu wählen, daß es das gewünschte Input/Output-Verhalten zeigt. Dazu wird es mit Trainingsbeispielen für

Netzwerk Input

Netzwerk Output

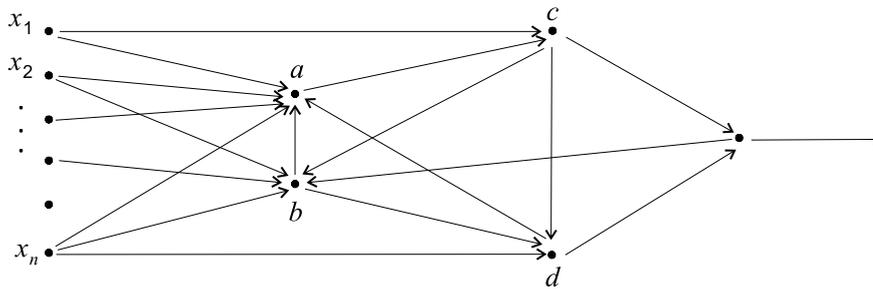


Abbildung 8.4. Beispiel für ein Neuronales Netz mit einer komplexeren Verbindungsstruktur. Der Netzwerk Input ist wieder links abgebildet, und der Netzwerk Output ist wieder rechts. Intern ist der Informationsfluß aber komplizierter. Das Neuron a erhält als Inputs nicht nur Netzwerk Inputs, sondern auch die Outputs der Neuronen b und d . Die Neuronen b und d bekommen als Input unter anderem den Output von Neuron c , das wiederum den Output von Neuron a als Input bekommt. Wegen der vielfältigen "Rückkopplung" im Netz ist es nicht mehr so einfach zu beschreiben, wie die einzelnen Rechenschritte im Netz koordiniert werden. Komplizierte Netzwerk-Strukturen mit "Rückkopplungen" sind typisch für biologische Netzwerke von Neuronen, werden aber für eine große Anzahl von Anwendungen auch künstlich im Rechner simuliert.

"gute" Kombinationen von Inputs und Outputs gefüttert. Mittels einem einprogrammierten Lernalgorithmus kann es diese Trainingsbeispiele benutzen um seine Gewichte und Reizschwellen, und damit seine Arbeitsweise, in geeigneter Weise zu verändern.

Man verwendet unter anderem die folgenden Design Ideen bei der Konstruktion von Lernalgorithmen für künstliche neuronale Netze:

- Ermittle für jedes Gewicht in welcher Richtung man es verändern soll, damit das Verhalten des Neuronalen Netzes "in die richtige Richtung" verändert wird
- Geduld: Viele richtige kleine Schritte bei Änderungen von Gewichten und Reizschwellen führen auch zum Ziel
- Sei gelegentlich "kreativ", z.B. probier einfach aus, was passiert, wenn man einem Gewicht einen ganz anderen Wert gibt (→ Zufallselemente).

Was leistet ein künstliches Neuronales Netz?

Mit genügend vielen Trainingsbeispielen kann ein Neuronales Netz zum Beispiel lernen

- handgeschriebene Zeichen zu lesen
- dem Arzt bei der Diagnose von Krankheiten zu helfen (z.B. Früherkennung von Brustkrebs)
- einen Motor so zu steuern, daß der Ausstoß von Schadstoffen minimiert wird
- einen Roboter so zu steuern, daß er Kollisionen mit Hindernissen vermeidet.

Abgesehen von ihrer Lernfähigkeit haben Neuronale Netze den Vorteil, daß alle praktisch wichtigen Funktionen mittels in einer Richtung ausgerichteten Netzen bestehend aus nur 2 Schichten (und daher in nur 2 parallelen Rechenschritten) mit nicht zu vielen künstlichen Neuronen berechnet werden können. Daher ergeben Neuronale Netze gleichzeitig ein sehr nützliches mathematisches Modell für Rechnen in "Echtzeit".¹⁸

Wie unterscheidet sich ein künstliches Neuronales Netz von seinem Vorbild in der Natur?

Wenn man die Arbeitsweise und Leistung der gegenwärtigen Generation von künstlichen Neuronalen Netzen mit der von biologischen Nervensystemen vergleicht, stellt sich heraus, daß sie weiter voneinander entfernt sind, als man früher gedacht hatte. Ein wesentlicher Unterschied zwischen biologischen und künstlichen Neuronen liegt in der Struktur ihres Outputs. Der Output eines biologischen Neurons besteht aus elektrischen Pulsen ("spikes"), die in unregelmäßigen Abständen, ca 1–100 mal pro Sekunde, ausgesendet werden. Wenn man die Zeiten protokolliert, wann ein biologisches Neuron einen spike aussendet, so schaut ein solcher "spike train" so aus wie eine der Zeilen von Abbildung 8.5. In Abbildung 8.5 wurden die spike trains von 30 (ziemlich willkürlich ausgewählten) Neuronen für einen Zeitraum von 4 Sekunden aufgezeichnet. Wenn wir zum Beispiel alle Informationen protokollieren würden, die unser Gehirn innerhalb von 4 Sekunden von unseren Augen erhält, so würde eine ähnliche Abbildung mit 1.000.000 Zeilen (anstatt 30) entstehen, weil alle visuellen Eindrücke in

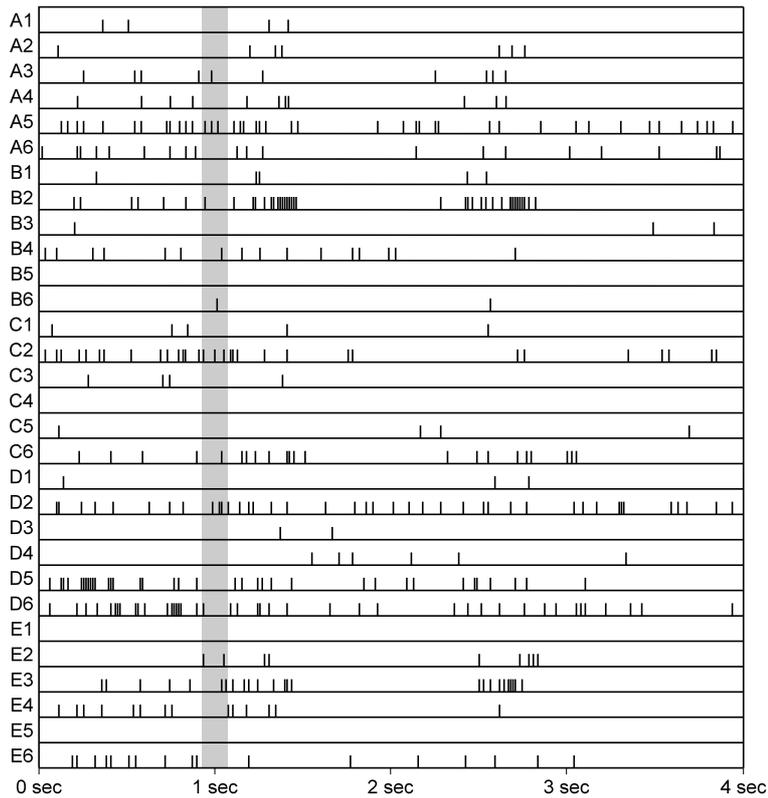


Abbildung 8.5. Aufnahme der Feuerzeiten von 30 Neuronen in der visuellen Kortex des Affen über 4 Sekunden [Krüger and Aiple, 1988]. Die 30 Neuronen wurden von den Experimentatoren mit A1 bis E6 bezeichnet (siehe linke Spalte der Abbildung). Das Feuerverhalten von jedem der 30 Neuronen ist jeweils in einer Zeile protokolliert. Die waagrechte Achse ist die Zeitachse. Jede Feuerzeit ist durch einen senkrechten Strich angezeigt. Man bezeichnet die jeweils in einer Zeile festgehaltene Folge von Feuerzeichen eines Neurons als "spike train". Zum Vergleich mit der Zeitdauer einer typischen Berechnung in einem Nervensystem haben wir ein Zeitintervall von 150 Millisekunden grau markiert. Innerhalb dieser Zeitdauer können biologische Nervensysteme (bestehend aus 10 und mehr Schichten) komplexe Aufgaben der Mustererkennung ausführen. Abdruck mit freundlicher Genehmigung von The American Physiological Society.

