Diese Arbeit wurde vorgelegt am Lehrstuhl für Mathematik - MATHCCES

Didaktisch-methodische Ausarbeitung eines Lernmoduls zum Thema Google im Rahmen eines mathematischen Modellierungstages für Schülerinnen und Schüler der Sekundarstufe II

Bachelorarbeit in Mathematik

vorgelegt von: Sarah Schönbrodt
[redacted]

Erstprüfer: Prof. Dr. Martin Frank
Department of Mathematics
RWTH Aachen

Zweitprüfer: Prof. Dr. Johanna Heitzer
Lehr- und Forschungsgebiet Didaktik der Mathematik
RWTH Aachen

Aachen, den 13. August 2015

Inhaltsverzeichnis

1. Motivation

Bereits seit Jahrzehnten spielen Anwendungen und Modellierung eine wichtige Rolle in den nationalen und internationalen didaktischen Diskussionen zum Mathematikunterricht. Jährlich kommen mehr und mehr fachdidaktische Publikationen, die das große Interesse an dem Forschungsgebiet der *Modellierung im Mathematikunterricht* unterstreichen, hinzu und auch die Fülle an Unterrichtsmaterialien zu mathematischer Modellierung wächst stetig[1] (vgl. [5], Blum, 2007, S. 5). Doch obgleich die Bedeutung der Einbeziehung von Anwendungen in den Mathematikunterricht vom fachdidaktischen Standpunkt vielfach betont wurde (vgl. [5], Blum, 2007, S. 3; [19], Westermann 2011, S. 148), ist die Einbindung von mehr Realitätsbezügen in die konkrete Unterrichtspraxis noch nicht im gewünschten Maße erfolgt. Insbesondere die Ergebnisse von PISA und TIMSS verdeutlichten, dass die Schülerinnen und Schüler[2] weltweit, insbesondere auch in Deutschland, große Probleme beim Bearbeiten von Aufgaben haben, welche die Anwendung von Mathematik zur Lösung eines außermathematischen Problems verlangen (vgl. [19], Westermann, 2011, S. 148). Dies wirft die Frage nach den Gründen auf, warum zwischen den didaktischen Diskussionen zur Modellierung und der Unterrichtsrealität eine derartige Abweichung besteht. Blum (2007) stellte diesbezüglich fest, dass Modellieren nicht nur für Schüler eine Herausforderung darstellt, sondern auch Lehrkräften schwer fallen kann. So kann die Einbindung außermathematischer Kontexte in den Mathematikunterricht diesen sowohl komplexer, als auch weniger gut planbar bzw. vorhersehbar machen. Vielfach kommt es zudem zu einem erhöhten Vorbereitungsaufwand der betreffenden Unterrichtsstunden, was sich unter anderem darauf zurückführen lässt, dass außermathematisches Sachwissen notwendig ist (vgl. [5], Blum, 2007, S. 5; [19], Westermann, 2011, S. 151).

Da das mathematische Modellieren sowohl Schülern als auch Lehrern offenbar relativ *schwer* fällt, sollte man an dieser Stelle zunächst einen Blick darauf werfen, warum Anwendungen und mathematische Modellierung dennoch bedeutsam sind. Welchen Nutzen bzw. Mehrwert bringen sie mit sich und legitimieren damit die aktuellen Bemühungen um mehr Anwendungsorientierung im Unterricht?
Die Bedeutung von Anwendungen und mathematischer Modellierung wird nach Westermann vor allem durch Realitätsbezüge, die einen Beitrag zur *Verbesserung des Bildes von Mathematik* leisten können, hervorgehoben. Darüber hinaus können sie als *Hilfe beim Mathematiklernen* dienen, indem sie zu einem tieferen Verständnis sowie zu einem nachhaltigeren Begreifen und Behalten mathematischer Begriffe und Verfahren und zudem zu einer höheren Motivation seitens der Schüler führen. Die Betonung des Anwendungsaspekts im Mathematikunterricht kann außerdem dazu dienen die spezifische Art, in der Mathematik die Welt beschreibt und beeinflusst, zu verdeutlichen und die

[1]Beispiele für eine solche Materialsammlung stellen die ISTRON-Reihe oder die MUED-Sammlung dar.

[2]Nachfolgend werden Schülerinnen und Schüler unter der Bezeichnung *Schüler* zusammengefasst. Analog wird mit den Personengruppen Lehrerinnen und Lehrer, Betreuerinnen und Betreuer u.a. verfahren.

Rolle, die Mathematik in der Welt spielt, hervorzuheben, und damit einen *Beitrag zu der Allgemeinbildung und zum Weltverständnis* der Schüler zu leisten (vgl. [19], Westermann, 2011, S. 149). Schließlich können Anwendungen auch als *Mittel zur Entwicklung von prozessbezogenen Kompetenzen* und zur*Ausbildung von Entscheidungs-* sowie *Handlungsstrategien* gesehen werden (vgl. [5], Blum, 2007, S. 4).
Sicherlich auch aus den genannten Gründen findet sich das Modellieren als eine der zentralen Kompetenzen in den verbindlichen Bildungsstandards wieder (vgl. [13], Kultusministerkonferenz, 2003, S. 7). Auch in den Kernlehrplänen des Landes Nordrhein-Westfalen (Sek I / Sek II) wird das Modellieren als eine der prozessbezogenen Kompetenzen benannt, welche die Schüler durch die Auseinandersetzung mit konkreten Fragestellungen und Lerninhalten des Fachs Mathematik erwerben sollen (vgl. [16], Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, 2007, S. 12).

Das Modellierungsprogramm CAMMP[3] der RWTH Aachen[4] hat es sich zur Aufgabe gemacht im Rahmen von einzelnen Modellierungstagen oder Modellierungswochen die Entwicklung und Förderung der Modellierungskompetenz von Schülern zu unterstützen. Indem sich die Schüler mit aktuellen Fragestellungen beschäftigen und zur Problemlösung mathematische Methoden und Computersimulationen verwenden, erlernen sie die Grundlagen der mathematischen Modellierung durch aktive Auseinandersetzung mit alltagsnahen Problemstellungen.
CAMMP bietet zudem interessierten Lehrern die Möglichkeit durch begleitende Teilnahme[5] an den von CAMMP angebotenen Veranstaltungen, neben Anregungen zur Einbindung außermathematischer Kontexte in ihren eigenen Unterricht, auch konkrete Materialien zu den Lernmodulen von CAMMP zu erhalten.

Das im Zuge dieser Arbeit ausgearbeitete Lernmodul zum Thema Google, welches zur Durchführung im Rahmen eines Modellierungstages von CAMMP für Schüler der Sekundarstufe II entwickelt wurde, soll seinerseits einen Beitrag zu Ausbildung der so wichtigen Modellierungskompetenz leisten.

In dieser Arbeit werden in Kapitel 2.1 zunächst das Modellieren, die Modellierungskompetenz, verschiedene Modellierungskreisläufe sowie die Ziele von Modellierung im Bezug auf den Mathematikunterricht thematisiert. Damit soll eine wichtige Grundlage für die Erklärung und Diskussion der entwickelten Materialien vom didaktisch-methodischen Standpunkt aus geschaffen werden. Weiterhin wird in Kapitel 2.2 die Philosophie von CAMMP, die unter anderem als Orientierungsbasis bei der Ausarbeitung des Lernmoduls diente, dargelegt. Zudem werden die von CAMMP angebotenen Modellierungstage vorgestellt, die den organisatorischen Rahmen des Lernmoduls feststeckten.

In Kapitel 3 werden die dem Thema des Lernmoduls zugrundeliegenden mathematischen

[3]**C**omputitional **A**nd **M**athematical **M**odeling **P**rogram (vgl. Abschnitt 2.2)
[4]www.rwth-aachen.de, Stand: 02.08.2015
[5]Die Schülergruppen, die die Angebote von CAMMP wahrnehmen, werden von Mathematiklehrern begleitet.

Inhalte erläutert, um anschließend zu dem didaktisch-methodischen Konzept, welches den Kern dieser Arbeit ausmacht, über zu leiten. Im Zuge dessen werden die entwickelten Materialien vorgestellt und vor dem theoretischen Hintergrund der mathematischen Modellierung diskutiert.

Schließlich wird in Kapitel 5 die Durchführung des Lernmoduls im Rahmen eines Modellierungstages beschrieben sowie die von mir gemachten Beobachtung dargelegt und die durchgeführten Schülerbefragungen ausgewertet. Abschließend wird bezugnehmend auf die Ziele von CAMMP sowie auf die allgemeinen Ziele der mathematischen Modellierung eine zusammenfassende Wertung des Lernmoduls vorgenommen und ein Ausblick auf Optimierungs- sowie Weiterentwicklungspotentiale des Lernmoduls und der entwickelten Materialien gegeben.

2. Theoretischer Hintergrund

2.1. Mathematische Modellierung

Wie der Name „Computergestütztes Mathematisches Modellierungsprogramm“ bereits aussagt, bildet das mathematische Modellieren das Fundament der von CAMMP angebotenen Projekte und somit auch des im Zuge dieser Arbeit konzipierten Lernmoduls. Aufgrund dessen sollen nachfolgend zunächst wichtige Begriffe geklärt, sowie die damit verbundenen theoretischen und didaktischen Hintergründe des mathematischen Modellierens erörtert werden.

2.1.1. Modelle und mathematisches Modellieren

Die mathematische Modellierung, die als Teilaspekt der angewandten Mathematik aufgefasst werden kann, „legt den Fokus auf den *Prozess* des Lösens von Problemen aus der Realität oder [...] aus dem Rest der Welt außerhalb der Mathematik“ (vgl. [8], Greefrath u.a., 2013, S. 11).
Ein wichtiger Bestandteil des mathematischen Modellierens ist die Entwicklung mathematischer Modelle. Ein Modell bezeichnet eine vereinfachte Darstellung der Realität, welche nur bestimmte, jedoch hinreichende, Teilaspekte berücksichtigt. Auf ein mathematisches Modell lassen sich mathematische Methoden anwenden, um schließlich mathematische Resultate zu erhalten. Die Beschreibung realer Situationen durch mathematische Modelle hat insofern ihre Grenzen, als dass die Realität auf Grund ihres Detailreichtums und ihrer Komplexität nicht vollständig erfasst werden kann. Dies ist jedoch meist auch nicht erwünscht, da Modelle gerade der übersichtlicheren Darstellung und schnelleren Verarbeitung realer Daten und Informationen dienen (vgl. [8], Greefrath u.a., 2013, S. 13).

2.1.2. Modellieren als Kompetenz

Modellierung und Anwendungsorientierung im Mathematikunterricht hat in allgemeinbildenden Schulen in den letzten Jahren zunehmend an Bedeutung gewonnen. Insbesondere die Einführung der national geltenden Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss zu Beginn des 21. Jahrhunderts hat einen wichtigen Anteil dazu beigetragen. In den Bildungsstandards ist *mathematisches Modellieren* eine der sechs allgemeinen mathematischen Kompetenzen, welche die Schüler im Laufe ihrer Schullaufbahn durch die aktive Auseinandersetzung mit mathematischen Inhalten erwerben sollen (vgl. [13], Kultusministerkonferenz, 2003, S. 6). Auch in den an den Bildungsstandards orientierten Kernlehrplänen des Landes Nordrhein-Westfalen ist das mathematische Modellieren in der Sekundarstufe I und II als eine von den Schülern zu erwerbende prozessbezogene Kompetenz fest verankert (vgl. [16] Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, 2007; [17], Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, 2014).

Die Modellierungskompetenz kann beschrieben werden als „die Fähigkeit, die jeweils

nötigen Prozessschritte beim Hin- und Herwechseln zwischen Realität und Mathematik problemadäquat auszuführen sowie gegebene Modelle zu analysieren oder vergleichend zu beurteilen" (Blum, 2007, zitiert nach [8], Greefrath u.a., 2013, S. 18). Schüler sollen demnach die Fähigkeit erwerben, den Wechsel zwischen Realität und Mathematik zu vollführen.
Die Modellierungskompetenz lässt sich in die Teilkompetenzen *Verstehen*, *Vereinfachen*, *Mathematisieren*, *Interpretieren*, *Validieren* und *Vermitteln* aufschlüsseln. Auf die Kennzeichen dieser Teilkompetenzen wird im nachfolgenden Abschnitt, bei der Erläuterung des Modellierungskreislaufes, genauer eingegangen (vgl. [8], Greefrath u.a., 2013, S. 19).

2.1.3. Der Modellierungskreislauf

Zur Beschreibung des gesamten mathematischen Modellierungsprozesses existieren zahlreiche idealisierte Modellierungskreisläufe.
Der in Abbildung 1 dargestellte siebenschrittige Modellierungskreislauf wurde vor dem Hintergrund der kognitiven Betrachtung des Modellierungsprozesses im Rahmen des DISUM-Projekts von Blum und Leiss (2005) entwickelt. Dieser Kreislauf ist insbesondere diagnostisch hilfreich, da er detailliert beschreibt welche Schritte der Lernende in idealtypischer Weise beim Lösen von Modellierungsaufgaben durchläuft (vgl. [5], Blum, 2007, S. 5).