der Retina in die spike trains von 1.000.000 Neuronen kodiert und in dieser Form über den optischen Nerven an das Gehirn weitergeleitet werden.

Man hat früher gemeint, daß das einzige relevante Signal im Output eines biologischen Neurons die *Häufigkeit* seines Feuerns ist, und daß diese

Häufigkeit dem Output $\sigma(S)$ eines künstlichen Neurons entspricht. Man sieht aber sofort aus Abbildung 8.5, daß sich die momentane Häufigkeit des Feuerns eines biologischen Neurons ständig ändert, und daß die zeitlichen Abstände zwischen dem Feuern eines Neurons viel zu unregelmäßig sind um aus 2 oder 3 spikes eine gute Abschätzung von dessen gegenwärtiger Häufigkeit des Feuerns zu ermöglichen. Neuere experimentelle Untersuchungen (siehe zum Beispiel [Rieke et al., 1997, Koch, 1999, Recce, 1999]) zeigen vielmehr, daß das gesamte raum-zeitliche Muster des Feuerns von biologischen Neuronen für deren Informationsverarbeitung relevant ist. Man kann den Output eines Systems von biologischen Neuronen also eher mit einem von einem großen Orchester gespielten Musikstück vergleichen, zu dessen Wiedererkennung es nicht ausreicht zu wissen *wie oft* jedes Instrument gewisse Töne spielt. Charakteristisch für ein Musikstück ist vielmehr, wie jeder Ton in eine Melodie oder in einen Akkord eingebettet ist. Man nimmt an, daß in ähnlicher Weise viele Gruppen von Neuronen in biologischen Systemen die von ihnen ausgesendeten Informationen durch das zeitliche Muster kodieren, in dem jedes Neuron in der Gruppe *relativ zu den anderen* feuert. Daher ist die Kommunikationsweise innerhalb unseres Gehirns einem Musikstück sehr viel ähnlicher als die von der gegenwärtigen Generation von Computern bevorzugte Kommunikationsweise.

Die Untersuchung der theoretischen und praktischen Möglichkeiten mittels raum-zeitlicher Muster von Pulsen zu rechnen und zu kommunizieren, hat in den letzten Jahren eine neue Generation künstlicher Neuronaler Netzwerke entstehen lassen: *pulsbasierte* künstliche Neuronale Netze.¹⁹ In diesen wird die künstliche Synchronisation der traditionellen künstlichen Neuronalen Netzen ersetzt durch die der Biologie abgeschauten Methode, Informationen in raum-zeitlichen Mustern (analog wie bei einem Handzeichen oder einer Melodie) zu kodieren. Dies eröffnet für den Informatiker eine faszinierende neue Welt, nämlich die Möglichkeit *Zeit* als eine bisher in unseren Rechnern brachliegende Dimension zu erschließen, und direkt *mit raum-zeitlichen Mustern zu rechnen* [Maass, 1999, Maass and Sontag, 1999].²⁰

Erfreulicherweise kann man diese Strategie auch relativ leicht in neu entwickelter elektronischer Hardware anwenden [Murray, 1999].²¹ Diese Kodierungsstrategie implementiert eine faszinierende Idee von [Mead, 1989]: "to let time represent itself", d.h. anstatt die zeitliche Struktur der Inputs (zum Beispiel visuelle Informationen über bewegte Objekte) durch angehängte künstliche "time-stamps" – vergleichbar den Eingangsstempeln in einem Büro – zu kodieren, bleibt die zeitliche Struktur des Inputs während

des "Rechnens" in der raum-zeitlichen Struktur der spike trains repräsentiert. In einer Analogie kann man diese Vorgangsweise vergleichen mit der Arbeitsweise in der Küche eines (noch nicht vollständig digitalisierten) Restaurants. Dort werden die vom Ober aufgenommenen Bestellzettel für die Köche gut sichtbar an einem Rad (oder gelegentlich an einer Schiene) befestigt. In diesem Fall kodiert die Position jedes einzelnen Bestellzettels relativ zu den anderen die Zeit seines Eingangs. Auch ohne Eingangsstempel ("time stamps") auf den Bestellzetteln können bei dieser raum-zeitlichen Organisationsstruktur die Bestellungen in der richtigen Reihenfolge ausgeführt werden. Der Ansatz, daß man traditionelle Rechenschritte ersetzt durch geeignete Manipulationen von raum-zeitlichen Strukturen, eröffnet neue Möglichkeiten um zum einen Energie-effiziente parallele Rechner zur Verarbeitung komplexer Information in Echtzeit zu entwerfen, und zum anderen möglicherweise ein wenig besser zu verstehen wie komplexe raum-zeitliche Informationen im menschlichen Gehirn kodiert und verarbeitet werden.²²

Konklusion

Ich habe die Frage nach der Informationsverarbeitung im menschlichen Gehirn bisher ausgeklammert. Dies geschah zum einen deshalb, weil ich kein Experte für dieses Gebiet bin, zum anderen, weil Gespräche mit Neurophysiologen und Neurobiologen deutlich machen, daß über die Informationsverarbeitung im menschlichen Gehirn sehr viel weniger gesichertes Wissen vorliegt, als in der typischen Populär-Literatur suggeriert wird. In manchen populärwissenschaftlichen Artikeln liest man, daß es schon jetzt möglich wäre, mittels geeigneter Apparate "Gedanken zu lesen" im menschlichen Gehirn, während andere Artikel das bestreiten. Darüberhinaus vermitteln zahlreiche Veröffentlichungen in wissenschaftlichen Fachzeitschriften den Eindruck, daß die wesentlichen Rätsel bezüglich der Informationsverarbeitung in biologischen Nervensystemen bereits gelöst seien. Leider gibt es zu fast allen solchen Veröffentlichungen andere, genauso seriöse Veröffentlichungen, die behaupten, dieselben Rätsel mit vollkommen anderen Antworten gelöst zu haben, oder daß diese Rätsel noch ungelöst wären.