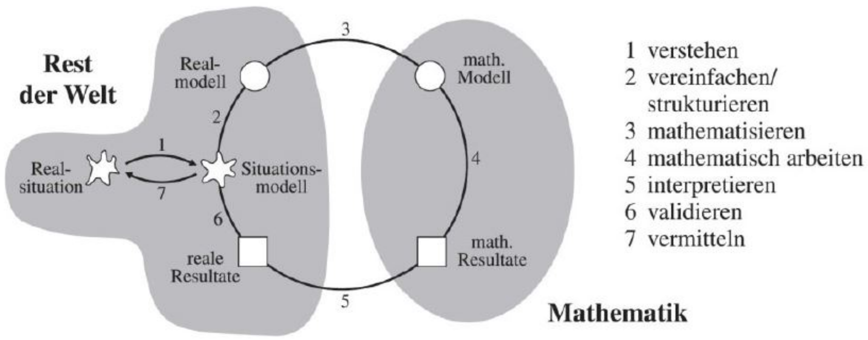


Abbildung 1: Modellierungskreislauf nach Blum und Leiss (vgl. [4], Blum, 2006, S. 9)

Der Modellierungskreislauf geht von der *Realsituation*, d. h. einer Problemstellung oder einer Aufgabe in der realen Welt, aus. Im 1. Schritt müssen die Schüler diese reale Situation zunächst *verstehen*, um so ein mentales Modell der Ausgangssituation, das sogenannte *Situationsmodell*, zu konstruieren. Dieses Modell enthält vielfach noch Angaben die irrelevant oder überflüssig sind. Im 2. Schritt muss der Lernende die gegebenen, als notwendig identifizierten Daten und Informationen *strukturieren* und *vereinfachende* oder *idealisierende Annahmen* treffen. Diese Annahmen helfen die Komplexität der Situation zu verringern, sodass das Problem schließlich in vereinfachter Form als sogenanntes *Realmodell* dargestellt werden kann. Dabei muss darauf geachtet werden, dass das Problem durch die getroffenen Annahmen nicht *zu einfach* und dadurch falsch abgebildet wird. Der 3. Schritt spannt nun die Brücke zwischen Realität und Mathematik, indem

der Lernende das Realmodell durch *Mathematisierung* in ein *mathematisches Modell* übersetzt, auf dem schließlich Mathematik anwendbar ist (vgl. [19], Westermann, 2011, S. 157). Im 4. Schritt, dem *mathematischen Arbeiten* auf dem mathematischen Modell, muss der Lernende geeignete mathematische Methoden anwenden, um schließlich zu einem *mathematischen Resultat* zu gelangen (vgl. [12], IQB, 2009, S. 4). In diesem Schritt spielt der Computer als Modellierungswerkzeug häufig eine zentrale Rolle. Aufgrund der hohen Komplexität der realen Probleme und der Quantität der zu verarbeitenden Daten ist dessen Einsatz vielfach sogar unverzichtbar. Indem der Computereinsatz den Lernenden vom Kalkül entlastet, kann zudem die hinter den Berechnungen liegende Bedeutung in den Vordergrund rücken (vgl. [15], Leuders, 2011, S. 207). Das ermittelte Resultat muss im 5. Schritt wieder in die Realität übertragen und *interpretiert* werden. Das auf diese Weise erhaltene *reale Resultat* wird schließlich anhand des Situationsmodells *validiert* (6. Schritt). Ist das Modell gut, d. h. liefert es brauchbare und sinnvolle Ergebnisse, so können diese im 7. und letzten Schritt verständlich *vermittelt* bzw. *kommuniziert* werden (vgl. [12], IQB, 2009, S. 4).
Wird nach einmaligem Durchlaufen des Kreislaufes jedoch noch kein zufriedenstellendes Ergebnis erzielt, wie es vielfach der Fall ist, müssen die Schritte 2 bis 6 des Modellierungskreislaufes noch weitere Male durchlaufen werden. Vereinfachende Annahmen werden dann auf ihre Gültigkeit und gegebene Daten auf ihre Richtigkeit überprüft. Schließlich werden weitere Annahmen getroffen oder Informationen hinzugenommen, weggelassen bzw. abgeändert und der Modellierungskreislauf erneut durchschritten.

Der oben beschriebene siebenschrittige Modellierungskreislauf bietet dem Lehrenden eine gute Möglichkeit die einzelnen Schritte, die der Lernende bei der Bearbeitung von Modellierungsaufgaben durchläuft, zu reflektieren. Aufgrund der hohen Komplexität des Modellierungskreislaufes ist dieser jedoch nicht zur gemeinsamen Thematisierung des Modellbildungsprozesses mit den Schülern und somit nicht als Werkzeug zur Metakognition geeignet (vgl. [12], IQB, 2009, S. 4). Stattdessen bietet es sich an mit einem vereinfachten Modellierungskreislauf zu arbeiten, der sich durch eine reduzierte Komplexität auszeichnet und der damit als Orientierungshilfe bei der Bearbeitung von Modellierungsaufgaben herangezogen werden kann.

Ein möglicher vierschrittiger und damit vereinfachter Modellierungskreislauf, angelehnt an Blum (1985), ist in Abbildung 2 dargestellt (vgl. [8], Greefrath u.a., 2013, S. 16).
Ausgangspunkt ist auch bei diesem idealen Kreislauf die *Reale Situation*. Diese wird zunächst strukturiert und vereinfacht (1. Schritt), um schließlich ein *Reales Modell* zu erhalten. Durch Mathematisieren (2. Schritt) wird das Reale Modell in ein *Mathematisches Modell* überführt, auf dem mathematisch gearbeitet werden kann (3. Schritt) bis ein *Mathematisches Resultat* erhalten wird. Dieses muss dann in Bezug auf die Reale Situation interpretiert werden (4. Schritt). Stellt das Mathematische Resultat noch keine zufriedenstellende Lösung des realen Problems dar, so wird der Kreislauf erneut durchlaufen.

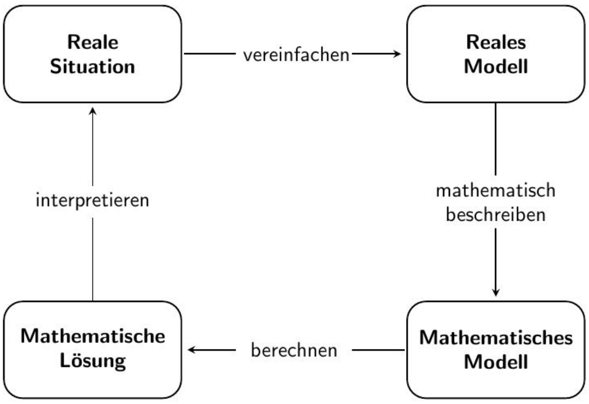


Abbildung 2: Vereinfachter Modellierungskreislauf (angelehnt an Blum, vgl. [8], Greefrath u.a., 2013, S. 17)

Dieser vereinfachte Kreislauf kann dem Lernenden als Hilfe zur bewussten Auseinandersetzung mit dem Prozess des mathematischen Modellierens und folglich mit der Vorgehensweise der Mathematik bei der Lösung realer Probleme dienen. Auf die Einbindung dieses vierschrittigen Modellierungskreislaufes in die Konzeption des entwickelten Lernmoduls wird in Abschnitt 4.3 ausführlich eingegangen.

2.1.4. Ziele des Modellierens

Mit der verstärkten Einbindung von Modellierung und Anwendungsorientierung in den Mathematikunterricht werden verschiedene Ziele verfolgt. Dazu zählen unter anderem inhaltsbezogene, prozessbezogene, lernpsychologische und allgemeine Ziele. Im Folgenden werden einige dieser Ziele, die insbesondere bei der Ausarbeitung des Lernmoduls als Orientierungsgrundlage dienten, aufgeführt. Inwieweit die im Rahmen des Lernmoduls entwickelten Materialien diesen Zielen tatsächlich genügen, wird in den Kapiteln 5 und 6 ausführlich diskutiert.

- Inhaltsbezogene Ziele:
  Durch die Beschäftigung mit Modellierungsaufgaben befassen sich die Schüler mit lebensweltlichen Problemen und sind in der Lage, sich ihre Umwelt mit mathematischen Mitteln zu erschließen. Modellierung und Anwendungsorientierung ermöglichen den Schülern somit ein besseres Verständnis zur Wahrnehmung unserer Welt zu entwickeln. Dies spricht insbesondere die Erste der von Winter (1995) formulierten Grunderfahrungen an, denen Mathematikunterricht genügen sollte (vgl. [19], Westermann, 2011, S. 55).

- Prozessbezogene Ziele:
  Durch die Beschäftigung der Schüler mit realen Problemen und anwendungsbezogenen Aufgaben sollen auch andere allgemeine Kompetenzen, wie die Problemlösefähigkeit, Kommunikation oder Argumentation, gefördert werden (vgl. [8], Greefrath u.a., 2013, S. 19 ff.).

- Lernpsychologische Ziele:
Anwendungsorientierung und Modellierung sollen zu einem tieferen Verständnis sowie zu einem besseren Behalten der mathematischen Inhalte führen und zudem das Interesse und die Motivation der Schüler an der Mathematik erhöhen (vgl. [8], Greefrath u.a., 2013, S. 19 ff.).

- Allgemeine Ziele:
Die Bearbeitung von Modellierungsaufgaben soll zu einer Verbesserung des Bildes von Mathematik führen, indem die Schüler Sinn, Nutzen und Bedeutung der Mathematik an einer alltagsnahen Problemstellung erfahren. Durch Modellierungsprobleme aus den verschiedensten Anwendungsbereichen soll der Satz „Mathematik ist überall" auch für Schüler mehr an Glaubwürdigkeit gewinnen (vgl. [19], Westermann, 2011, S. 149).

2.2. CAMMP

2.2.1. Allgemeines

CAMMP[6], **C**omputitional **A**nd **M**athematical **M**odeling **P**rogram, ist ein mathematisches Schülerlabor an der RWTH Aachen. Organisiert wird CAMMP vom Lehrstuhl Mathematik CCES[7] (Prof. Dr. Martin Frank, Dr. Christina Roeckerath), der Arbeitsgruppe Molecular Simulations und Transformations (Prof. Dr. Ahmed E. Ismail) und der Graduiertenschule AICES (Dr. Nicole Faber). CAMMP bietet zwei verschiedene Veranstaltungsformate an, in deren Rahmen sowohl Schüler als auch Lehrer die Möglichkeit erhalten, aktiv in die Problemlösung mit Hilfe von mathematischer Modellierung unter Computereinsatz einzusteigen. Diese beiden Veranstaltungformate, CAMMP week und CAMMP day, sollen im Folgenden kurz vorgestellt werden.

Die CAMMP week ist eine Modellierungswoche, in der sich an Mathematik interessierte Schüler der Sekundarstufe II eine Woche lang mit einer realen Problemstellung befassen. Diese Problemstellungen stammen aus der aktuellen Forschung von Firmen oder Universitätsinstituten. Beispiele solcher Problemstellungen, die im Rahmen von CAMMP weeks bereits zum Tragen kamen, sind die *Optimierung des Buchungssystems einer Carsharing-Firma* oder die *Messung der Geschwindigkeit mithilfe einer Handykamera*. Während der gesamten Modellierungswoche arbeiten die Schüler in kleinen Teams gemeinsam mit einem Wissenschaftler an der Lösung der Problemstellung und werden zusätzlich von zwei Lehrkräften betreut. Am Ende der Modellierungswoche findet eine Abschlussveranstaltung statt, bei der die Schüler ihre Ergebnisse unter anderem den Problemstellern präsentieren.

[6] *www.cammp.rwth-aachen.de*, Stand: 27.07.2015

[7] Das Center for Computational Engineering Science ist eine Kooperation verschiedener Institute der RWTH aus fünf Bereichen (u.a. Mathematik, Informatik / Naturwissenschaften und Maschinenbau). *www.cces.rwth-aachen.de*, Stand: 27.07.2015

Das zweite von CAMMP angebotene Veranstaltungsformat ist der CAMMP day, ein Modellierungstag, der erstmalig im Jahr 2012 angeboten wurde. Mathematikkurse der Mittel- bis Oberstufe kommen im Rahmen eines CAMMP days mit ihren Lehrern an die RWTH Aachen. Dort beschäftigen sie sich einen Tag lang mit einer praxisorientierten Fragestellung und werden dabei von Wissenschaftlern unterstützt. In kleinen Gruppen arbeiten die Schüler an der Lösung eines realen herausfordernden Problems, welches für die Durchführung im Rahmen der CAMMP days bereits didaktisch-methodisch ausgearbeitet wurde.

Bei der Lösung der Probleme, sowohl bei den CAMMP days, als auch bei den CAMMP weeks, kommen mathematische Methoden sowie Computersimulationen zum Einsatz. Damit bietet CAMMP den Schülern einen Einblick in den Berufsalltag von Mathematikern, Informatikern und Ingenieuren und insbesondere in die Welt der mathematischen Modellierung.

2.2.2. Philosophie und Ziele von CAMMP

Die von CAMMP verfolgten Ziele, die sich sowohl an Schüler, Lehrer als auch an die Veranstalter richten, werden nachfolgend aufgeführt:

CAMMP dient der Berufs- und Studienorientierung, indem es die gesellschaftliche Relevanz von Mathematik und Simulationstechnik an realen Problemstellungen heraushebt und die Schüler auf diese Weise für ein Studium der Mathematik oder einer mathematiknahen Studienrichtung, wie beispielsweise CES[8], sensibilisiert und begeistert.

CAMMP soll keinen aus der Schule ausgelagerter Unterricht darstellen, sondern kann vielmehr als Forschung auf dem Gebiet der Didaktik verstanden werden. Es ist damit nicht als starres Progamm aufzufassen, sondern kann und soll fortwährend weiterentwickelt werden und bietet damit gleichermaßen die Möglichkeit neue didaktische Methoden und Prinzipien zu erproben.

CAMMP soll allen Beteiligten Spaß machen. Es kann als Lehr-Lern-Labor aufgefasst werden, welches auch die Veranstaltenden als Angebot zur Weiterbildung nutzen können.

CAMMP dient der Aus- bzw. Fortbildung von Lehrern, die im Rahmen eines CAMMP days oder einer CAMMP week Ideen für die Integration von komplexen und anwendungsorientierten Problemstellungen in den Schulunterricht erlangen können und zudem für konkrete Projekte Materialien an die Hand bekommen.

[8] CES steht für Computational Engineering Science und ist ein an der RWTH angebotener Studiengang. *www.ces.rwth-aachen.de*, Stand: 27.07.2015

2.2.3. Organisatorischer Rahmen eines CAMMP days

Das Lernmodul zum Thema Google, welches von mir im Zuge dieser Arbeit entwickelt wurde, ist für die Durchführung im Rahmen eines CAMMP days ausgerichtet. Aufgrund dessen wird der organisatorische Rahmen eines CAMMP days nachfolgend beschrieben.

Abhängig von dem Thema des behandelten Moduls richten sich die CAMMP days an Schülergruppen aus der Mittel- oder Oberstufe[9]. Bei den Schülergruppen handelt es sich meist um einzelne Mathematikkurse[10] einer allgemeinbildenden Schule. Das Thema des Modellierungstages können die Lehrer, vorzugsweise in Absprache mit den Schülern, aus einem Modulangebot wählen. Dieses Angebot umfasst derzeit die folgenden Lernmodule:

- Wie funktioniert eigentlich GPS und was hat das mit Mathe zu tun?
- Spiegelaufstellung in einem Solarkraftwerk
- Vom Lotfällen bis zum JPEG-Format
- Wie funktioniert eigentlich Google und was hat das mit Mathe zu tun?
- Von Mickey Mouse bis Buzz Light Year – wie Mathematik die Filmfiguren zum Leben erweckt

Die CAMMP days finden am Center for Computational Engineering Science in der RWTH Aachen statt. Bei der Bearbeitung der Lernmodule, die vorzugsweise in kleinen Gruppen oder in Partnerarbeit erfolgt, stehen den Schülern Computer zur Verfügung. Zudem werden die Schüler von mehreren Wissenschaftlern unterstützt.

Jeder CAMMP day beginnt nach einer kurzen Begrüßung mit einem Vortrag, der die Schüler in die Welt der mathematischen Modellierung einführt und die Bedeutung von Modellierung und Simulation für Industrie, Wirtschaft und Forschung anhand von ausgewählten Beispielen veranschaulicht. Die weitere Gestaltung und auch die Dauer des CAMMP days ist abhängig von dem jeweiligen Lernmodul.
Zusätzlich zu der Bearbeitung des Lernmoduls wird den Schülern im Laufe des Modellierungstages durch den Besuch der AixCave[11] oder der Powerwall[12] ein Einblick in das Forschungsgebiet der virtuellen Realität ermöglicht.

[9] Die Module erfordern unterschiedliches Vorwissen, sodass nicht jedes Modul für jede Altersgruppe geeignet ist.

[10] Im Rahmen von *Freien CAMMP days* werden auch gemischte Schülergruppen unterschiedlicher Schulen an CAMMP days teilnehmen können.

[11] Eine fünfseitige Virtual-Reality-Installation zur Darstellung immersiver, virtueller Umgebungen.

[12] Virtual reality system.

3. Mathematischer Hintergrund

Google ist eine Suchmaschine, die das WWW innerhalb von Sekundenbruchteilen durchsucht und dem Internetnutzer schließlich die *wichtigsten* Seiten zu seiner Suchanfrage ganz oben auflistet. Wie Google die Wichtigkeit einer Internetseite festlegt und welche elementaren mathematischen Ideen hinter der Berechnung dieser Wichtigkeit stecken, soll nachfolgend erläutert werden.

3.1. Der Erfolg der Suchmaschine Google

Der Prototyp der Suchmaschine Google wurde Mitte der 90er Jahre von den beiden Informatikstudenten Lawrence Page und Sergej Brin an der Stanford University entwickelt. Nachdem Brin und Page im Jahr 1998 das Unternehmen Google Inc. gegründet hatten, ging die Suchmaschine das erste mal online. Mit einem Marktanteil von über 70%[13] ist Google mittlerweile die unangefochtene Nummer eins unter den Suchmaschinen und die Erfinder gehören laut Forbes-Liste[14] unterdessen zu den zwanzig reichsten Menschen weltweit (vgl. [10], Humenberger, 2009, S. 665). Google hob sich damals einerseits durch eine wesentlich schnellere Beantwortung der Suchanfragen gegenüber konkurrierenden Suchmaschinen und andererseits durch das Erstellen von Rankings, bei denen dem Nutzer die scheinbar interessantesten Internetseiten stets ganz oben präsentiert wurden, ab (vgl. [6], Bryan und Leise, 2006, S. 1).
Was Google so erfolgreich machte, ist der auf der Verlinkungsstruktur des Internets basierende PageRank-Algorithmus. Dieser berechnet monatlich die relative Wichtigkeit einer Internetseite. Diese Wichtigkeit wird durch den PageRank ausgedrückt, einer Zahl zwischen 0 und 1, nach der die Internetseiten bei Suchanfragen schließlich sortiert werden. Dabei gilt: Umso höher der PageRank, umso wichtiger die Seite.
Im nachfolgenden Abschnitt werden die grundlegenden Schritte der Suchmaschine erläutert. Der Schwerpunkt wird dabei auf den mathematischen Hintergründen des PageRank-Algorithmus liegen.

3.2. Die grundlegenden Schritte der Suchmaschine Google

Google führt drei grundlegende Schritte aus. Im ersten Schritt durchsuchen sogenannte Webcrawler, auch Spider genannt, das Internet. Dies sind Softwareprogramme, die das Internet durchforsten und dabei alle öffentlich zugänglichen Internetseiten erfassen. Das Crawling beginnt dabei mit einer Liste von Internetadressen, die aus vorherigen Durchsuchungen des Internets bereits bekannt sind. Zunächst besuchen die Spider alle diese Internetseiten und folgen dann sämtlichen Links, auf die diese Seiten verweisen, um dann wiederum den Links auf diesen Seiten zu folgen. So arbeiten sich die Spider durch das Internet (vgl. [7], Cho u.a., 1998, S. 1).
Im zweiten Schritt indiziert Google die Daten aus Schritt 1, um eine effiziente Durchsuchung nach relevanten Suchbegriffen, zusammen mit deren Position im Internet zu

[13] *www.netmarketshare.com*, Stand: 01.07.2015
[14] *www.forbes.com/forbes-400/list*, Stand: 26.06.2015

ermöglichen (vgl. [6], Bryan und Leise, 2006, S. 1). Der dabei entstehenden Index ist ganz ähnlich aufgebaut wie der Index eines Buches, in dem zu jedem Stichwort aufgelistet steht auf welchen Seiten des Buches dieser Begriff auftaucht. Google speichert in seinem Index somit Informationen über einzelne Suchbegriffe sowie deren Position im Internet[15]. An dieser Stelle ist anzumerken, dass die Schritte 1 und 2 nicht nacheinander, sondern bereits parallel zueinander, ablaufen.
Im dritten Schritt wird schließlich mithilfe des PageRank-Algorithmus die Wichtigkeit einer jeden Internetseite berechnet. Nach dieser können die Seiten bei Suchanfragen dann sortiert werden (vgl. [6], Bryan und Leise, 2006, S. 1).

3.3. Der PageRank-Algorithmus

Ziel des nachfolgenden Abschnitts ist es die grundlegenden mathematischen Ideen hinter dem PageRank-Algorithmus zu verdeutlichen. An dieser Stelle soll jedoch festgehalten werden, dass dieser Algorithmus in der Praxis noch mit wesentlich weitreichenderen Nebenbedingungen verbunden und insbesondere technisch von weitaus höherer Komplexität ist, als im Rahmen dieser Arbeit dargestellt werden kann.
Die nachfolgende Darlegung des PageRank-Algorithmus ist strukturell stark an dem Aufbau des Lernmoduls orientiert. Dies hat den Vorteil, dass sich bei der Diskussion der entwickelten Materialien leicht auf die mathematischen Hintergründe bezogen werden kann.

Das Random Surfer Modell

Der PageRank-Algorithmus modelliert das Verhalten eines idealen, zufällig durch das Internet navigierenden Users. Dieser *Random Surfer* besucht zu Beginn irgendeine öffentlich zugängliche Internetseite. Dann folgt er zufällig einem beliebigen ausgehenden Link auf eine andere Seite. Von dieser Seite navigiert er dann wiederum über einen beliebigen ausgehenden Link auf eine weitere Webseite. Auf diese Weise bewegt der Random Surfer sich durch das Internet fort. Von Zeit zu Zeit folgt er jedoch keinem Link, sondern springt zufällig auf eine beliebige Internetseite, indem er beispielsweise in der Adresszeile eine Internetadresse eingibt. Der Surfvorgang des Random Surfers ist nicht terminiert, d. h. er wird nicht müde immer und immer weiter durch das Internet zu surfen (vgl. [20], Wills, 2006, S. 4; [18], Page u.a., 1999, S. 5).

Diese Betrachtung des Internetnutzers als Random Surfer legt folgende Definition der relativen Wichtigkeit einer Internetseite nahe:

Definition der Wichtigkeit:
Eine Internetseite ist wichtig, wenn es wahrscheinlich ist, dass der Random Surfer sie besucht (vgl. [20], Wills, 2006, S. 4).

Das Modell zu Berechnung des PageRanks einer Internetseite, dem das Verhalten des

[15]www.google.com/insidesearch/howsearchworks/crawling-indexing.html, Stand: 02.07.2015

Random Surfers zugrunde liegt, soll in den nachfolgenden Abschnitten schrittweise erarbeitet werden. Dabei wird zunächst ein *erstes Modell* zur Berechnung des PageRanks entwickelt, welches jedoch nicht für alle beliebigen Netzwerke anwendbar ist, sodass wichtige *Modellverbesserungen* eingebaut werden. Bei der Entwicklung des PageRank-Algorithmus werden die Modellierungsschritte und auftretenden Probleme an kleinen Beispielnetzwerken veranschaulicht. Dabei handelt es sich um eben die Netzwerke, mit denen auch die Schüler im Rahmen des Modellierungstages arbeiten.

Das Internet als gerichteter Graph

Zur Entwicklung eines ersten mathematischen Modells, das die Struktur des Internets möglichst genau darstellt, wird das Internet zunächst als gerichteter Graph modelliert. Ein gerichteter Graph besteht aus Ecken (den Internetseiten) und Kanten (den Links), die alle eine Richtung aufweisen.
Angenommen das betrachtete Internet bestehe aus N Webseiten, die alle bekannt sind. Diese werden nun bezeichnet mit

$$G_i, i = 1, \ldots, N.$$

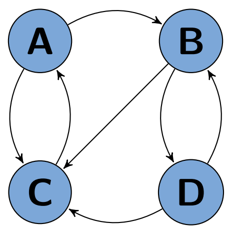


Abbildung 3: Kleines Netzwerk (angelehnt an [10], Humenberger, 2009, S. 668)

Ein kleines Beispielnetzwerk aus 4 Internetseiten ist in Abbildung 3 dargestellt. Gemäß der obigen Bezeichnungen gilt für dieses:

$$G_1 = A,\ G_2 = B,\ G_3 = C,\ G_4 = D.$$

Ein Pfeil von Seite C zu Seite A steht für einen Link von Seite C zu A usw.
Nun kann man folgende Modellannahme treffen:

1. Modellannahme:
Die Internetnutzer bewegen sich über Links fort und alle Links einer Internetseite werden mit der gleichen Wahrscheinlichkeit genutzt.

Der gerichtete Graph lässt sich dann als Übergangsgraph darstellen, bei dem jede Kante ein Kantengewicht besitzt. Die Kantengewichte stellen die Übergangswahrscheinlichkeiten dar, mit denen ein Internetnutzer von einer Internetseite zu einer anderen wechselt. Berechnen lassen sich diese, indem man die Anzahl der ausgehenden Links einer Internetseite betrachtet und gleichzeitig die erste Modellannahme berücksichtigt. Bei einer Internetseite mit m ausgehenden Links betragen die Übergangswahrscheinlichkeiten von jedem dieser ausgehenden Links folglich $\frac{1}{m}$.

Nun wird die Gesamtmenge aller Internetnutzer und deren relative Anteile auf den einzelnen Seiten betrachtet. Wir bezeichnen die relativen Anteile zu Beginn des Surfvorgangs mit $(G_1^0, G_2^0, \ldots, G_N^0)$ bzw. das Beispielnetzwerk betrachtend mit (A_0, B_0, C_0, D_0). Die relativen Anteile der Internetnutzer auf den jeweiligen Seiten nach dem n-ten Klick auf einen Link werden dann durch $(G_1^n, G_2^n, \ldots, G_N^n)$ bzw. (A_n, B_n, C_n, D_n) beschrieben. Mit diesen Bezeichnungen lassen sich rekursive Formeln aus dem Übergangsgraphen ablesen, welche die Verteilung der Internetnutzer auf die Internetseiten angeben.
Für das obige kleine Beispielnetzwerk findet man folgende Formeln, die die Verteilung der Internetnutzer nach n Zeitschritten, d. h. nach n Klicks auf beliebige Links, beschreiben:

$$\begin{array}{rlllll}
A_n = & & & 1\ C_{n-1} & \\
B_n = & 1/2\ A_{n-1} & & & + 1/2\ D_{n-1} \\
C_n = & 1/2\ A_{n-1} & + 1/2\ B_{n-1} & & + 1/2\ D_{n-1} \\
D_n = & & 1/2\ B_{n-1} & &
\end{array}$$

Das entstehende lineare Gleichungssystem lässt sich in Matrix-Vektor-Schreibweise wie folgt darstellen:

$$\underbrace{\begin{pmatrix} A_n \\ B_n \\ C_n \\ D_n \end{pmatrix}}_{=:v_n} = \underbrace{\begin{pmatrix} 0 & 0 & 1 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 1/2 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 & 0 \end{pmatrix}}_{=:U} \cdot \underbrace{\begin{pmatrix} A_{n-1} \\ B_{n-1} \\ C_{n-1} \\ D_{n-1} \end{pmatrix}}_{=:v_{n-1}}$$

bzw. kurz:

$$v_n = U \cdot v_{n-1}.$$

Allgemeiner lässt sich die Matrix U zum Übergangsgraphen aus seiner Adjazenzmatrix A entwickeln. Die Adjazenzmatrix A ist eine NxN-Matrix mit den Einträgen 0 und 1, wobei N die Anzahl der Ecken (Internetseiten) angibt. Es gilt:
Gibt es eine gerichtete Verbindung von Ecke j zu Ecke i, dann ist der Eintrag (i, j) der Matrix A gleich 1, sonst 0. Kurz:

$$A_{ij} = \begin{cases} 1, & \text{es existiert eine gerichtete Kante von } j \text{ nach } i \\ 0, & \text{sonst.} \end{cases}$$

Die Matrix U zu einem beliebigen Übergangsgraphen ergibt sich dann, indem die Einträge der Adjazenzmatrix durch die Einträge des Spaltensummenvektors geteilt werden. Die Einträge des Spaltensummenvektors lauten:

$$s_j = \sum_{i=1}^{N} A_{ij}.$$

Die Matrix U kann dann wie folgt definiert werden:

$$U_{ij} = \begin{cases} \dfrac{A_{ij}}{s_j}, & s_j \neq 0 \\ A_{ij}, & s_j = 0. \end{cases}$$

Die Einträge der Matrix U lassen sich folgendermaßen interpretieren:
Der Eintrag U_{ij} gibt die Wahrscheinlichkeit an, mit der ein Internetnutzer, der sich gerade auf Seite G_j befindet, im nachfolgenden Schritt über einen Link auf Seite G_i navigiert. Betrachtet man die Übergänge der Reihe nach, so lässt sich mithilfe der Matrixschreibweise eine einfache Beschreibung der Verteilung nach dem n-ten Zeitschritt, d. h. nach dem n-ten Klick aller Internetnutzer auf einen Link, finden (vgl. [10], Humenberger, 2009, S. 669):

$$v_1 = U \cdot v_0, \quad v_2 = U \cdot v_1 = U \cdot (U \cdot v_0) = U^2 \cdot v_0 \, , \ldots , \quad v_n = U^n \cdot v_0.$$

Um nun die Wichtigkeit einer Internetseite festlegen zu können, betrachtet man die langfristige Verteilung der Internetnutzer auf die einzelnen Seiten.
Zu betrachten ist somit der folgende Grenzwert:

$$\lim_{n \to \infty} v_n = \lim_{n \to \infty} U^n \cdot v_0.$$

Um einen eindeutigen PageRank für jede Internetseite zu erhalten, sollte dieser Grenzwert zu jeder beliebigen Startverteilung bzw. zu jedem stochastischen Vektor[16] eine eindeutige Grenzverteilung, d. h. einen eindeutigen stochastischen Vektor v liefern, dessen Einträge schließlich die Wichtigkeiten der einzelnen Internetseiten widerspiegeln. Wann tatsächliche zu jeder beliebigen Startverteilung eine eindeutige Grenzverteilung existiert, liefert der folgende Satz:

Satz von Markov

Sei $A \in \mathbb{R}^{nxn}$ eine stochastische Matrix[17]. Hat eine Matrixpotenz A^n für $n \geq 1$ nur positive Einträge, so existiert die Grenzmatrix

$$A_\infty = \lim_{n \to \infty} A^n$$

mit identischen Spalten.
Schnell lässt sich dann einsehen, dass die Grenzverteilung v mit

$$v = A_\infty \cdot v_0$$

dann unabhängig von einem beliebigen stochastischen Startvektor v_0 durch diese Spalte gegeben ist. Es gilt also

$$A_\infty = \begin{pmatrix} a_1 & a_1 & . & . & . & a_1 \\ a_2 & a_2 & . & . & . & a_2 \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ a_n & a_n & . & . & . & a_n \end{pmatrix} \text{ und } v = \begin{pmatrix} a_1 \\ a_2 \\ . \\ . \\ . \\ a_n \end{pmatrix}.$$

[16]Ein stochastischer Vektor hat reelle nicht-negative Einträge, die in der Summe 1 ergeben.
[17]Eine quadratische Matrix heißt spalten-/zeilenstochastisch, wenn alle Einträge zwischen 0 und 1 liegen und die Spalten-/Zeilensummen 1 betragen. Eine quadratische Matrix heißt stochastisch, wenn sie spaltenstochastisch ist.