Dieser verwirrende Zustand wird verständlich wenn man sich die experimentellen Ergebnisse genauer anschaut, die diesen Artikeln zugrunde liegen. Wir erleben gegenwärtig gewaltige Fortschritte bei *nichtinvasiven* Methoden zur Messung von Gehirnaktivitäten, wie zum Beispiel EEG, PET, MRI (für allgemeinverständliche Beschreibungen dieser Techniken siehe [Schmidt, 1993, Purves et al., 1997]). Mit diesen Methoden erhält man

raum-zeitliche Muster von Gehirnaktivitäten, wobei aber die räumliche Auflösung so gering ist, daß man nur Informationen über das durchschnittliche Aktivitätsniveau von neuronalen Schaltkreisen bestehend aus Milliarden von Neuronen erhält, aber nicht erfährt wie dort im Einzelnen Informationen kodiert und verarbeitet werden. Man kann aber mittels solcher Methoden zum Beispiel feststellen, ob die Gehirnregion, die für die Steuerung der rechten Hand zuständig ist, überdurchschnittlich aktiviert ist. Da die für die Steuerung der rechten Hand zuständige Gehirnregion schon dadurch aktiviert werden kann, daß die betreffende Person nur intensiv daran *denkt* die rechte Hand zu bewegen, kann man diesen *Gedanken* in der Regel im EEG der betreffenden Person ablesen. Es ist aber nicht vorstellbar, daß man mit solchen Apparaten abstraktere Gedanken eines Menschen lesen kann, also zum Beispiel an welche Zahl eine Person gerade denkt; oder welche Partei sie am kommenden Wahlsonntag wählen möchte.

Mittels *invasiver* Methoden erhält man bei Tieren zusätzlich Informationen über das genaue Feuerverhalten einer kleineren Anzahl von Neuronen – siehe zum Beispiel Abbildung 8.5. Aber hier hat man das Problem, daß unbekannt bleibt, welchen *Input* diese Neuronen gerade von ihren bis zu 10 000 “Vorgänger“-Neuronen bekommen, und welche intrazellulären biochemischen Substanzen ihr gegenwärtiges Verhalten beeinflussen. Daher erhält man bei diesen Experimenten nur in recht indirekter Weise Informationen über die von diesen Neuronen ausgeführten Berechnungen. In wirbellosen Tieren gelingt es gelegentlich, *einzelne* Neuronen zu identifizieren, die eine Schlüsselrolle bei der internen Kodierung von sensorischem Input (also zum Beispiel von dem, was das Auge sieht) spielen. In einzelnen Fällen, wie zum Beispiel bei den beiden H1-Neuronen in der Fliege, konnte sogar der “Code“ geknackt werden, mit dem diese Information im Feuerverhalten dieser Neuronen kodiert ist [Rieke et al., 1997, Recce, 1999].

Leider ist es bisher nicht einmal bei relativ einfachen Tieren gelungen, den “Neuronalen Kode“ zu identifizieren, mit dem *Zwischenergebnisse* bei der weiteren Verbreitung der sensorischen Informationen kodiert sind, sodaß selbst hier der Zugang zum Verständnis der eigentlichen Informationsverarbeitung in den darauffolgenden neuronalen Schaltkreisen versperrt bleibt. Bei höheren Tieren ist die interne neuronale Kodierung noch sehr viel komplexer, weil Informationen in verteilter Weise kodiert sind, also aufgespalten über sehr große Gruppen von Neuronen.

Ein Gedankenexperiment hilft uns vielleicht, die Schwierigkeit zu verstehen, aus experimentellen Daten der vorher geschilderten Art auf die Organi-

sation der Informationsverarbeitung im menschlichen Gehirn zu schließen. Nehmen wir einmal an, ein künstlicher Rechner der gegenwärtigen Generation wäre aufgrund eines time-warps versehentlich im Jahr 1920 "vom Himmel gefallen", und man hätte in einem Forschungslabor mittels aller damals bekannten Meßmethoden zu ermitteln versucht, wie die Informationsverarbeitung in dieser Maschine funktioniert. Mittels Messungen der Temperatur an verschiedenen Stellen dieser Maschine (welche Informationen liefern über den lokalen Energieverbrauch, analog zu PET und MRI beim Gehirn) und Messungen elektrischer Felder (analog zum EEG) hätte man vielfältige Aktivitätsmuster und deren Abhängigkeit von Eingaben über die Tastatur studieren können. Zusätzlich hätte man vielleicht auch das zeitliche Muster des Schaltens einzelner Transistoren (entsprechend dem Feuerverhalten einzelner Neuronen im Gehirn) ermitteln können. Trotzdem glaube ich nicht, daß man mittels solcher "bottom-up Methoden" die Struktur des Betriebssystems dieses Computers hätte ermitteln können. Vielmehr hätte man diese bottom-up Methoden ergänzen müssen durch top-down Modelle, bei denen man fortschreitend feinere und adäquatere Hypothesen über die Organisationsstruktur des Betriebssystems entwickelt, und in künstlichen Maschinen zur Informationsverarbeitung erprobt. In anderen Worten: man hätte die Funktionsweise des im Jahr 1920 vom Himmel gefallenen Rechners nur dann ermitteln können, wenn man begleitend zu den experimentellen Untersuchungen wesentliche Teile der Rechner-Theorie vorausgedacht und an entsprechend konstruierten Prototypen *ausprobiert* hätte.

Im Vergleich zu diesem Gedankenexperiment ist das Problem der Entschlüsselung der Arbeitsweise des menschlichen Gehirns ungleich größer. Anstatt durch begrifflich strukturierte Design-Prinzipien von menschlichen Ingenieuren ist das menschliche Gehirn durch die Evolution, also im Laufe einer langen Kette von größtenteils zufallsgesteuerten trial-and-error Experimenten geformt worden. Daher ist es hier noch schwieriger als im vorhergehenden Gedankenexperiment, in direkter Weise experimentell zu ermitteln, wie Berechnungen strukturiert sind. Deshalb muß man auch im Bereich der Neurowissenschaften die bottom-up Methode durch top-down Ansätze ergänzen, bei denen man versucht die vielen bruchstückhaften aus Experimenten gewonnenen Erkenntnisse, Indizien und Vermutungen mittels *theoretischer Modelle und Computersimulationen* zu globalen Hypothesen über die Struktur neuronaler Berechnungen in konkreten Lebewesen zu kombinieren. Jede solche globale Hypothese kann gleichzeitig aufgefaßt werden als grober Bauplan für die Organisation eines (möglicherweise stark speziali-

sierten) künstlichen Rechners, und die Tragfähigkeit solcher Hypothesen kann eigentlich nur dadurch ermittelt werden, daß man solch einen neuartigen künstlichen Rechner baut.

Man sieht also, daß die Entschlüsselung der Informationsverarbeitung im menschlichen Gehirn und die Entwicklung intelligenter Rechner Hand in Hand gehen. In der Tat sind alle im zweiten Abschnitt diskutierten innovativen Ideen zur Entwicklung "intelligenter" Rechner aus dem Bemühen, Informationsverarbeitung in konkreten biologischen Nervensystemen zu verstehen (siehe [Arbib, 1995]), entstanden. Aus dem vermeintlichen Gegensatz zwischen unseren Erkenntnissen über das menschliche Gehirn und dem Entwurf von leistungsfähigen künstlichen Rechnern entsteht also bei genauerem Hinschauen eine Symbiose von zwei sehr verschiedenartigen Wissensbereichen, wobei Fortschritte im Verständnis des ersten Bereichs untrennbar verbunden sind mit innovativen technischen Ideen im zweiten Bereich. In anderen Worten: *Je mehr wir von der Organisation der Informationsverarbeitung im menschlichen Gehirn verstehen werden, umso "gehirnartiger" werden gleichzeitig die besten künstlichen Rechner werden, und umso weniger befremdlich wird die Vorstellung sein, daß das menschliche Gehirn "nur" ein Rechner sei.*

Nachsatz: Ich möchte Peter Auer, Margot Goettsberger, Gerold Muhr und Thomas Natschläger für hilfreiche Kommentare zum ersten Entwurf dieses Aufsatzes danken.