Im Falle des kleinen Beispielnetzwerkes (vgl. Abb. 3) hat bereits die fünfte Potenz der Übergangsmatrix U nur positive Einträge. Nach dem Satz von Markov existiert für dieses Netzwerk somit unabhängig vom Startvektor die eindeutige Grenzverteilung v:

$$v = (\frac{3}{9}, \frac{2}{9}, \frac{3}{9}, \frac{1}{9})^T.$$

Die Grenzverteilung v stellt selbst stets einen stochastischen Vektor dar, für den insbesondere gilt:

$$A \cdot v = A \cdot (A_\infty \cdot v_0) = A_\infty \cdot v_0 = v.$$

Die Grenzverteilung v stellt also gleichermaßen die eindeutig bestimmte stationäre Verteilung zum Eigenwert 1 von A dar und die Berechnung des PageRanks lässt sich demgemäß als Eigenwertproblem auffassen (vgl. [2], Büchter und Henn, 2007, S. 332). Der Beweis zu Existenz und Eindeutigkeit eines Eigenvektors zum Eigenwert 1 einer positiven stochastischen Matrix ist Abschnitt 3.5 zu entnehmen.

Es lässt sich festhalten, dass das bisherige Modell für das oben gewählte Beispielnetzwerk (vgl. Abb. 3) eine eindeutige Grenzverteilung und somit einen eindeutigen PageRank liefert. Dies ist jedoch nicht für alle beliebigen Netzwerke der Fall, wie das nachfolgende Beispiel verdeutlicht.

Modellverbesserung 1: Senke

Betrachtet man das in Abbildung 4 dargestellte Netzwerk, so liefert die Berechnung des PageRanks mit dem bisherige Algorithmus über die zum Netzwerk gehörende Matrix U_1

$$U_1 = \begin{pmatrix} 0 & 0 & 1/3 & 0 & 0 & 0 \\ 1/2 & 0 & 1/3 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1/2 & 0 \end{pmatrix}$$

als Ergebnis für die Grenzverteilung den Nullvektor. Dies ist weder ein sinnvolles, noch ein brauchbares Ergebnis für den PageRank-Vektor.
Das Problem besteht darin, dass von Seite B keine Links ausgehen. Das Netzwerk enthält somit eine Sackgasse, die auch Senke genannt wird. Dies äußert sich in der Matrix durch eine Nullspalte. Die Matrix ist folglich nicht stochastisch und die Voraussetzungen für den Satz von Markov sind somit nicht erfüllt, sodass sich mit dem bisherigen Algorithmus kein stochastischer Grenzvektor berechnen lässt.

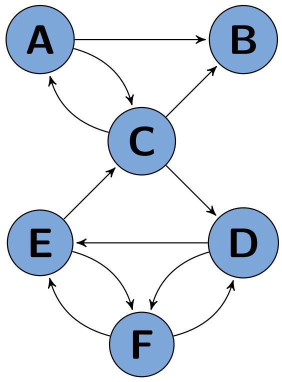


Abbildung 4: Netzwerk zur Modellverbesserung *Senke* (angelehnt an [14], Langville und Meyer, 2004, S. 3)

Um dieses Problem zu lösen wird folgende Modellannahme getroffen:

2. Modellannahme:
Navigiert der User beim Surfen durch das Internet in eine Senke, so kehrt er zu der Liste aller Internetseiten zurück und klickt nun eine beliebige der möglichen N Internetseiten an und zwar jede Seite mit der selben Wahrscheinlichkeit (vgl. [9], Humenberger, 2009, S. 154).

Dies führt dazu, dass die Nullspalte nun durch die Spalte

$$(\frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N})^T$$

ersetzt wird. Für das 2. Beispielnetzwerk ergibt sich damit die stochastische Übergangsmatrix U_1

$$U_1 = \begin{pmatrix} 0 & 1/6 & 1/3 & 0 & 0 & 0 \\ 1/2 & 1/6 & 1/3 & 0 & 0 & 0 \\ 1/2 & 1/6 & 0 & 0 & 0 & 0 \\ 0 & 1/6 & 1/3 & 0 & 0 & 1/2 \\ 0 & 1/6 & 0 & 1/2 & 0 & 1/2 \\ 0 & 1/6 & 0 & 1/2 & 1 & 0 \end{pmatrix}.$$

Nun kann der Satz von Markov angewandt werden und das erneute Berechnen der Grenzverteilung liefert ein sinnvolles Ergebnis:

$$v = (0.0755, 0.1132, 0.1698, 0.1887, 0.2264, 0.2264)^T.$$

Die Modellannahme lässt außer Acht, dass es unwahrscheinlicher ist wieder auf dieselbe Seite zu klicken. Dies fällt jedoch bei der Berechnung des PageRanks mit der zum realen Internet gehörende Matrix, die aus mehreren Billionen Zeilen besteht, kaum ins Gewicht (vgl. [10], Humenberger, 2009, S. 673).

Ein weiteres Problem, das auftreten kann und zu dem der Algorithmus in der bisherigen Version kein brauchbares Ergebnis liefert, wird am Beispiel des in Abbildung 5 dargestellten Netzwerkes deutlich.

Modellverbesserung 2: Unzusammenhängende Netzwerke

Das Ausführen des bisherigen Algorithmus mit der zu dem Netzwerk gehörenden Übergangsmatrix

$$U_2 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 1 \\ 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

liefert für unterschiedliche Startvektoren unterschiedliche Grenzverteilungen. So ergibt sich beispielsweise mit dem Startvektor

$$v_0 = (1/5, 1/5, 1/5, 1/5, 1/5)^T$$

die Grenzverteilung

$$v = (0.2, 0.2, 0.2667, 0.1333, 0.2)^T,$$

mit dem Startvektor

$$v_0 = (1, 0, 0, 0, 0)^T$$

hingegen die Grenzverteilung

$$v = (1, 0, 0, 0, 0)^T.$$

Folglich kann den einzelnen Internetseiten nicht unabhängig von der gewählten Startverteilung ein eindeutiger PageRank zugeordnet werden.

Problematisch ist an dieser Stelle, dass der zu dem Netzwerk gehörende Graph nicht zusammenhängend ist[18]. Dieses Problem wird durch die folgenden Modellannahme aufgegriffen:

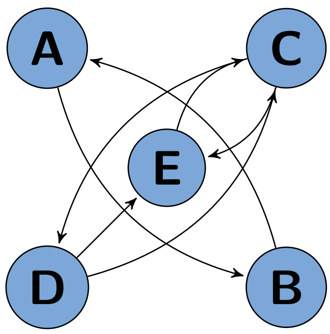


Abbildung 5: Netzwerk zur Modellverbesserung *Unzusammenhängende Netzwerke*

[18] Genau genommen tritt neben dem Problem des unzusammenhängenden Netzwerkes zudem das Problem eines kreisförmigen (Teil-)Netzes zwischen den Seiten A und B auf. Dieses wird jedoch ebenfalls durch die nachfolgende 3. Modellannahme aufgegriffen, weshalb darauf nicht näher eingegangen wird.

3. Modellannahme:
Der User folgt mit einer Wahrscheinlichkeit von $0 < p < 1$ den Links und mit der Wahrscheinlichkeit $(1 - p)$ springt er zufällig auf eine beliebige Internetseite (vgl. [9], Humenberger, 2009, S. 154).

Diese Modellannahme lässt sich folgendermaßen mathematisch beschreiben:
Wird den Links auf den Internetseiten gefolgt, so ist die Übergangsmatrix gemäß des bis hierhin entwickelten Algorithmus, d. h. im Beispiel gemäß der Übergangsmatrix U_1, gegeben. Betrachtet man nur das zufällige Springen des Internetnutzers auf beliebige Internetseiten, beispielsweise indem er oben in der Adresszeile eine Internetadresse eingibt, so lässt sich dies durch eine weitere Übergangsmatrix U_3 beschreiben. Jeder Eintrag dieser Matrix beträgt $\frac{1}{N}$, wobei $N > 0$ der Gesamtzahl aller Internetseiten, bzw. allgemeiner der Anzahl der Ecken des gerichteten Graphen, entspricht. Verknüpft man diese beiden Matrizen unter Berücksichtigung der angegebenen Wahrscheinlichkeiten, so ergibt sich eine neue Übergangsmatrix T gemäß:

$$T = p \cdot U_1 + (1 - p) \cdot U_3$$

Da p zwischen 0 und 1 liegt und zudem sowohl die Matrix U_1, als auch die Matrix U_3 stochastisch sind, ist auch T stochastisch. Des Weiteren hat die Matrix T nur positive Einträge. Der Satz von Markov ist somit erfüllt und der Algorithmus liefert stets eine von der Startverteilung unabhängige eindeutige stochastische Grenzverteilung.
Die Erfinder Page und Brin verwendeten bei der Berechnung des PageRanks anfänglich den Wert $p = 0.85$. Mit welcher Wahrscheinlichkeit die Suchmaschine jedoch heutzutage arbeitet, ist ein gut gehütetes Geheimnis des Unternehmens Google Inc. (vgl. [9], Humenberger, 2009, S. 63; [20], Wills, 2006, S. 7).

3.4. Iterative Berechnung des PageRanks in der Praxis

Sowohl die Berechnung des PageRanks über das Lösen des Eigenwertproblems

$$v = T \cdot v$$

als auch die Berechnung über die Bestimmung riesiger Matrixpotenzen gemäß

$$\lim_{n \to \infty} T^n$$

ist für eine gigantische Übergangsmatrix mit mehr als 25 Billionen von Zeilen und Spalten, wie es im Falle des realen Internets zutrifft, ungeeignet.
In der Praxis wird die Grenzverteilung deswegen iterativ angenähert. Dazu startet der Algorithmus mit einer beliebigen Startverteilung v_0 und berechnet dann schrittweise

$$v_{r+1} = T \cdot v_r$$

solange, bis der Fehler

$$\|v_{r+1} - v_r\|$$

genügend klein geworden ist (vgl. [20], Wills, 2006, S. 9).

3.5. Beweise zum PageRank-Algorithmus

In diesem Kapitel soll der Beweis dafür geliefert werden, dass der PageRank-Algorithmus in der oben dargelegten endgültigen Form immer ein brauchbares Ergebnis liefert. Dazu muss gezeigt werden, dass zu der Übergangsmatrix T, aus der finalen Fassung des PageRank-Algorithmus, stets eine eindeutige positive stochastische Grenzverteilung existiert.
Da die Erfinder Page und Brin den PageRank ursprünglich als Lösung eines Eigenwertproblems definierten, wird nachfolgend nicht der Beweis des Grenzwertsatzes (Markov) geführt, sondern der Beweis von Satz 1, der die Existenz eines eindeutigen positiven stochastischen Eigenvektors $\mathbf{v}$ zum Eigenwert 1 der Googlematrix T garantiert. Die Beweisführung orientiert sich dabei an dem Artikel *The \$25,000,000,000 Eigenvector: The Linear Algebra behind Google* von Bryan und Leise (vgl. [6], Bryan und Leise, 2006).

Satz 1

Sei $A \in \mathbb{R}^{nxn}$ eine positive[19] spaltenstochastische Matrix. Dann hat die Matrix A den Eigenwert 1. Weiter existiert zu diesem ein eindeutiger Eigenvektor $\mathbf{v}$. Der Vektor $\mathbf{v}$ ist stochastisch und hat nur positive Einträge. Zudem lässt sich $\mathbf{v}$ über

$$\mathbf{v} = \lim_{n\to\infty} A^n x_0$$

für jeden beliebigen positiven Startvektor x_0 mit $\|x_0\| = 1$ berechnen.

Für den Beweis dieses Satzes, der das Funktionieren des PageRank-Algorithmus garantiert, werden zunächst einige vorbereitende Lemmata benötigt.

Lemma 2

Jede spaltenstochastische Matrix $A \in \mathbb{R}^{nxn}$ hat den Eigenwert 1.

Beweis.
Da A spaltenstochastisch ist, ist die transponierte Matrix A^T zeilenstochastisch und es gilt für

$$e = (1, 1, \ldots, 1)^T \text{ mit } e \in \mathbb{R}^n,$$

dass

$$A^T e = e.$$

Somit hat die transponierte Matrix A^T den Eigenwert 1 und damit auch die Matrix A. Es existiert folglich ein Eigenvektor v zum Eigenwert 1. □

Nachfolgend bezeichne $V_1(A)$ den Eigenraum zum Eigenwert 1 einer spaltenstochastischen Matrix A. Es ist zu zeigen, dass der Eigenraum stets eindimensional ist. Dazu werden zunächst die folgenden Lemmata gezeigt.

[19] Eine Matrix heißt positiv, wenn sie nur positive Einträge hat.

Lemma 3

Sei $A \in \mathbb{R}^{nxn}$ eine positive spaltenstochastische Matrix. Dann hat jeder Eigenvektor $v \in V_1(A)$ nur positive oder nur negative Einträge.

Beweis.
Angenommen, es existiert ein $w \in V_1(A)$, sodass w Einträge mit unterschiedlichen Vorzeichen hat.
Aus

$$w = Aw$$

folgt

$$w_i = \sum_{j=1}^{n} A_{ij} w_j.$$

Die Verwendung der Dreiecksungleichung führt zu folgender Ungleichung:

$$|w_i| = |\sum_{j=1}^{n} A_{ij} w_j| < \sum_{j=1}^{n} A_{ij} |w_j|.$$

Die strikte Ungleichung gilt, da A positiv ist und somit mindestens einer der Summanden $A_{ij} w_j$ ein negatives Vorzeichen hat.
Durch Aufsummieren der Einträge von $i = 1$ bis n der obigen Ungleichung, sowie Vertauschen der Summen über i und j ergibt sich:

$$\sum_{i=1}^{n} |w_i| < \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} |w_j| = \sum_{j=1}^{n} |w_j| \sum_{i=1}^{n} A_{ij} = \sum_{j=1}^{n} |w_j|.$$

Hierbei wurde verwendet, dass A spaltenstochastisch ist und somit $\sum_{i=1}^{n} A_{ij} = 1$ für alle j gilt. Damit ergibt sich insgesamt, dass

$$\sum_{i=1}^{n} |w_i| < \sum_{j=1}^{n} |w_j|,$$

was einen Widerspruch darstellt. □

Für den Beweis der Eindimensionaliät des Eigenraums $V_1(A)$ und damit der Eindeutigkeit des Eigenvektors benötigt man noch folgendes Lemma:

Lemma 4

Für zwei linear unabhängige Vektoren v und w aus dem $\mathbb{R}^n$ mit $n \geq 2$ existieren zwei Zahlen $s, t \in \mathbb{R}$, die nicht beide null sind, sodass der Vektor

$$x = sv + tw$$

sowohl positive als auch negative Einträge besitzt.

Beweis.
Aus der linearen Unabhängigkeit der Vektoren v, w folgt:

$$v \neq 0, w \neq 0.$$

Sei die Summe der Einträge des Vektors v

$$d = \sum_{i=1}^{n} v_i.$$

1. Fall: $d = 0$
Dann muss der Vektor v Einträge mit unterschiedlichem Vorzeichen haben. Wählt man $s = 1$ und $t = 0$, so folgt die Behauptung.
2. Fall: $d \neq 0$
Dann definiere man

$$s = -\frac{\sum_{i=1}^{n} w_i}{d}, t = 1 \text{ und } x = sv + tw.$$

Da v und w linear unabhängig sind, ist $x \neq 0$ und damit gilt:

$$\sum_{i=1}^{n} x_i = 0.$$

Somit hat x sowohl positive als auch negative Einträge, was den Fall $d \neq 0$ beweist. □

Lemma 5
Sei $A \in \mathbb{R}^{nxn}$ eine positive spaltenstochastische Matrix. Dann ist $V_1(A)$ eindimensional.

Beweis.
Angenommen, es gibt linear unabhängige Vektoren v und w mit $v, w \in V_1(A)$.
Für zwei beliebige reelle Zahlen s und t, von denen nicht beide null sind, gilt dann:

$$0 \neq x = sv + tw \in V_1(A).$$

Mit Lemma 3 folgt dann, dass alle Einträge von x das gleiche Vorzeichen haben. Aus Lemma 4 folgt jedoch, dass es s und t geben muss, sodass x sowohl negative als auch positive Einträge hat, was einen Widerspruch darstellt. Somit folgt

$$\dim(V_1(A)) = 1.$$ □

Insgesamt hat man also bereits gezeigt, dass der Eigenvektor zum Eigenwert 1 bis auf einen Faktor eindeutig ist und alle Einträge des Eigenvektors das gleiche Vorzeichen haben. Für das Funktionieren des PageRank-Algorithmus bleibt zu zeigen, dass das iterative Verfahren für beliebige Startvektoren gegen diesen eindeutigen Eigenvektor konvergiert. Dazu betrachte man zunächst das folgende Lemma:

Lemma 6

Sei $A \in \mathbb{R}^{nxn}$ eine positive spaltenstochastische Matrix und sei V der Teilraum des $\mathbb{R}^n$ der alle Vektoren enthält, sodass

$$\sum_{j=1}^{n} v_j = 0.$$

Dann gilt für alle $v \in V$:

$$Av \in V \text{ und } \|A \cdot v\|_1 \leq c\|v\|_1,$$

wobei c eine Konstante ist, für die gilt

$$c = \max_{1\leq j\leq n} |1 - 2 \min_{1\leq i\leq n} A_{ij}| < 1.$$

Beweis.

Um zu zeigen, dass für beliebige $v \in V$ gilt, dass $Av \in V$, setzt man

$$w = Av.$$

Dann ist

$$w_i = \sum_{j=1}^{n} A_{ij} v_j.$$

Durch Aufsummieren über alle Einträge i erhält man

$$\sum_{i=1}^{n} w_i = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij} v_j = \sum_{j=1}^{n} v_j \Big(\sum_{i=1}^{n} A_{ij}\Big) = \sum_{j=1}^{n} v_j = 0.$$

Somit gilt $w \in V$.

Um die Ungleichung bezüglich der 1-Norm zu zeigen betrachte man zunächst

$$\|w\|_1 = \sum_{i=1}^{n} e_i w_i = \sum_{i=1}^{n} e_i (\sum_{j=1}^{n} A_{ij} v_j),$$

wobei

$$e_i = \operatorname{sgn}(w_i).$$

Dabei sei zu beachten, dass e_i nicht für alle Einträge von w das gleiche Vorzeichen haben kann, da $\sum_{i=1}^{n} w_i = 0$ gilt. Durch Vertauschen der Summen ergibt sich

$$\|w\|_1 = \sum_{j=1}^{n} v_j (\sum_{i=1}^{n} e_i A_{ij}) = \sum_{j=1}^{n} b_j v_j,$$

wobei

$$b_j = \sum_{i=1}^{n} e_i A_{ij}.$$

Da der Vektor v Einträge mit unterschiedlichen Vorzeichen hat und $\sum_{i=1}^{n} A_{ij} = 1$ gilt, wobei $0 < A_{ij} < 1$ ist, lässt sich folgende Abschätzung machen:

$$-1 < -1 + 2 \min_{1 \leq i \leq n} A_{ij} \leq b_j \leq 1 - 2 \min_{1 \leq i \leq n} A_{ij} < 1.$$

Damit ergibt sich, dass

$$|b_j| \leq |1 - 2 \min_{1 \leq i \leq n} A_{ij}| < 1.$$

Sei nun

$$c := \max_{1 \leq j \leq n} |1 - 2 \min_{1 \leq i \leq n} A_{ij}|,$$

dann gilt $c < 1$ und $|b_j| \leq c$ für alle j. Mit der obigen Gleichung folgt dann

$$\|w\|_1 = \sum_{j=1}^{n} b_j v_j = |\sum_{j=1}^{n} b_j v_j| \leq \sum_{j=1}^{n} |b_j||v_j| \leq c \sum_{j=1}^{n} |v_j| = c\|v\|_1,$$

was zu zeigen war. □

Damit hat man nun alle Voraussetzungen um den Satz 1 zu beweisen:

Beweis.
Nach Lemma 2 hat die Matrix A den Eigenwert 1. Mit Lemma 5 folgt zudem, dass der Eigenraum $V_1(A)$ eindimensional ist. Des Weiteren gilt nach Lemma 3 für jeden Vektor $w \in V_1(A)$, der nicht der Nullvektor ist, dass dieser nur positive oder nur negative Einträge hat. Weiterhin ist klar, dass es einen eindeutigen Vektor $\mathbf{v} \in V_1(A)$ gibt, der nur positive Einträge hat, sodass gilt

$$\sum_{i=1}^{n} \mathbf{v}_i = 1.$$

Damit ist $\mathbf{v}$ der eindeutige positive stochastische Eigenvektor zum Eigenwert 1, dessen Einträge im PageRank-Algorithmus die relativen Wichtigkeiten der Internetseiten darstellen.
Um zu zeigen, dass dieser sich gemäß

$$\mathbf{v} = \lim_{n \to \infty} A^n x_0$$

berechnen lässt, betrachte man einen beliebigen positiven stochastischen Vektor $x_0 \in \mathbb{R}^n$. Dann lässt sich x_0 schreiben als

$$x_0 = \mathbf{v} + w \text{ mit } w \in V \text{ (mit } V \text{ wie in Lemma 6).}$$

Dann erhält man

$$A^n x_0 = A^n \mathbf{v} + A^n w = \mathbf{v} + A^n w$$

und somit

$$A^n x_0 - \mathbf{v} = A^n w.$$

Induktion und Anwenden von Lemma 6 liefert dann

$$\|A^n w\|_1 \leq c^n \|w\|_1 \text{ für } 0 \leq c < 1,$$

wobei c wie in Lemma 6 gewählt wird. Aus

$$\lim_{n\to\infty} c^n \|w\| = 0$$

und durch Anwenden des Einschnürungskriteriums erhält man

$$\lim_{n\to\infty} \|A^n w\|_1 = 0.$$

Damit erhält man schließlich

$$\lim_{n\to\infty} A^n x_0 = \mathbf{v},$$

was zu zeigen war (vgl. [6], Bryan und Leise, 2006). □

4. Didaktisch-methodisches Konzept

> „Mathematik entwickelt sich im Wechselspiel von Theorie und praktischer Anwendung, sie trägt zum Verständnis und zur Gestaltung der uns umgebenden Welt bei. Das Modellieren ist der Prozess der Strukturierung von Sachsituationen, der Beschreibung außermathematischer Realität durch mathematische Begriffe und Zusammenhänge (Mathematisierung) sowie der Nutzung mathematischer Zusammenhänge zur Lösung realer Probleme, der anschließenden Interpretation des Ergebnisses und der Validierung des Modells" (vgl. [17], Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, 2014, S. 16).

Das entwickelte Lernmodul soll Schülern der Sekundarstufe II die Möglichkeit bieten eigene Erfahrungen vom praktischen Nutzen der Mathematik zu sammeln und damit ihr Verständnis von der Mathematikhaltigkeit ihrer Umwelt zu erweitern. Indem sich die Schüler im Rahmen eines Modellierungstages aktiv mit der alltagsnahen Fragestellung *Wie funktioniert eigentlich Google und was hat das mit Mathe zu tun?* auseinandersetzen, soll zudem der Aufbau der mathematischen Modellierungskompetenz unterstützt werden - einer vor dem Hintergrund der Herausbildung mündiger Bürger zentral erscheinenden mathematischen Kompetenz, wie es sowohl die Kernlehrpläne des Landes Nordrhein-Westfalen, als auch die verbindlichen Bildungsstandards mehrfach betonen (vgl. Abschnitt 2.1.2).

In diesem Kapitel soll das Konzept des ausgearbeiteten Lernmoduls sowie die entwickelten Materialien und deren Einbindung in einen mathematischen Modellierungstag im Rahmen eines CAMMP days vorgestellt werden.

4.1. Ziele

Bei der Konzipierung des Lernmoduls zum Thema *Google* im Rahmen dieser Arbeit wurden insbesondere die nachfolgend dargelegten Ziele verfolgt, die sowohl mit den Zielen von CAMMP (vgl. Abschnitt 2.2.2) als auch mit den allgemein von mathematischer Modellierung verfolgten Zielen (vgl. Abschnitt 2.1.4) eng verbunden sind.
Zum einen sollen die Schüler die Schritte des *Modellierungsprozesses*, der in Abschnitt 2.1.3 beschrieben ist, anhand eines realen Problems, das sich an ihrer Lebenswelt orientiert, *durchlaufen*. Die einzelnen Modellierungsschritte sollen dabei soweit wie möglich *eigenständig* von den Schülern absolviert werden. Weiterhin sollen die zu bearbeitenden *Aufgaben problemorientiert* sein und hinsichtlich des Lösungsweges möglichst *offen*. Das Lernmodul soll so ausgearbeitet sein, dass es von Lerngruppen mit *unterschiedlichen inhaltlichen Voraussetzungen*[20] absolviert werden kann und zudem die *Heterogenität* innerhalb einer Lerngruppe gefördert wird. Des Weiteren sollen die Schüler durch die Auseinandersetzung mit dem realen Problem der Internetseitensortierung, aber auch durch

[20]Das Lernmodul soll sich sowohl an Leistungskurse als auch an Grundkurse richten.

gemeinsame Reflexion mit den Betreuern des Modellierungstages ihr *Wissen über Modellierung und den Modellbildungsprozess erweitern*. Übergeordnet soll das Lernmodul dazu beitragen Sinn, Bedeutung und auch Nutzbarkeit von Mathematik für das alltägliche Leben herauszuheben und damit das Bild von Mathematik zu verbessern (vgl. Abschnitt 2.1.4).

Nachdem im nachfolgenden Abschnitt kurz der Ablauf des Modellierungstages erläutert wird, werden in Abschnitt 4.3 die entwickelten Materialien beschrieben. Anschließend werden die durchgeführten Modellierungstage dargelegt und auf der Grundlage einer Schülerbefragung sowie meiner eigenen Beobachtungen diskutiert, inwieweit die mit dem ausgearbeiteten Lernmodul angestrebten Zielen erreicht wurden.

4.2. Ablauf des Modellierungstages

Für die reine Bearbeitung des Lernmoduls durch die Schüler sind 2,5 Zeitstunden anberaumt. Zudem gibt es eine halbstündige Einführung sowie eine halbstündige Nachbesprechung, sodass für den Modellierungstag in Summe 3,5 Zeitstunden eingeplant sind[21].
Ablauf:

1. Der Modellierungstag beginnt am Morgen mit einer kurzen *Begrüßung*. CAMMP und die unterschiedlichen Veranstaltungsangebote von CAMMP werden vorgestellt. Direkt im Anschluss wird ein *Modellierungvortrag*[22] gehalten, der einen Einblick in die Welt der mathematischen Modellierung ermöglicht.

2. Ein *Problemeinführungsvortrag*, der die Suchmaschine Google und die Problemstellung vorstellt, wird von den Betreuern gehalten. Die Schüler erhalten einen Überblick über die Geschichte und den Erfolg von Google. Zudem werden die grundlegenden Schritte der Suchmaschine erläutert. Anschließend erhalten die Schüler eine kleine Einführung in das Programm MATLAB[23].

3. Die Schüler beginnen mit der Bearbeitung der folgenden Modellierungsschritte:
   Schritt 1: Bestimmung der relativen Anteile der Internetnutzer
   Schritt 2: Umschreiben in Matrix-Vektor-Schreibweise
   Schritt 3: Aufstellen von Netzwerken
   Schritt 4: Bestimmung der Grenzverteilung
   Schritt 5: Unabhängigkeit von der Startverteilung
   Schritt 6: Manipulation der Wichtigkeit
   Zur Bearbeitung erhalten die Schüler das *Arbeitsblatt 1*, welches Aufgabenstellungen zu allen sechs Schritten beinhaltet. Außerdem erhalten die Schüler die *Kleine*

[21] Die tatsächliche Bearbeitungsdauer des Lernmoduls hängt natürlich stark von der jeweiligen Schülergruppe und deren Motivation bzw. deren Vorwissen ab.
[22] Dieser Vortrag ist für alle Lernmodule, die im Rahmen von CAMMP days durchgeführt werden, gleich und musste nicht mehr von mir entwickelt werden.
[23] MATLAB ist eine höhere Programmiersprache und interaktive Umgebung für numerische Berechnungen, Programmierung und Visualisierung. *www.de.mathworks.com*, Stand: 28.07.2015

Einführung in Matrizen. Zudem stehen *Hilfekarten* zu den Modellschritten 1 und 3 zu Verfügung, die der Betreuer den Schülern bei Bedarf aushändigt.
Zur Bearbeitung der Modellierungsschritte arbeiten die Schüler mit den vorbereiten MATLAB-Skripten *kleines_netzwerk.m* und *google.m*. Schülern, die die einzelnen Schritte besonders schnell bearbeiten, kann ein *Infoblatt* zum Satz von Markov ausgeteilt werden.

4. Vor der Mittagspause findet eine Sicherung der Ergebnisse statt. Dazu werden die Modellierungsschritte von einzelnen Schülern an der Tafel vorgestellt. Dann wird die anfängliche Problemstellung der Internetseitensortierung aufgegriffen und im Plenum diskutiert, wie die Wichtigkeit einer eigenen Seite erhöht werden kann.

5. Mittagspause.

6. Als Zusammenfassung des bis hierhin entwickelten Modells zur Berechnung des PageRanks wird der *Vortrag zur Sicherung des ersten Modells* gehalten. Anschließend erhalten die Schüler das *Arbeitsblatt 2*, mit dem die folgenden fünf Schritte bearbeitet werden:
Schritt 1 + 2: Erarbeitung der Modellverbesserung *Senke*. In Schritt 2 sammeln die Schüler in Kleingruppen zunächst Ideen zur Lösung des Problems *Senke*, die dann in einer kurzen frontalen Phase gesammelt werden. Der Betreuer stellt schließlich die von den Google Erfindern getroffene Modellannahme vor.
Schritt 3: Erfassung des Problems *Unzusammenhängende Netzwerke*.
Schritt 4 + 5: Erarbeitung des endgültigen Modells.
Zur Bearbeitung der Modellverbesserungsschritte arbeiten die Schüler mit dem vorbereiteten MATLAB-Skript *komplexes_netzwerk.m*. Zu dem dritten Modellverbesserungsschritt steht den Schülern eine *Hilfekarte* zur Verfügung.