Anmerkungen

- ¹ Eine genauere Betrachtung ergibt ein differenziertes Bild. Vom Gesichtspunkt der Computerindustrie ist es logisch und sinnvoll die Rechnerentwicklung so anzulegen, daß der Kunde alle paar Jahre seine software und hardware ersetzen muß. Unter diesem Gesichtspunkt wäre es kontraproduktiv, den Schwerpunkt in der Rechnerentwicklung umzustellen auf die Entwicklung von Rechnern die sich selbständig auf Änderungen in ihrer Umgebung einstellen und ihre eigene Leistungsfähigkeit weiterentwickeln, und dabei möglicherweise sogar eine "Individualität" entwickeln, an der der Kunde hängt.
- ² Die Turing Maschine ist ein von Turing vorgeschlagenes mathematisches Rechnermodell, das "universell" ist in dem Sinn, daß es jedes andere Modell für einen digitalen und deterministischen Rechner simulieren kann (wobei die Turing Maschine aber wesentlich mehr Rechenschritte benutzen darf als der simulierte Rechner), siehe [Sipser, 1997, DePauli und Weibel, 1997].
- ³ P ist ein mathematisches Modell für die Klasse aller tatsächlich auf einem Computer lösbarer Berechnungsprobleme. Ein Problem ist dann in der Klasse P wenn es einen Algorithmus zur Lösung dieses Problems gibt, und zusätzlich die benötigte Rechnerzeit nur polynomiell (also nicht "explosionsartig" wie bei der Exponentialfunktion $n \mapsto 2^n$) mit der Input-Länge n anwächst. NP ist die Klasse aller auf einer fiktiven "nichtdeterministischen" Turing Maschine in polynomieller Rechenzeit lösbarer Probleme. Von vielen für die Praxis besonders wichtigen Berechnungsprobleme, zum Beispiel aus den Bereichen Operations Research, Scheduling,

Kryptographie, ist bekannt, daß sie zu der Klasse NP gehören. Weil nur $P \subseteq NP$ bekannt ist, besagt das aber nicht, daß diese Probleme tatsächlich auf einem Computer lösbar sind, also zur Klasse P gehören.

⁴ Ein weiterer Grund weshalb Turing Maschinen keinen geeigneten Vergleichsmaßstab für das menschliche Gehirn liefern ist ein rein mathematischer: Turing Maschinen sind als mathematisches Rechnermodell nur dann von Interesse, wenn beliebig lange Bit-Folgen als Inputs zu verarbeiten sind, weil sie sich sonst wie endliche Automaten verhalten. Es haben aber alle durch ein menschliches Gehirn verarbeiteten Inputs eine durch eine fixe Zahl beschränkte Anzahl von Bits.

⁵ Wir bezeichnen hier mit Hintergrundwissen das bei genauerem Hinsehen sehr umfangreiche Wissen, das uns – ohne daß wir uns dessen in der Regel bewußt werden – hilft, zielführende Entscheidungen zu treffen. Um ein ganz einfaches Beispiel zu nennen: wenn wir die Route planen, um einen Raum zu durchqueren, benutzen wir unbewußt vielfältiges Hintergrundwissen, zum Beispiel, daß Möbel nicht wirklich zweidimensional sind, wie sie auf unserer Retina erscheinen, daß sich Beleuchtung, Fenster- und Türstellungen schnell ändern können, aber Möbel keine spontanen Sprünge ausführen, und daß wir etwaige Stufen am Boden aber keine Kaffeetische betreten dürfen. Das klingt alles recht einfach, aber erklären Sie dieses Hintergrundwissen einmal Ihrem PC, ohne sich in Widersprüche oder Mehrdeutigkeiten zu verwickeln!

⁶ Die hier angesprochene Schwierigkeit bewirkt interessanterweise, daß das berühmte Gödel'sche Theorem über die Unbeweisbarkeit der Widerspruchsfreiheit eines formalen Systems S mit Mitteln dieses formalen Systems S kein zusätzliches Hindernis für praktische maschinelle Intelligenz bedeutet. Die in diesem Kontext relevanten formalen Systeme S sind sowieso nicht widerspruchsfrei.

⁷ <http://www.ai.mit.edu/people/brooks/>

⁸ Sein Buch "Vehicles: Experiments in Synthetic Psychology" [Braitenberg, 1984] ist noch heute überaus lesenswert und anregend. Eine Zusammenfassung der wichtigsten Abschnitte ist online erhältlich von <http://www.cis.tu-graz.ac.at/igi/STIB/WS98/gruppe3/welcome.html>.

⁹ Man denke zum Beispiel an die Konflikte zwischen Es, Ich und Überich in der Psychoanalyse von Sigmund Freud.

¹⁰ <http://www.ai.eecs.umich.edu/cogarch0/subsump/index.html>

¹¹ <http://lamiwww.epfl.ch/lami/team/mondada/index.html>;
<http://diwww.epfl.ch/lami/team/floreato/>

¹² http://www.cs.bham.ac.uk/~anp/ai_ml.html;
<http://www.sgi.com/Technology/mlc/>; <http://www.ics.uci.edu/AI/ML/MLDBRepository.html>; <http://www.cis.tu-graz.ac.at/igi/STIB/WS98/>

¹³ <http://www.creatures2.com>

¹⁴ Anzumerken ist, daß gesicherte und allgemein anerkannte Kenntnisse über neurobiologische Mechanismen des Lernens im menschlichen Gehirn kaum vorhanden sind, und daß es sich bei den bisher in künstlichen Rechnern realisierten Lernmethoden um Lernmechanismen relativ einfacher Tiere handelt (siehe [Arbib, 1995]).

¹⁵ genauer gesagt: über den Wert ihres gegenwärtigen Zustands hinsichtlich der von ihr angestrebten Ziele

¹⁶ <http://www.cs.brown.edu/people/lpk/rl-survey/rl-survey.html/>

- ¹⁷ Carver Mead, Professor am California Institute of Technology und maßgeblicher Pionier bei der Entwicklung der all unseren gegenwärtigen Rechnern zugrunde liegenden integrierten Schaltkreisen, hat für diese Entwicklungsrichtung den Namen "neuromorphic engineering" geprägt, und in seinem Buch [Mead, 1989] zugleich einen der wichtigsten Beiträge hierzu geliefert.
- ¹⁸ Eine schöne Einführung in künstliche Neuronale Netze ist das Tutorial von Gerstner "Supervised Learning for Neural Networks: A Tutorial with JAVA exercises": http://diwww.epfl.ch/lami/team/gerstner/wg_pub.html
- ¹⁹ [Maass and Bishop, 1999] enthält Übersichtsartikel zum gegenwärtigen Stand der Forschung über puls-basierte Neuronale Netze.
- ²⁰ Einige Arbeiten zu diesem Thema sind online erhältlich von <http://www.cis.tu-graz.ac.at/igi/maass/Welcome.html>.
- ²¹ Im Prinzip kann man sich auch damit behelfen, daß man so einen neuartigen Rechner auf einem herkömmlichen digitalen Rechner simuliert [Jahnke et al., 1999]. Praktisch scheitert dies oft daran, daß die erforderliche Rechenzeit zu groß ist. Daher schlägt man stattdessen an der ETH Zürich den von Carver Mead vorgezeichneten Weg ein und baut eine analoge "silicon cortex" [Deiss et al., 1999].
- ²² Für weitere Informationen über puls-basierte Neuronale Netze verweise ich auf die allgemeinverständliche Einführung von Natschläger in deutscher Sprache: <http://www.cis.tu-graz.ac.at/igi/tnatschl/3gen/3genlang.html> sowie auf die detaillierten Übersichtsartikel "Spiking neurons" von Gerstner http://diwww.epfl.ch/lami/team/gerstner/wg_pub.html sowie auf meinen Übersichtsartikel "Paradigms for Computing with Spiking Neurons" von <http://www.cis.tu-graz.ac.at/igi/maass/#Publications>.
Einige links zu weiterer Literatur, Forschungsberichten und Simulatoren für Neuronale Netzwerke findet man unter <http://www.cis.tu-graz.ac.at/igi/maass/#SomeLinks>, http://diwww.epfl.ch/w3mantra/mantra_links.html, http://diwww.epfl.ch/w3mantra/mantra_journals.html.

Literatur

- [Arbib, 1995] Arbib, M. A., editor (1995). *The Handbook of Brain Theory and Neural Networks*. MIT-Press (Cambridge, MA).
- [DePauli und Weibel, 1997] DePauli-Schimanovich, W., Weibel, P. (1997). *Kurt Gödel: ein Mathematischer Mythos*. Hölder-Pichler-Tempsky (Wien).
- [Deiss et al., 1999] Deiss, S. R., Douglas, R. J., and Whatley, A. M. (1999). A pulse-coded communications infrastructure for neuromorphic systems. In Maass, W. and Bishop, C., editors, *Pulsed Neural Networks*. MIT-Press, Cambridge.
- [Jahnke et al., 1999] (1999). Digital Simulation of Spiking Neural Networks. In Maass, W. and Bishop, C., editors, *Pulsed Neural Networks*. MIT-Press (Cambridge, MA).

- [Koch, 1999] Koch, C. (1999). *Biophysics of Computation: Information Processing in Single Neurons*. Oxford University Press (Oxford).
- [Krüger and Aiple, 1988] Krüger, J., and Aiple, F. (1988). Multielectrode investigation of monkey striate cortex: Spike train correlations in the infragranular layers. *Neurophysiology*, 60:798–828.
- [Maass, 1999] Maass, W. (1999). Computing with spiking neurons. In Maass, W. and Bishop, C., editors, *Pulsed Neural Networks*. MIT-Press (Cambridge, MA).
- [Maass and Bishop, 1999] Maass, W., and Bishop, C., editors (1999). *Pulsed Neural Networks*. MIT-Press (Cambridge, MA).
- [Maass and Sontag, 1999] Maass, W., and Sontag, E. D. (1999). Neural systems as nonlinear filters. *Neural Computation*, in press.
- [Mead, 1989] Mead, C. (1989). *Analog VLSI and Neural Systems*. Addison-Wesley (Reading).
- [Mitchell, 1997] Mitchell, T. M. (1997). *Machine Learning*, McGraw-Hill (New York).
- [Murray, 1999] Murray, A. F. (1999). Pulse-based computation in VLSI neural networks. In Maass, W. and Bishop, C., editors, *Pulsed Neural Networks*. MIT-Press (Cambridge, MA).
- [Picard, 1997] Picard, R. W. (1997). *Affective Computing*, MIT Press (Cambridge, MA).
- [Purves et al., 1997] Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., La Mantia, A.-S., McNamara, J. O., eds. (1997). *Neuroscience*, Sinauer Associates (Sunderland, MA).
- [Recce, 1999] Recce, M. (1999). Encoding Information in Neuronal Activity. In Maass, W. and Bishop, C., editors, *Pulsed Neural Networks*. MIT-Press (Cambridge, MA).
- [Rieke et al., 1997] Rieke, F., Warland, D., Bialek, W., and de Ruyter van Steveninck, R. (1997). *SPIKES: Exploring the Neural Code*. MIT-Press (Cambridge, MA).

- [Schmidt, 1993] Schmidt, R. F. (1993). *Neuro- und Sinnesphysiologie*, Springer (Berlin).
- [Sipser, 1997] Sipser, M. (1997). *Introduction to the Theory of Computation*, PWS Publishing Company (Boston, MA).
- [Sutton and Barto, 1998] Sutton, R. S., Barto, A. G. (1998). *Reinforcement Learning*, MIT Press (Cambridge, MA).
- [Turing, 1950] Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59, no. 236.
- [Weiss and Kulikowski, 1991] Weiss, S. M., Kulikowski, C. A. (1991). *Computer Systems that Learn*, Morgan Kaufmann (San Mateo, CA)

Brain Drain oder Internationalisierung: Überlegungen zur Situation der Mathematik und Theoretischen Informatik in Österreich

Wolfgang Maass und Peter Weibel

Es ist allgemein bekannt, daß Wien vor dem Zweiten Weltkrieg eines der international führenden Zentren für die Wissenschaft des formalen Denkens gewesen war. Von der darauffolgenden Vertreibung der besten Köpfe unter dem Nationalsozialismus hat sich Österreich bis heute nicht ganz erholt.

Viele der aus Österreich vertriebenen Gelehrten waren maßgeblich am rasanten Aufstieg der formalen Wissenschaften in den USA und England beteiligt und haben dort hervorragende Forschungszentren im Bereich der Kultur und Naturwissenschaften aufgebaut. Darum gehen auch heute noch junge WissenschaftlerInnen gerne ins Ausland, nicht weil sie vertrieben werden, sondern weil sie dort die attraktiveren Arbeitsbedingungen finden, zumal sich einige wissenschaftliche Disziplinen in Österreich bis heute nicht von dem Exodus erholt haben. Dies hat zu der paradoxen Situation geführt, daß österreichische Forscher im Bereich der Formalwissenschaften zwar zu den in der Welt Führenden gehören, daß aber die Mehrzahl unter ihnen als Professoren oder Assistenten im Ausland arbeitet, mit wenigen oder gar keinen Kontakten zu ihrem Wissenschaftsbereich in Österreich.