7. Kurz vor Ende des Modellierungstages findet eine erneute Sicherung statt. Die Schüler stellen das endgültige Modell vor.

8. Es werden *Evaluationsbögen* an die Schüler ausgeteilt.

9. Bei einer kurzen Verabschiedung erhalten die Schüler Informationsmaterialien zu weiteren Angeboten von CAMMP und dem Studiengang CES.

4.3. Materialien

Das entwickelte Lernmodul ist so aufgebaut, dass die Schüler den gesamten Modellierungsprozess vollständig, jedoch in kleineren Teilschritten durchlaufen.
Die Aufteilung des gesamten Prozesses in einzelne Teilprozesse wurde bewusst vorgenommen, um die Komplexität des Modellierungsprozesses zu reduzieren und um auf diese Weise geeignete Materialien und Aufgaben zu konzipieren, die von den Schülern im Laufe eines einzigen Modellierungstages bewältigt werden können.

Eine solche Herangehensweise an das Modellieren ermöglicht die Schulung einzelner Teilkompetenzen und damit langfristig den Aufbau einer umfassenden Modellierungskompetenz (vgl. [8], Greefrath u.a., 2013, S. 19). Wie in Abschnitt 4.2 bereits deutlich wurde, arbeiten die Schüler vorrangig mit zwei Arbeitsblättern und zugehörigen MATLAB-Skripten. Überdies stehen dem Betreuer zur Einarbeitung in das Thema weiterführende Materialien sowie Lösungen zu den einzelnen Aufgaben zur Verfügung.
Nachfolgend werden zunächst die für die Schüler relevanten Materialien einschließlich der Präsentationen in der Reihenfolge dargelegt, wie sie während des Modellierunstages zum Tragen kommen. Dies soll dazu dienen, die mit den Materialien intendierten Ziele und die Orientierung dieser am Modellierunskreislauf übersichtlicher darzustellen. Lediglich die ebenfalls für die Schüler vorgesehenen MATLAB-Skripte sowie die weiterführenden Materialien für die Betreuer werden anschließend gesondert beschrieben.

4.3.1. Modellierungsvortrag

Der Modellierungsvortrag lag bereits in der angehängten Form vor (vgl. Anhang C.1) und musste nicht mehr von mir entwickelt werden. Dennoch soll er an dieser Stelle kurz beschrieben werden, da er einen wichtigen Bestandteil des Modellierungstages darstellt. Der Vortrag, der vorzugsweise von einem an der Organisation von CAMMP beteiligten Professoren gehalten wird, liefert den Schülern eine Antwort auf die Frage, was mathematische Modellierung ist. Dazu werden die einzelnen Schritte des Modellierungsprozesses am Modellierungskreislauf erläutert und mit kleinen Beispielen aus dem Bereich der Astrophysik[24] veranschaulicht. Überdies hebt der Vortrag die große Bedeutung von mathematischer Modellierung für zahlreiche Berufe aus Forschung, Industrie und Wirtschaft hervor.

4.3.2. Problemeinführungsvortrag

Der von den Betreuern gehaltene Problemeinführungsvortrag (vgl. Anhang C.2) liefert zunächst allgemeine Informationen zu der Suchmaschine Google. Die Erfinder der Suchmaschine werden vorgestellt und die Geschichte der Entstehung von Google wird kurz umrissen. Zudem werden interessante Fakten, wie die Tatsache, dass das Verb *googlen* mittlerweile Einzug in den Duden[25] gehalten hat, dass die Erfinder Googles zu den Top 20 der weltweit reichsten Menschen gehören oder, dass das Original Paper, in dem die Erfinder der Suchmaschine die Funktionsweise und die mathematischen Ideen erläutern, frei zugänglich im Internet zu finden ist.
Dieser allgemein informierende Einstieg dient dazu, zunächst das Interesse und die Motivation der Schüler an dem Thema des Modellierungstages zu wecken.
Anschließend werden die grundlegenden Schritte der Suchmaschine Google erläutert. Im Zuge dessen wird der PageRank-Algorithmus erstmalig erwähnt und die Problemstellung der Sortierung von Internetseiten nach deren Wichtigkeiten und die damit auf den Plan

[24] Ein Beispiel stellt die fehlgeschlagene Modellierung der Flugbahn des Mars Polar Landers, einer Sonde der NASA, dar (vgl. Anhang C.1).
[25] www.duden.de/rechtschreibung/googeln, Stand: 03.08.2015

gerufene Fragestellung, wann eine Seite überhaupt wichtig sein sollte, wird benannt. Eben diese Fragestellung wird dann gemeinsam mit den Schülern diskutiert. Der Betreuer sammelt dabei die Vorschläge der Schüler und hält schließlich die Definition der Wichtigkeit nach den Erfindern Google's fest. Als übergeordnetes Ziel wird schließlich die Bestimmung der *Verteilung der Internetnutzer mit idealisiertem Surfverhalten* auf die einzelnen Internetseiten benannt.

Abschließend wird zu dem Modellierungskreislauf, der im Modellierungsvortrag bereits detailliert vorgestellt wurde, Bezug genommen. Die Fragestellung der sinnvollen Internetseitensortierung wird als *Reales Problem* festgehalten und der erste Schritt des Modellierungskreislaufes (vgl. Abb. 2) vorgenommen. Dazu wird das reale Problem vereinfacht, indem anstelle des gesamten Internets nur ein kleines Netzwerk aus vier Internetseiten betrachtet wird. Das *Vereinfachte Problem* wird dann in ein *mathematisches Modell* überführt, indem das Netzwerk als gerichteter Graph dargestellt wird. Dazu werden die definierenden Eigenschaften eines gerichteten Graphen kurz erläutert und der Bezug zu der gegebenen Situation hergestellt. Diese Schritte des Modellierungsprozesses werden bewusst von den Betreuern übernommen, da insbesondere die Darstellung als gerichteter Graph ein neues Konzept für die Schüler darstellt, das in dieser Form nicht in der Schule behandelt wird.
Die Einbettung des Modellierungskreislaufes in das Lernmodul und die gemeinsame Thematisierung des Kreislaufes mit den Schülern soll als Orientierungshilfe im Verlaufe des Modellierungstages dienen. Des Weiteren kann der Einbau des Modellierungskreislaufes als metakognitives, strategisches Element betrachtet werden, welches die Reflektion des Modellierungsprozesses durch die Schüler unterstützen und damit den Aufbau der mathematischen Modellierungskompetenzen erleichtern soll. Obgleich bisherige Studien zum Einsatz des Modellierungskreislaufes als Lernhilfe noch keine einheitlichen Befunde aufzeigen konnten, stellt dies dennoch einen vielversprechenden Ansatz bei der Ausbildung der Modellierungskompetenz dar (vgl. [3], Besser u.a., 2015, S. 50).
Anschließend öffnet der Betreuer das MATLAB-Skript *google.m*[26] (vgl. Anhang G.4) und gibt an der gekennzeichneten Stelle

```
%Startwebseite
site = 'http://www.aachen.de/';
```

eine beliebige Internetseite ein. Durch aktivieren des Skripts[27], öffnet sich eine Grafik, die ein Teilnetz d. h. einen Teil der realen Verlinkungsstruktur der eingegebenen Seite darstellt. Dies dient der Veranschaulichung des Netzwerkbegriffs und der Bezugnahme zum Aufbau der Verlinkungsstruktur von Internetseiten, wie sie tatsächlich im Internet existieren. An diesem Teilnetz kann die Idee der Darstellung von Netzwerken als gerichtete Graphen überdies nochmal verdeutlicht werden.
Bevor die Schüler im Anschluss an den Vortrag mit der Bearbeitung der Aufgaben be-

[26] Dieses musste nicht mehr von mir entwickelt werden, da es bereits in der angehängten Form vorlag.
[27] Dieses Skript ruft mit einem Befehl automatisch das Skript *surfer.m* auf, welches die Verlinkungsstruktur der gewählten Seite erfasst.

ginnen, erhalten sie eine kleine Einführung zum Umgang und zur Funktionsweise der Software MATLAB. Darauf wird in Abschnitt 4.3.7 genauer eingegangen.

4.3.3. Arbeitsblatt 1 und Zusatzmaterialien

Nachdem mit dem Einführungsvortrag die Problemstellung klar dargelegt wurde, ist das erste Arbeitsblatt (vgl. Anhang A.1) so aufgebaut, dass an einem kleinen Beispielnetzwerk ein erstes Modell zur Berechnung der Wichtigkeiten entwickelt wird.
Das Arbeitsblatt beginnt mit einer kurzen Zusammenfassung der wichtigsten Informationen aus dem Einführungsvortrag, sowie einer kurzen Problembeschreibung. Auf diese Weise haben die Schüler sowohl die gegeben Situation als auch die konkrete Problemstellung jederzeit vor Augen.

Zunächst sollen die Schüler mithilfe des MATLAB-Skripts *google.m*, welches in der Einführung bereits von dem Betreuer vorgestellt wurde, eigenständig zu selbst gewählten Internetseiten Teilnetze erstellen lassen. Anschließend beginnen die Schüler mit der Bearbeitung der Aufgaben, welche sich bei diesem Arbeitsblatt allesamt an dem im Einführungsvortrag dargestellten vereinfachten Problem, der Sortierung der Internetseiten eines kleinen Netzwerks aus nur 4 Internetseiten, orientieren.
Die erste Aufgabe des Arbeitsblattes dient dazu ein Verständnis für die relativen Anteile von Internetnutzern und von Übergangswahrscheinlichkeiten zu entwickeln. Dazu wird eine 1. Modellannahme formuliert, die das Verhalten eines idealisierten Internetnutzers beschreibt. Anhand des auf dem Arbeitsblatt abgebildeten Übergangsgraphen ermitteln die Schüler zunächst Formeln zur Berechnung der relativen Anteile nach genau einem Internetseitenwechsel, um dann im nächsten Schritt allgemeine Formeln für die rekursive Berechnung der Verteilung nach beliebig vielen Internetseitenwechseln aufzustellen.
Für Schüler, die mit dem Ablesen der Formeln aus dem Übergangsgraphen Schwierigkeiten haben, wurde eine Hilfekarte entwickelt, die bei Bedarf von dem Betreuer ausgeteilt wird (vgl. Anhang B). Auf dieser Hilfekarte finden die Schüler ein Netzwerk aus nur drei Seiten mit einem kleinen Beispiel zur Berechnung der relativen Anteile. Das Aufstellen der Formeln für die relativen Anteile können die Schüler dann von diesem Beispiel auf die erste Aufgabe übertragen.

Das im ersten Schritt entstehende lineare Gleichungssystem soll in der zweiten Aufgabe in Matrix-Vektor-Schreibweise umgeschrieben werden. Da das entwickelte Lernmodul dem Anspruch der Heterogenität innerhalb einer Lerngruppe genügen soll und zudem für den Einsatz in verschiedenen Jahrgangsstufen der Oberstufe, die unterschiedliches Vorwissen mitbringen, geeignet sein soll, kann an dieser Stelle nicht davon ausgegangen werden, dass Matrizen und Vektoren allen Schülern bekannt sind. Aufgrund dessen erhalten die Schüler das Informationsblatt *Einführung in Matrizen* (vgl. Anhang A.2). Dieses beschreibt alle für dieses Lernmodul wichtigen Eigenschaften und Rechenregeln zu Matrizen und veranschaulicht diese an kleinen Beispielen. Dabei wurde darauf geachtet nur das Mindestmaß an notwendigen Informationen darzustellen, um die Schüler nicht mit an dieser Stelle belanglosen Informationen zu überfordern. Anhand dieser Einführung in Matrizen können die Schüler das Gleichungssystem aus der ersten Aufgabe nun in Matrix-Vektor-Schreibweise umformulieren.

Um sicher zu stellen, dass den Schülern der Zusammenhang zwischen aufgestellter Matrix und Netzwerk bewusst ist, werden in der dritten Aufgabe zwei Matrizen vorgegeben. Anhand derer sollen die Schüler beurteilen, ob diese ein Netzwerk beschreiben und dieses dann gesetzt dem Fall aufzeichnen. Ohne den Begriff *Übergangsmatrix* explizit zu nennen, wird an dieser Stelle dennoch die Bedeutung der Matrix als Vermittlerin eines Überganges deutlich. Mithilfe dieser Aufgabe soll insbesondere der Bezug zwischen der Matrix, einem abstrakten Objekt der Mathematik, und dem Netzwerk, als Beschreibung der gegebenen Situation und damit als Teil der Wirklichkeit, vertiefend aufgezeigt werden. Schüler, die noch große Probleme bei der Interpretation der Matrizen haben, erhalten eine zu dieser Aufgabe entwickelte Hilfekarte (vgl. Anhang B). Auf dieser ist eine Matrix sowie die Interpretation einer Matrixspalte in Worten dargestellt.

In der vierten Aufgabe soll schrittweise die langfristige Verteilung der Internetnutzer, die dann als Maß für die Wichtigkeit einer Seite dient, bestimmt werden. Dazu ermitteln die Schüler zunächst die geschlossene Formel[28] für die Verteilung nach einem konkreten[29] und schließlich nach beliebig vielen Zeitschritten.
Mit Blick auf das Ziel, die relativen Wichtigkeiten zu bestimmen, welches zu Beginn der Aufgabe vier auch schriftlich noch einmal hervorgehoben wird, berechnen die Schüler mithilfe der zuvor entwickelten Formel die Verteilung der Internetnutzer nach sehr vielen Zeitschritten. Dazu bestimmen sie hohe Potenzen der zu dem Netzwerk gehörenden Übergangsmatrix U und nähern die Grenzverteilung an.
Um die Interpretation der Ergebnisse zu sichern, sollen die Schüler die Einträge des erhaltenen Verteilungsvektors als Wichtigkeiten deuten und die vier Seiten des Netzwerkes nach ihrem PageRank anordnen. Zu diesem Zeitpunkt des Modellierungstages haben die Schüler ein *erstes Modell* zur Berechnung der relativen Wichtigkeiten entwickelt und den Modellierungskreislauf, wenngleich stark angeleitet, vollständig durchschritten.

Die fünfte Aufgabe dient dazu wichtige Eigenschaften des ersten Modells zu verdeutlichen. Indem die Schüler in der ersten Teilaufgabe die Startverteilung der Internetnutzer variieren[30], entdecken sie, dass die langfristige Verteilung und damit auch die Wichtigkeit der Seiten in diesem Fall nicht von der Startverteilung abhängt.
Die zweite Teilaufgabe lässt die Schüler auf die Möglichkeit der Berechnung des Verteilungsvektors als Bestimmung eines Eigenvektors stoßen. Dabei wird jedoch der Begriff des Eigenwertes bzw. des Eigenvektors nicht eingeführt, wiederum aufgrund unterschiedlichen Vorwissens verschiedener Lerngruppen bzw. variierender Kenntnisstände innerhalb einer Schülergruppe. Die an dieser Stelle auftretende Frage, inwiefern ein kurzes

[28] $v_n = U^n \cdot v_0$ (vgl. Abschnitt 3.3)
[29] Nachdem alle Internetnutzer 4 Links gefolgt sind.
[30] Sie müssen sich dazu zunächst bewusst machen, welche Eigenschaften eine Startverteilung erfüllen muss.