Das Beispiel von Ungarn zeigt, daß dies nicht so sein muß. Auch Ungarn erlitt den Exodus einer großen Zahl von SpitzenforscherInnen und KünstlerInnen in der Zwischenkriegszeit und während der kommunistischen Okkupation mit den Spätfolgen eines bis in die Gegenwart anhaltenden "brain-drain" beim Nachwuchs. Aber im Unterschied zu Österreich hält eine große Anzahl ungarischer Gelehrter im Ausland enge wissenschaftliche Kontakte zu ihren Kollegen im Heimatland aufrecht, vor allem mittels zahlreicher kürzerer oder längerer Arbeitsbesuche an ungarischen Universitäten und Forschungsinstituten. Weiters üben diese Auslandsungarn durch ihre Besuche einen äußerst fruchtbaren Einfluß auf die Ausbildung der Studenten in Ungarn aus. Sie sind Mitveranstalter einer nicht geringen Zahl von Workshops in Ungarn und können durch persönliche Beratung während ihres Aufenthaltes in Ungarn zusätzliches Know how, Anregungen und wissenschaftliche Kontakte vermitteln. Darüber hinaus hat diese enge institu-

tionelle Verbindung der ungarischen Spitzenforscher in ihrer Heimat dazu geführt, daß die internationale Fachwelt diese nicht nur als hervorragende Professoren in den USA betrachtet, sondern eben auch als ungarische Gelehrte. Dieser Umstand trägt sicherlich nicht unwesentlich zum fast mythischen Ruf bei, den Ungarn gegenwärtig im Bereich der Formalwissenschaften besitzt.

So stellt sich die Frage, ob es nicht auch möglich wäre, unter Einsatz von relativ geringen Mitteln eine engere Anbindung von österreichischen Spitzenforschern, die im Ausland arbeiten, an den Wissenschaftsbetrieb in ihrer Heimat zu erreichen. Als naheliegende Lösung schlagen wir die Gründung eines Wissenschaftszentrums vor, das ausschließlich die Aufgabe hat, eine enge wissenschaftliche Verbindung zwischen Forschung und Lehre in Österreich und Österreichern, die im Ausland eine Professur im Bereich der Formalwissenschaften innehaben, zu schaffen. Dieses Wissenschaftszentrum wäre zu verstehen als Ideenlabor, das auch in die Bereiche technologische Kultur und Kunst übergreifen würde, weil in einem Jahrhundert, das zutiefst von den Errungenschaften einer Technokultur geprägt ist, eine Trennung von sogenannten Geistes- und Naturwissenschaften und von Wissenschaft und Kunst nicht nur historisch obsolet, sondern auch schädlich wäre. Die wissenschaftliche Forschung ist heute ebenso wie die Kunst in vielen Bereichen durch den Einsatz von Computern geprägt. Die digitale Technologie und deren Folgewirkungen für neue Medien sowie für die Wissenschaften von der Gesellschaft, vom Menschen, vom Gehirn, von den physikalischen Mikro- und Makrosystemen etc., bilden daher eine gemeinsame Achse für die Forschungsrichtung des Zentrums. Aus dieser gemeinsamen Basis sind besonders fruchtbare Ansätze zu erwarten. Dieses Zentrum würde kürzere und längere Heimataufenthalte der Auslandsösterreicher, insbesondere aus den USA und Kanada, verbunden mit Arbeitsmöglichkeiten an diversen Instituten ihrer Fachrichtung, unterstützen und organisieren. Durch eine Reihe von begleitenden Maßnahmen würde sichergestellt sein, daß diese Gastaufenthalte neue Anregungen und fruchtbare Ergänzungen für Forschung und Lehre in Österreich ergeben. Im Prinzip soll es für diese Gastaufenthalte kein Gehalt geben, sondern nur die Reise- und Aufenthaltskosten sollten vom Institut finanziert werden. Dieses Institut sollte für Auslandsösterreicher der favorisierte Ort werden, wo sie ihr Forschungsfreisemester (Sabbatical) verbringen. Da ein Sabbatical normalerweise die Weiterzahlung des Gehalts bedeutet, ist der finanzielle Anreiz nicht so wichtig, sondern viel wichtiger sind die günstigen Arbeitsbedin-

gungen und das geistige Milieu. Durch eine Vielfalt größerer und kleinerer Workshops, die von den Gästen am Wissenschaftszentrum organisiert werden, kann eine engere Zusammenarbeit mit Forscherkollegen in Österreich erreicht werden, aber auch der direkte Kontakt zu fortgeschrittenen Studierenden in Österreich. Diese Veranstaltungen wären eine große Bereicherung der Ausbildung und würden nicht nur helfen, die Kluft zwischen dem Ausbildungsniveau in Österreich und an den besten Universitäten im Ausland auszugleichen, sondern würden auch die beruflichen Chancen dieser Studierenden erhöhen. Auf diese Weise wird sicher der eine oder andere Auslandsösterreicher wieder Geschmack an der wissenschaftlichen Arbeit in Österreich finden und sich bei einer frei werdenden Professur bewerben. Durch das Einrichten von Internet-facilities am Wissenschaftszentrum wird ein globaler Informationsfluß möglich, insbesondere zwischen den Inlands- und Auslandsösterreichern. Darüber hinaus bilden diese Auslandsösterreicher am Wissenschaftszentrum einen zusätzlichen Pool von hervorragend qualifizierten und weitgehend unparteiischen Ratgebern und Gutachtern. Daher kann das Wissenschaftszentrum in vielfältiger Weise dazu beitragen, daß erstens Forschung und Lehre in Österreich wieder dasselbe Niveau erreichen, das sie einmal vor der "Vertreibung der Vernunft" besaßen, und zweitens, daß der Forschungsbetrieb in Österreich sich innerhalb der Europäischen Union optimal profilieren kann.

Appendix

Im Ausland tätige Universitätsprofessoren aus Österreich in den Gebieten Mathematik und Theoretische Informatik:

Prof. Dr. Martin Aigner
Institut für Mathematik
FU Berlin

Dr. Anton Arnold
Fachbereich Mathematik
Universität des Saarlandes
Saarbrücken

Prof. Dr. Leo Bachmair
Department of Computer Science
State University of New York at Stony Brook

Prof. Dr. Karl Doppel

Institut für Mathematik
FU Berlin

Prof. Dr. Herbert Edelsbrunner
Department of Computer Science
Duke University, Durham

Prof. Dr. Rainer Göttfert
Department of Mathematical Sciences
University of Nevada, Las Vegas

Ph.D. Prof. Helmut Grömer
Department of Mathematics
University of Arizona

Prof. Alexander Hahn
Department of Mathematics
University of Notre Dame

Prof. Ernst Hairer
Section de Mathématiques
Université de Genève

Prof. Dr. Susanne Hambruch
Department of Computer Science
Purdue University

Prof. Dr. Wilfried Hazod
Institut für Mathematik
Universität Dortmund

Prof. Dr. Thomas Henzinger
The Center for Applied Mathematics
Cornell University

Prof. Dr. Gerhard Jank
Lehrstuhl II für Mathematik
RWTH Aachen

Prof. Dr. Erich Kaltofen
Department of Mathematics
North Carolina State University

Prof. Dr. Hans Lausch
Department of Mathematics
Monash University

Dr. Peter Lesky, Akad. Rat
Universität Stuttgart

Prof. Dr. Rudolf Lidl
University of Tasmania
Dept. of Mathematics

Prof. Dr. Peter Markowich
TU Berlin
Fachbereich Mathematik

Prof. Dr. Johann Pfanzagl
Mathematisches Institut
Universität Köln

Prof. Dr. Walter Philipp
Department of Statistics
University of Illinois at Urbana-Champaign Champaign