Anreißen der Betrachtung des PageRanks als Lösung des linearen Gleichungssystems[31] für das Verständnis des PageRank-Algorithmus nützlich ist bzw. zu nachhaltigen Erkenntnisse führt, wird in der Auswertung (vgl. Kapitel 6) diskutiert.

Die sechste und letzte Aufgabe dieses Arbeitsblattes bietet den Schülern die Möglichkeit das erworbene Wissen zu Definition und Berechnung der Wichtigkeit kreativ einzusetzen, indem sie das gegebene kleine Netzwerk manipulieren. Dazu können sie Links hinzufügen oder entfernen um die Wichtigkeit einer ausgesuchten Internetseite zu erhöhen.

Zusätzlich zum ersten Arbeitsblatt kann den Schülern ein *Informationsblatt* (vgl. Anhang A.2) zum Satz von Markov ausgeteilt werden. Dieses Informationsblatt liefert den Schülern die mathematische Begründung dazu, warum und wann das entwickelte Modell die gewünschten Resultate liefert. Indem dieses zusätzliche Informationsblatt an Schüler ausgeteilt wird, die schneller als die übrigen Schüler ihrer Lerngruppe das erste Arbeitsblatt vollständig bearbeitet haben, soll ein individuelles Lern- und Arbeitstempo ermöglicht werden.
An dieser Stelle ist jedoch zu beachten, dass die in das Infoblatt integrierte Aufgabenstellung, der Überprüfung einer Folgerung aus dem Satz von Markov, nur von Schülern bearbeitet werden kann, die bereits vor dem Modellierungstag Kenntnisse über Rechenoperationen für Matrizen besessen haben.

Alle Berechnungen zu den Aufgaben des ersten Arbeitsblattes werden mithilfe des vorbereiteten MATLAB-Skripts *kleines_netzwerk.m* durchgeführt. Auf den genauen Aufbau dieses Skripts wird in Abschnitt 4.3.7 explizit eingegangen.

4.3.4. Vortrag zur Sicherung des ersten Modells

Nachdem die Schüler mithilfe des ersten Arbeitsblattes ein erstes Modell zur Berechnung der Wichtigkeit bzw. des PageRanks bestimmt haben, findet nun eine von den Betreuern moderierte Sicherung im Plenum statt. Dazu werden die Aufgaben von einzelnen Schülern an der Tafel vorgestellt. Während der Diskussion der Ergebnisse ist seitens der Betreuer darauf zu achten, dass insbesondere die Interpretation der erhaltenen mathematischen Resultate bezogen auf die reale Problemstellung im Vordergrund steht, um ein tiefgehendes und problemorientiertes Verständnis der Schüler zu sichern.
Bei der Diskussion der Ergebnisse wird sich an dem *Vortrag zur Sicherung des ersten Modells* (vgl. Anhang C.4) orientiert, welcher die Abfolge der Modellierungsschritte visualisiert. Dies dient während der Diskussion sowohl dem Betreuer als auch den Schülern als Orientierungshilfe und unterstützt gleichzeitig die Reflexion der absolvierten Modellierungsschritte durch die Schüler, indem die Ergebnisse nicht nur auf der symbolischen, sondern auch auf der ikonischen Ebene repräsentiert werden[32].

[31] $U \cdot v = v$ (vgl. Abschnitt 3.3)
[32] Man vergleiche dazu die drei von Bruner festgehaltenen Repräsentationsmodi.

4.3.5. Arbeitsblatt 2 und Zusatzmaterialien

Die erste Aufgabe des Arbeitsblattes (vgl. Anhang A.3) lässt die Schüler auf das Problem *Senke* stoßen. Dazu wenden sie ihren bisherigen Algorithmus an, um die Wichtigkeiten der Internetseiten eines komplexeren Netzwerkes, welches eben dieses Problem beinhaltet, zu berechnen. Da sie mit dem bisherigen Modell kein sinnvolles Ergebnis für die Wichtigkeit der Seiten erhalten, müssen Modellverbesserungen eingebaut werden. Dazu sollen die Schüler in der nächsten Aufgabe, das Netzwerk betrachtend, eigenständig überlegen worauf das auftretende Problem zurück zuführen ist und welche Möglichkeiten sie zur Problemlösung sehen.
Anschließend findet eine kurze Diskussion im Plenum statt, in der die Schüler ihre Modellannahmen, die zur Lösung des Problems führen könnten, vorstellen und bewerten. Schließlich präsentiert der Betreuer die Idee der Google Erfinder, die dann von den Schülern eigenständig in das Modell eingebaut wird.

An dieser Stelle gelangt man an einen kritisch zu reflektierenden Punkt des entwickelten Lernmoduls. Einerseits sollen die Schüler möglichst kreativ und schöpferisch arbeiten können, was einschließen würde, dass sie mit ihren eigenen Modellannahmen weiterarbeiten, andererseits ist es das übergeordnete Ziel des Lernmoduls den PageRank-Algorithmus in seiner tatsächlich bestehenden Form zu erarbeiten. Um die Modellannahmen und Vorschläge der Schüler dennoch zu würdigen, sollte während der Diskussion insbesondere auf die jeweiligen Vor- und Nachteile bzw. die Umsetzungsmöglichkeiten der Schülervorschläge eingegangen werden.
Nachdem die Schüler die Modellannahme in das Modell eingebaut haben, erhalten sie durch erneute Berechnung der Grenzverteilung ein sinnvolles Ergebnis für den PageRank der einzelnen Seiten.

In der dritten Aufgabe erkunden die Schüler an einem weiteren Netzwerk ein zweites Problem, zu dem der bisherige Algorithmus noch kein sinnvolles Ergebnis liefert. Indem die Schüler zu einem abgebildeten Netzwerk die Grenzverteilung mithilfe unterschiedlicher selbstgewählter Startverteilungen berechnen, erfassen sie, dass die Grenzverteilung in diesem Fall nicht unabhängig vom gewählten Startvektor ist. Diese Problematik sollen die Schüler anhand des Netzwerkes interpretieren und darauf zurückführen, dass das Netzwerk unzusammenhängend ist. Haben Schüler Schwierigkeiten dieses Problem zu erfassen, kann ihnen eine Hilfekarte ausgeteilt werden (vgl. Anhang B).

Im Anschluss an diese Aufgabe wird eine dritte Modellannahme auf dem Arbeitsblatt festgehalten, die das zufällige Springen des Internetnutzers mit einbezieht. Diese Modellannahme löst das Problem der *Unzusammenhängende Netzwerke* und führt schließlich zu dem finalen Modell des PageRank-Algorithmus.

In Aufgabe vier stellen die Schüler zunächst den ersten Aspekt der Modellannahme, das zufällige Springen des Internetnutzers auf beliebige Internetseiten in mathematischer Form als Matrix dar. Die fünfte und letzte Aufgabe dient dazu, die bereits getroffenen

Modellannahmen zu verknüpfen und das endgültige Modell des PageRank-Algorithmus zu ermitteln. Dazu berechnen die Schüler die Wichtigkeiten des Netzwerkes aus der ersten Aufgabe dieses Arbeitsblattes erneut und halten explizit fest, welches die wichtigste Seite darstellt.
Bei der Bearbeitung des zweiten Arbeitsblattes arbeiten die Schüler mit dem vorbereiteten MATLAB-Skript *komplexes_netzwerk.m*.

Im Anschluss an die Bearbeitung des zweiten Arbeitsblattes findet eine weitere Sicherung im Plenum statt. Dazu leitet der Betreuer wiederum die Diskussion zu den Aufgaben und regt die Präsentation der Ergebnisse durch die Schüler an der Tafel an. Wie bei der Sicherung der Ergebnisse des ersten Arbeitsblattes, sollte der Betreuer auch an dieser Stelle den Fokus auf die Diskussion der erhaltenen Ergebnisse im Bezug auf die reale Problemstellung legen.
Zudem nimmt der Betreuer in dieser abschließenden Phase des Modellierungstages erneut Bezug zum Modellierungskreislauf, indem er die Bedeutung des Überprüfens von Resultaten und die ggf. resultierende Notwendigkeit des Einbauens von Modellverbesserungen hervorhebt. Dies kann den Schülern helfen den Modellierungsprozess im Ganzen zu reflektieren und zudem verdeutlichen, dass das bestmögliche mathematische Resultat zur Beschreibung einer gegebenen realen Problemstellung häufig nicht nach einmaligen Durchlaufen des Modellierunskreislaufes erhalten wird.

4.3.6. Materialien für die Betreuer

Der übergeordnete Anspruch an die entwickelten begleitenden Materialien für die Betreuer war diese so aufzubauen, dass sie alle notwendigen allgemeinen, mathematischen sowie didaktisch-methodischen Informationen und Hinweise zu Thematik und Durchführung liefern, sodass sich die betreuende Lehrperson schnell einarbeiten kann und sich zudem ausreichend vorbereitet fühlt. Eine über die entwickelten Materialien hinausgehende Einarbeitung oder Recherche sollte nicht notwendig sein.

Das zusätzliche Material für die Betreuer umfasst zunächst das *Basic Paper* (vgl. Anhang D), in dem die Hintergründe der Suchmaschine Google sowie alle mathematischen Grundlagen des Lernmoduls in anschaulicher und verständlicher Weise beschrieben werden. Zudem erhalten die Betreuer ein *Methodisches Konzept* (vgl. Anhang E), welches den genauen Ablauf des Modellierungstages und den Einsatz der einzelnen Schülermaterialien erläutert.
Des Weiteren gibt es zu beiden Arbeitsblättern *Lösungsblätter* (vgl. Anhang A.4 und A.5), an denen sich die Betreuer bei der Besprechung der Aufgaben orientieren können. Auch gibt es Lösungen zu den MATLAB-Skripten, auf welche die Betreuer bei Schwierigkeiten der Schüler, bezüglich der Eingabe ihrer Überlegungen in MATLAB, zurückgreifen können. Überdies gibt es *Hinweise zum Problemeinführungsvortrag* (vgl. Anhang C.3), die den Betreuenden Anregungen geben, welche Informationen sie an welcher Stelle des Vortrages an die Schüler weitergeben können.

Durch diese Materialien soll die Einarbeitungs- bzw. Vorbereitungszeit zur Durchführung des Modellierungstages minimiert und dem Lehrenden ein maximales Gefühl von Sicherheit bei der Durchführung ermöglicht werden. Weiterführend ist diese Materialsammlung auch dieserart entwickelt und strukturiert worden, um interessierten Mathematiklehrern diese Sammlung weiterreichen und ihnen mögliche Bedenken zur Durchführung des Lernmoduls im eigenen Unterricht nehmen zu können.

4.3.7. Die MATLAB-Codes

> „[...] die Ergebnisse internationaler Vergleichsstudien (PISA / IGLU) [deuten] darauf hin, dass deutsche Schüler im Reproduzieren formaler Kenntnisse und bei der Anwendung von Routineverfahren zwar zufriedenstellende Leistungen erzielen – sobald es aber um Problemlösung und um das Entwickeln und Beschreiben eigener Lösungsansätze geht, schneiden sie deutlich schwächer ab. Anscheinend finden gerade die prozessbezogenen Kompetenzen immer noch nicht genügend Beachtung im Mathematikunterricht. Nach wie vor scheint ein Unterricht weit verbreitet zu sein, in dem die Mathematik auf den Erwerb formaler Rechenfertigkeiten reduziert wird und die Möglichkeiten aktiv-entdeckenden Lernens nicht ausgeschöpft werden. Gerechtfertigt wird dieser fertigkeitsorientierte Unterricht seitens der Lehrer meist mit der enormen Stofffülle, die nur wenig Raum für „entdeckende und forschende“ Problemaufgaben lässt” (vgl. [1], Basten, 2011).

Das Lernmodul wurde mit Computerunterstützung entwickelt, da gerade die weitere Schulung formaler Fertigkeiten, die im Mathematikunterricht häufig überbetont wird, nicht im Vordergrund des Modellierungstages stehen soll.
Unter Verwendung einer geeigneten Computersoftware soll der Fokus des Lernmoduls auf der problemorientierten Bearbeitung der Aufgaben durch die Schüler, die auf diese Weise von der Rechenarbeit entlastet werden, liegen. Damit soll insbesondere die Förderung der prozessbezogen Kompetenz des Modellierens, aber auch die des Problemlösens in den Mittelpunkt rücken.

Dass der Rechenaufwand ohne Computereinsatz bei einem Großteil der Aufgaben, welche die Schüler im Rahmen des Lernmodul bearbeiten, zu zeitaufwendig und damit nicht zu leisten wäre, ist ein weiteres Argument, das den Einsatz des Computers bei diesem Workshop begründet.

Das Lernmodul wurde im MATLAB erstellt, welches eine hochentwickelte Programmiersprache und interaktive Umgebung darstellt, die unter anderem für numerische Berechnungen sowie Datenanalyse und insbesondere zum disziplinübergreifenden Arbeiten genutzt wird.

In der Regel kommen die Schüler im Rahmen des Modellierungstages zum ersten Mal in Kontakt mit MATLAB, sodass sie zu Beginn des Workshops eine kurze, interaktive Einführung in die Software erhalten (vgl. Anhang G.1). Dazu erklärt der Betreuer frontal die wichtigsten Funktionsweisen und den Aufbau von MATLAB und zeigt an kleinen Beispielaufgaben, wie Befehle eingegeben werden und die Berechnung aktiviert wird. Simultan sollen auch die Schüler die vom Betreuer gezeigten Befehle ausführen, um ein erstes Gefühl für den Umgang mit MATLAB zu erhalten.