Prof. Dr. Helmut Ratschek
Mathematisches Institut
Universität Düsseldorf

Prof. Dr. Christian Ringhofer
Department of Mathematics
Arizona State University

Prof. Dr. Gert Sabidussi
Département de Mathématiques et de Statistique
Université de Montréal

Prof. Dr. Hans Sagan
Department of Mathematics
Univ. of North Carolina

Prof. Dr. Norbert Sauer
Department of Mathematics
University of Calgary

Prof. Dr. Wolfgang Schmidt
Department of Mathematics
University of Colorado, Boulder

Prof. Dr. Raimund Seidel
FB Informatik
Universität des Saarlandes

Prof. Dr. Siegfried Steiner
Mathematisches Institut
Universität Stuttgart

Prof. Dr. Gottfried Tinhofer
Mathematisches Institut
TU München

Prof. Dr. Manfred Trummer
Department of Mathematics
Simon Fraser University

Univ.Doiz. Dr. Gerhard Turnwald
Wilhelm-Schickard Institut für Informatik
Tübingen

Prof. Dr. Anton Wakolbinger
Johann Wolfgang Goethe Universität Frankfurt/Main
FB Mathematik

Prof. Dr. Gerhard Wanner
Section de Mathematiques
Universite de Geneve

Prof. Dr. Wolfgang Watzlawek
Fakultät für Mathematik und Informatik
Universität Konstanz

Prof. Dr. Emo Welzl
Institut für Theoretische Informatik
ETH Zentrum

Prof. Dr. Harald Wimmer
Mathematisches Institut
Universität Würzburg

Internet Links für weitere Informationen

Nähere Informationen aus dem Kunstbereich findet man online auf den folgenden Homepages:

Neue Galerie Graz: <http://www.neuegalerie.at/titel.html>

Zentrum für Kunst und Medientechnologie: <http://www.zkm.de/>

Die Homepage der Technischen Universität Graz befindet sich unter <http://www.tu-graz.ac.at>,

wo unter

<http://www.tu-graz.ac.at/studium>

allgemeine Informationen zum Studium zusammengefaßt sind.

Das Mathematikstudium wird im wesentlichen von folgenden Instituten betreut, die auch jeweils eine Homepage anbieten:

Institut für Mathematik <http://www.math.tu-graz.ac.at>

Institut für Statistik <http://www.cis.tu-graz.ac.at/stat>

Institut für Geometrie <http://www.cis.tu-graz.ac.at/ig>

Die Homepage der Studienrichtungsvertretung Technische Mathematik befindet sich auf <http://oeh.tu-graz.ac.at/mathematik/>. Im Aufbau befindet sich eine Newsgroup tu-graz.mathematik, die am Newsserver news.tu-graz.ac.at zu finden ist.

Interessante Informationen zur Zahlentheorie findet man unter

<http://www.cs.unb.ca/~alopez-o/math-faq/node21.html>.

Einen Startpunkt für Informationen über Mathematikinstitute im In- und Ausland gibt

<http://euclid.math.fsu.edu/Science/Servers.html>,

Historisches zur Mathematik findet man unter

<http://www-groups.dcs.st-and.ac.uk:80/history/index.html>.

Für weitere Informationen wenden Sie sich bitte an Prof. Dr. Rainer E. Burkard (burkard@opt.math.tu-graz.ac.at) oder Dipl.-Ing. Clemens Heuberger (heuberger@weyl.math.tu-graz.ac.at).

Informationen über das Fach Theoretische Informatik findet man unter
<http://www.nada.kth.se/theory/links.html>
<http://dimacs.rutgers.edu/OtherLinks/tcs-res.html>.

Das Electronic Colloquium on Computational Complexity (ECCC) offeriert online neue research reports aus der Theoretischen Informatik, insbesondere über Komplexität von Berechnungen auf <http://www.eccc.uni-trier.de/eccc/>.

Ein anderes Archiv mit neuen Arbeiten über die Theorie des Maschinellen Lernens und Neuronaler Netze ist die Homepage des EU-Projekts NeuroCOLT: <http://www.neurocolt.com/>.

An der Technischen Universität Graz kann Theoretische Informatik im Rahmen eines Studiums der Technischen Mathematik oder der Telematik studiert werden. Verantwortlich für das Fach Theoretische Informatik ist an der Technischen Universität Graz das Institut für Grundlagen der Informationsverarbeitung <http://www.cis.tu-graz.ac.at/igi> .

Die Homepage von Prof. Maass ist
<http://www.cis.tu-graz.ac.at/igi/maass/>.

Die Homepage der Studienrichtungsvertretung Telematik befindet sich auf <http://oeh.tu-graz.ac.at/telematik/>.

Kurzbiographien der Autoren

Rainer E. Burkard

Rainer E. Burkard ist Professor für Mathematik an der Technischen Universität Graz. 1967 promovierte er an der Universität Wien in Mathematik. Von 1973-1981 war er ordentlicher Professor für Angewandte Mathematik an der Universität Köln. Seine Forschungsinteressen liegen auf dem Gebiet der Diskreten Optimierung. Er ist Autor des ersten deutschsprachigen Buches zur Ganzzahligen Optimierung, Mitautor und Herausgeber von 6 weiteren Büchern zur diskreten Optimierung und Autor von 120 wissenschaftlichen Arbeiten über strukturelle und algorithmische Aspekte in der Diskreten Optimierung. Er erhielt den Förderpreis der Österr. Mathematischen Gesellschaft, 1971, den Wissenschaftlichen Preis der Gesellschaft für Mathematik, Ökonomie und Operations Research 1991 und die EURO Gold Medaille 1997. 1998 wurde er zum Ehrenmitglied der Ungarischen Akademie der Wissenschaften gewählt.

Homepage: <http://www.opt.math.tu-graz.ac.at/burkard>

Herbert Edelsbrunner

Herbert Edelsbrunner promovierte 1982 an der Technischen Universität in Graz, Österreich, und arbeitete bis Sommer 1999 an der University of Illinois in Urbana-Champaign, Illinois, im Institut für Computer Wissenschaften. 1991 erhielt er den Alan T. Waterman Preis von der National Science Foundation in den U.S.A. Er ist Gründer und Direktor von Raindrop Geomagic, einer Firma für drei-dimensional geometrische Modellierungs Software. Seit August 1999 ist er Arts and Sciences Professor für Computer Wissenschaften an der Duke University in North Carolina. Er forscht in den Gebieten Datenstrukturen, Algorithmen, Geometrie, und Topologie. Er hat über hundert wissenschaftliche Arbeiten veröffentlicht und ist Autor eines Buches in algorithmischer Geometrie, erschienen 1987 bei Springer-Verlag.

Homepage: <http://www.cs.duke.edu/~edels>

Wolfgang Maass

Wolfgang Maass studierte an der Ludwig-Maximilians-Universität in München, an der er 1973 das Diplom, 1974 den Doktor und 1978 die Habilitation in Mathematik erhielt. Ein Heisenberg Stipendium der Deutschen Forschungsgesellschaft ermöglichte es ihm, in den 5 Jahren von 1979 bis 1984 Forschungsarbeiten am MIT in Cambridge, an der University of Chicago und an der University of California in Berkeley durchzuführen. Während dieses Stipendiums erhielt er 1982 einen Ruf als Associate Professor an die University of Illinois in Chicago und wurde dort 1986 zum ordentlichen Professor ernannt. Seit 1991 ist Wolfgang Maass ordentlicher Universitätsprofessor an der Technischen Universität Graz und seit 1992 Vorstand des neuen Instituts für Grundlagen der Informationsverarbeitung. Im Jahr 1997/98 wurde er als Sloan Fellow an das Computational Neurobiology Lab des Salk Instituts (USA) eingeladen.

Wolfgang Maass hat 115 Forschungsarbeiten auf den Gebieten Mathematische Logik, Komplexität von Berechnungen, Maschinelles Lernen, Neuronale Netze und Informationsverarbeitung in Nervensystemen veröffentlicht. Er ist Editor der Zeitschriften Machine Learning und Archive for Mathematical Logic, Associate Editor des Journal of Computer and System Sciences und Mitglied des Editorial Board der Zeitschrift Neurocomputing und des Handbook of Brain Theory and Neural Networks. Gemeinsam mit Chris Bishop hat er das Buch "Pulsed Neural Networks" (MIT Press) herausgegeben.

Homepage: <http://www.tu-graz.ac.at/igi/maass/>

Raimund Seidel

In Graz 1957 geboren; Schulausbildung am Akademischen Gymnasium in Graz und an der Hudson High School in Hudson, Bundesstaat Wisconsin, USA; Studium der Mathematik und Informatik an der Technischen Universität Graz, an der University of British Columbia in Vancouver, Kanada, und an der Cornell University in Ithaca, Bundesstaat New York, USA; Gastforscher beim DEC Systems Research Center und beim IBM Almaden Research Center; 1986–1996 in der Computer Science Division der University of California, Berkeley, als Assistant-, Associate- und schließlich Full-Professor;

seit 1994 Professor für Theoretische Informatik an der Universität des Saarlandes. Hauptinteressensgebiete: Algorithmentheorie, algorithmische und kombinatorische Geometrie, Randomisierung.

Homepage: <http://www-tcs.cs.uni-sb.de/seidel/>

Wolfgang M. Schmidt

Ph.D., University Vienna, 1995

University of Colorado since 1964, Full-Professor since 1965

Department of Mathematics, University of Colorado, Boulder, CO 80309-0395

Some Publications:

On normal numbers. *Pacific J. Math.* 10 (1960), 661–772.

Simultaneous approximation to algebraic numbers by rationals. *Acta Math.* 125 (1970), 189–201.

Irregularities of Distribution VII. *Acta Arith.* 21 (1972), 45–50.

Diophantine inequalities for forms of odd degree. *Advances in Math.* 38 (1980), 128–151.

Heights of points on subvarieties of G_m^n . *London Math. Soc. Lecture Notes Series* 235 (1996), 157–187.

On continued fractions and diophantine approximation in power series fields, *Acta Arith.* (to appear).

The zero multiplicity of linear recurrence sequences, *Acta Math.* (to appear).

Some Honors:

American Mathematical Society, Cole Prize in Number Theory, 1972

Dr. hon. causa, University of Ulm, Germany, 1981

Corresponding member, Austrian Academy of Science, 1982

Distinguished Professor, University of Colorado, 1988

Fellow, National Academy of Arts and Sciences, 1944

Dr. hon. causa, University of Paris VI, 1994

Member, Polish Academy of Science, 1997. Dr. hon. causa, University of Marburg, Germany, 1999.

Homepage: <http://www.colorado.edu/math/children/faculty/schmidt/>

Gottfried Tinhofer

geboren 1.10.1938, Kundl, Tirol, nach Volks- und Hauptschule Ausbildung als Volksschullehrer an der Lehrerbildungsanstalt Innsbruck, Matura 1959. Ein Jahr Studium der Germanistik und Romanistik, Ab 1961 Studium der Mathematik und theoretischen Physik in Innsbruck, 1966 Promotion in Mathematik mit einem Thema aus Differentialgeometrie, anschließend zunächst einige Monate Wehrdienst beim Österreichischen Bundesheer, ab 1967 Assistent am damaligen Institut für Rechentechnik der Universität Innsbruck, erste Kontakte mit Graphentheorie durch Mitarbeit an Projekten des mathematischen Instituts der Technischen Universität Graz, 1973 Habilitation an der Universität Innsbruck mit einem Thema aus Analysis, 1974 Berufung an die Technische Universität München, seit dieser Zeit intensive Beschäftigung mit Graphentheorie, die ihren Niederschlag in zahlreichen Zeitschriftenartikeln, Büchern und Tagungsorganisationen findet.

Homepage: <http://www-geo.mathematik.tu-muenchen.de/m9/algograph/homepages/tinhofer/>

Peter Weibel

geboren 5.3.1944 in Odessa, lebt in Karlsruhe, Studien der Literatur, Medizin, Logik, Philosophie und des Films in Paris und Wien. Dissertation über mathematische Logik (Modallogik). Polyartist, Kunst- und Medientheoretiker, Ausstellungskurator.

1976 - 1981 Lektor für "Theorie der Form" und

1981 - 1984 Gastprofessor für Gestaltungslehre und Bildnerische Erziehung an der Hochschule für angewandte Kunst in Wien

1981 Gastprofessor am College of Art and Design, Halifax, Canada

1979/80 Gastprofessor für "Medienkunst" und

1981 Lektor für "Wahrnehmungstheorie" und

1982 - 1985 Professor für Fotografie an der Gesamthochschule Kassel Gastprofessor für "Medienkunst" und

1981 Lektor für "Wahrnehmungstheorie" und

1982 - 1985 Professor für Fotografie an der Gesamthochschule Kassel

Associate Professor for Video and Digital Arts, Center for Media Study, State University of New York at Buffalo

1989 - 1994 Direktor des Instituts für Neue Medien an der Städelschule in Frankfurt/M.

Seit 1984 Professor für visuelle Mediengestaltung an der Hochschule für angewandte Kunst in Wien
1986 - 1995 Künstlerischer Berater der Ars Electronica in Linz.
1992 - 1995 Künstlerischer Leiter der Ars Electronica in Linz.
Seit 1993 Kurator der Neuen Galerie am Landesmuseum Joanneum in Graz.
Seit 1993 Österreich-Kommissär der Biennale von Venedig.
Seit 1999 Direktor des ZKM in Karlsruhe.

Wolfgang Woess

geboren 1954

1973-1978 Studium der Technischen Mathematik (TU Wien)

1978-1980 Doktoratsstudium (Univ. München und Salzburg)

1980-1981 Forschungsstipendium an der Univ. Salzburg

1982-1988 Universitätsassistent an der Montanuniv. Leoben

1984-1985 Forschungsaufenthalt in Rom

1986 Habilitation für Mathematik an der Univ. Salzburg

1988-1994 Professore associato für Math. Analysis an der Univ. Mailand

1994-1999 Professore ordinario für Wahrscheinlichkeitstheorie an der Univ. Mailand

Seit 1.9.1999 Univ.Prof. für Mathematik an der TU Graz

Arbeitsgebiet: Irrfahrten (Random Walks) auf Graphen und Gruppen, Diskrete Potentialtheorie, Sturkturtheorie unendlicher Graphen ca 60 Publikationen, 2 Bücher

Homepage: <http://www.cis.tu-graz.ac.at/mathc/woess/>