Das Lernmodul soll zwar die Möglichkeit bieten den Computer als hilfreiches Modellierungswerkzeug kennen und nutzen zu lernen, dennoch soll das Erlernen der verwendeten Software MATLAB nicht im Vordergrund des Lernmoduls stehen und den Modellierungstag dominieren. Aufgrund dessen wurde bei der Entwicklung der MATLAB-Skripte darauf geachtet diese so übersichtlich, intuitiv und einfach wie möglich zu gestalten.
Anhand des nachfolgenden Auszuges aus dem MATLAB-Skript *kleines_netzwerk.m* (vgl. Anhang G.2), welches die Schüler bei der Bearbeitung des ersten Arbeitsblattes verwenden, soll der Aufbau der Skripte kurz erläutert werden.

```
%% Schritt 1
% Zunaechst speichern wie die Anfangsverteilung
$A_0 = 0.25;
B_0 = 0.25;
C_0 = 0.25;
D_0 = 0.25;

%Gib hier deine Formeln fuer A_1 bis D_1 ein
A_1 = NaN
B_1 = NaN
C_1 = NaN
D_1 = NaN

% Um deine Loesung zu ueberpruefen druecke auf "Run Section"
ueberpruefe_Loesung('Aufgabe 1',A_1,B_1,C_1,D_1);
```

Die MATLAB-Skripte dokumentieren genau an welchen Stellen im Code die Schüler ihre Lösungen zu den jeweiligen Aufgaben eingeben sollen. Dabei ist von den Schülern der Ausdruck *NaN*[33] durch eine Formel oder eine Matrix zu ersetzen.
Wenn die Schüler die Berechnungen ausführen lassen, werden nicht nur ihre Ergebnisse ausgeben, sondern sie erhalten zudem eine Rückmeldung darüber, ob ihre Lösung korrekt ist oder nicht. Zum Abgleich der Lösungen wurde ein weiteres MATLAB-Skript geschrieben, welches die korrekten Lösungen enthält. Durch Aufruf des Befehls

```
ueberpruefe_Loesung,
```

der in das MATLAB-Skript der Schüler integriert ist, wird dieses zusätzliche Skript, welches die Lösungen enthält, im Hintergrund ausgeführt. Die Schülerlösungen werden automatisch mit der korrekten Lösung verglichen und die Schüler erhalten schließlich einen der folgenden Hinweise:

[33]NaN steht für *Not a Number*.

- Bitte korrigiere deine Lösung.
- Ein Teil deiner Lösung ist korrekt.
- Super. Deine Lösung ist richtig.

Indem das zusätzliche MATLAB-Skript in eine sogenannte *.p-Datei*[34] konvertiert wurde, ist es den Schülern jedoch nicht möglich diese Datei zu öffnen und die Lösungen einzusehen.

Durch Verwendung der *Überprüfe-Funktion* erhalten die Schüler zu jeder Aufgabe unmittelbar nach der Bearbeitung eine individuelle Rückmeldung zu ihren Überlegungen. Eine gemeinsame Sicherung nach jeder einzelnen Aufgabe wird damit überflüssig und die Schüler können eigenständig und in selbstbestimmten Lerntempo weiterarbeiten.

[34] Dies ist ein sogenannter *protected function file*, der durch den Befehl *pcode(Beispieldatei.m)* erzeugt wird und der dann nicht mehr einsehbar ist.

5. Durchführung

Im Rahmen dieser Arbeit wurde das Lernmodul im Zuge von CAMMP days mit zwei verschiedenen Schülergruppen durchgeführt. Die beiden durchgeführten Modellierungstage lagen in einem Zeitraum von drei Wochen, sodass nach der ersten Durchführung die Möglichkeit bestand notwendig erscheinende Verbesserungen an den Arbeitsblättern sowie am Code vorzunehmen und auch die Präsentationen umzustrukturieren.

5.1. Leitende inhaltliche Gesichtspunkte

Nachfolgend wird die Durchführung der CAMMP days, sowie die dabei festgehaltenen Beobachtungen beschrieben. Im Zuge dessen wird insbesondere auf die folgenden leitenden inhaltlichen Gesichtspunkte eingegangen, die sich von den in Abschnitt 4.1 dargelegten Zielen ableiten lassen.

- Welche Verbesserungen waren nach der 1. Durchführung notwendig und inwieweit haben die Verbesserungen bei der 2. Durchführung das Gewünschte geleistet?
- Inwieweit war das Vorwissen der Schüler ausreichend für die Bearbeitung der Aufgaben?
- Inwieweit hatten die Schüler Schwierigkeiten die Aufgaben zu lösen?
- Wie viel Hilfestellung durch die Betreuer war notwendig? Inwieweit konnten die Schüler die Aufgaben eigenständig lösen? Kamen die Hilfekarten zum Einsatz und inwieweit haben sie den Schülern geholfen?
- Inwieweit wurde die Heterogenität innerhalb der Lerngruppen gefördert bzw. ermöglicht?
- Inwieweit sind die Schüler in der Lage, die durch das mathematische Modell erhaltenen Ergebnisse auf die Realsituation zu beziehen?
- Gab es noch weitere auffällige Beobachtungen, die die Modellierungstage prägten?

Die leitenden Gesichtspunkte dienten dazu, bei der Durchführung fokussierte Beobachtung und bei der Auswertung eine strukturierte Darstellung der Beobachtungsergebnisse zu ermöglichen.

5.2. An den leitenden Gesichtspunkten orientierte Beobachtungen

Die oben genannten Leitfragen sollen nachfolgend zunächst aus meiner Sicht als Betreuerin der beiden Modellierungstage beantworten werden, um dann in Abschnitt 5.3 die Rückmeldungen durch die Schüler auszuwerten. Zunächst werden jedoch kurz die Rahmenbedingungen bei der Durchführung der beiden CAMMP days beschrieben.

5.2.1. Durchgeführte Modellierungstage im Überblick

1. Durchführung

Die Lerngruppe: Der erste Modellierungstag wurde mit einem Mathematikleistungskurs der Jahrgangsstufe Q1 des Rhein-Maas-Gymnasiums[35] durchgeführt. Der Kurs bestand aus 17 Schülern und wurde von einem Lehrer begleitet.

Vorwissen der Lerngruppe: Die Schüler kannten sowohl die Matrix-Vektor-Schreibweise für lineare Gleichungssysteme, als auch Übergangsmatrizen und stochastische Prozesse bereits aus dem Mathematikunterricht. Der Satz von Markov sollte laut dem Lehrer in einer der nächsten Mathematikstunden thematisiert werden. Der Umgang mit MATLAB war für alle Schüler neu.

Geplante Dauer des CAMMP days: Für den CAMMP day war ein Zeitrahmen von 9:00 bis 15:00 Uhr mit einer einstündigen Mittagspause angesetzt. Als reine Bearbeitungszeit des Lernmoduls durch die Schüler waren 3,5 Zeitstunden anberaumt.

2. Durchführung

Die Lerngruppe: Der zweite Modellierungstag wurde mit einem Mathematikleistungskurs der Jahrgangsstufe Q1 des Gymnasiums Herzogenrath durchgeführt. Der Kurs bestand aus 24 Schülern und wurde von einem Lehrer begleitet.

Vorwissen der Lerngruppe: Die Schüler kannten bereits Vektoren (als Elemente des $\mathbb{R}^3$) aus dem Mathematikunterricht. Matrizen sowie die Matrix-Vektor-Schreibweise für lineare Gleichungssysteme waren noch nicht bekannt. Der Umgang mit MATLAB war für alle Schüler neu.

Geplante Dauer des CAMMP days: Für den CAMMP day war ein Zeitrahmen von 9:00 bis 14:30 Uhr mit einer einstündigen Mittagspause angesetzt. Als reine Bearbeitungszeit des Lernmoduls durch die Schüler waren 3,0 Zeitstunden anberaumt.

5.2.2. Darlegung der konkreten Beobachtungsergebnisse zu den leitenden Gesichtspunkten

Bei der 1. Durchführung des Modellierungstages wurde die Einführung in MATLAB noch nicht frontal gestaltet. Dies führte zu großer Verwirrung seitens der Schüler, die augenscheinlich mit dem Umgang der für sie neuen Software überfordert waren. Bei der 2. Durchführung wurde die Einführung aufgrund dessen frontal vom Betreuer gestaltet und simultan von den Schüler ausgeführt. Außerdem wurde betont, dass MATLAB als ein großer Taschenrechner aufgefasst werden kann, der aufwendige Rechenarbeit erspart. Bei der anschließenden Bearbeitung der Aufgaben hatten die Schüler weniger Fragen und Probleme beim Umgang mit MATLAB.

[35] Ein Aachener Gymnasium mit naturwissenschaftlichem Schwerpunkt.

Bei dem 1. Modellierungstag sollten die Schüler die Datei *google.m* noch eigenständig und ohne vorherige Erläuterung durch die Betreuer erkunden. Viele Schüler kamen damit nicht gut zurecht, was ich an den zahlreichen Fragen festmachte, die die Schüler zu der Bedeutung und dem Nutzen der generierten Grafiken hatten.
Nachdem bei der 2. Durchführung die Datei *google.m* zunächst von den Betreuern im Zuge des Problemeinführungsvortrages erläutert wurde, war den Schülern anschließend beim eigenständigen Erstellen der Teilnetze klar, was die erzeugten Grafiken aussagen.

Die Diskussion zur Wichtigkeit einer Internetseite bereits in den Problemeinführungsvortrag einzubauen, war eine weitere Verbesserung die nach der 1. Durchführung vorgenommen wurde. Diese Änderung wurde eingebaut, um die Schüler dazu anzuregen eigene Ideen und Vorschläge zu entwickeln wann eine Seite wichtig ist. Auf diese Weise wurden die Schüler schon bei der Problemeinführung aktiv eingebunden. Außerdem sollte so sichergestellt werden, dass die Definition der relativen Wichtigkeit einer Internetseite von den Schülern verstanden und überdies als sinnvoll erachtet wird.

Das entwickelte Lernmodul soll so aufgebaut sein, dass verschiedene Lerngruppen der Oberstufe mit unterschiedlichem inhaltlichen Vorwissen mit den Aufgaben weder unter- noch überfordert sind. Da die erste Lerngruppe mehrstufige durch Übergangsmatrizen vermittelte Prozesse bereits im Unterricht behandelt hatte, die zweite Lerngruppe hingegen noch nicht über Wissen zu Matrizen oder Austauschprozessen verfügte, lässt sich dies an dieser Stelle gut vergleichen.
Die Schüler der ersten Lerngruppe kamen sehr gut mit den Aufgaben zurecht. Sie haben diese konzentriert bearbeitet und es gab wenige Fragen oder Unklarheiten. Die Bearbeitungszeit wurde um 45 verkürzt, da bereits alle Schüler das Lernmodul vollständig bearbeitet hatten. Es wurden keine Hilfekarten benötigt.
Die zweite Schülergruppe hingegen hatte zu den einzelnen Aufgaben einige Fragen und zudem wurden vereinzelt Hilfekarten ausgeteilt. Die Bearbeitungszeit wurde wie angesetzt benötigt. Mit dem Umschreiben von linearen Gleichungssystemen in Matrix-Vektor-Schreibweise und auch bei der Bearbeitung der nachfolgenden Aufgaben, bei denen Rechenoperationen mit Matrizen durchgeführt wurden, hatten die Schüler kaum Probleme. Dies wurde daran festgemacht, dass die Schüler wenige Fragen hatten und die Aufgaben zudem von allen Schülern korrekt gelöst werden konnten. Die zum ersten Arbeitsblatt ausgeteilten Hilfekarten lieferten den Schülern, die mit einzelnen Aufgaben nicht gut zurecht kamen, konkrete Orientierungsbeispiele, sodass sie die jeweilige Aufgabe im Anschluss eigenständig lösen konnten.

Bei der 1. Durchführung waren einige Schüler früher mit den Aufgaben fertig als der Rest ihrer Lerngruppe, sodass sie das *Infoblatt* zum Satz von Markov erhielten. Bei der zweiten Lerngruppe hingegen konnte das Infoblatt nicht an schnellere Schüler ausgeteilt werden, da es Kenntnisse über die konkrete Berechnung eines Matrix-Vektor Produktes voraussetzt. Schüler aus dieser Lerngruppe, die die Aufgaben schneller lösten, mussten demzufolge warten, bis auch alle übrigen Schüler das Arbeitsblatt vollständig bearbeitet hatten.

Bei der gemeinsamen Sicherung der Ergebnisse nach dem ersten und dem zweiten Arbeitsblatt waren die präsentierenden Schüler beider Lerngruppen in der Lage die erhaltenen Lösungen in Bezug zu der realen Situation zu setzen. Sie konnten die Einträge der erhaltenen Grenzverteilungen als Wichtigkeiten der einzelnen Internetseiten interpretieren. Auch konnten sie erläutern, wie die vorgenommenen Änderungen an den Matrizen mit der geänderten Betrachtung des Internetnutzerverhaltens bzw. den im Zuge dessen getroffenen Modellannahmen zusammenhängen.

5.3. Schülerbefragung

Jeweils am Ende der beiden durchgeführten Modellierungstage wurde eine Evaluation in Form einer schriftlichen anonymen Befragung durchgeführt. Die im Zuge dessen an die Schüler ausgeteilten Fragebögen sind Anhang F.1 zu entnehmen[36]. Diese Befragung sollte einerseits dazu dienen, den Erfolg des Lernmoduls hinsichtlich der in Abschnitt 4.1 festgelegten Ziele zu evaluieren, andererseits sollte sie Anhaltspunkte für die Überarbeitung und Optimierung des Lernmoduls mit Blick auf weitere Durchführungen liefern.

Die Schülerbefragung zeigte, dass die Schüler durch den Workshop ein besseres Verständnis von mathematischer Modellierung entwickelt haben. So gaben mehr als die Hälfte der insgesamt 41 befragten Schüler an, dass es „voll zutrifft", dass sie durch den Workshop mathematische Modellierung besser begriffen haben. Dies ist auch zahlreichen Kommentaren der Schüler zu der Frage *Was hast du für dich persönlich durch die Teilnahme am Workshop gelernt?* zu entnehmen, zu der die Antwort einige Male „neues Wissen über mathematische Modellierung" lautete. Neben dem Verständnis von mathematischer Modellierung geben einige Schüler an, dass die Bedeutung und der Nutzen von Mathematik für die Lösung realer Probleme durch das Lernmodul deutlich geworden sind. So schrieb ein Schüler, dass er / sie gelernt habe, dass „viele Probleme mathematisch lösbar" sind.

Einige Schüler kritisierten auf die Frage *Was würdest du am Workshop verändern bzw. verbessern wollen?*, dass sie sich „mehr offene Arbeit, weniger Aufgaben" gewünscht hätten. Auch der / die begleitende Lehrer / in des Gymnasiums Herzogenrath gab mir gegenüber an, dass „Möglichkeiten zum individuellen, freien Weiterarbeiten für Schüler fehlen, die die kleinschrittigen Aufgaben schnell bewältigen".

Die Schülerbefragung ergab zudem, dass der Schwierigkeitsgrad der Aufgaben für die beiden Schülergruppen mit unterschiedlichem Vorwissen angemessen war. So gab aus beiden Schülergruppen niemand die anzukreuzende Möglichkeit „trifft voll zu", bezüglich der Aussage, dass die Aufgaben zu schwierig oder zu einfach waren, an.

[36] Diese bestanden bereits in der vorliegenden Form und wurden nicht von mir entwickelt.

6. Zusammenfassende Auswertung

In diesem Kapitel sollen die wichtigsten Ergebnisse der Schülerbefragung sowie meiner eigenen Beobachtungen zusammenfassend festgehalten und Vor- und Nachteile des Lernmoduls diskutiert werden. Schlussendlich wird kritisch reflektiert, inwieweit die in Abschnitt 4.1 aufgeführten Ziele an das ausgearbeitet Lernmodul erreicht wurden.

Die im Verlauf des Modellierungstages von den Schülern zu bearbeitenden Aufgaben stellen überwiegend die Interpretation der Ergebnisse im Bezug auf das Internetnutzerverhalten in den Vordergrund. Auch mithilfe der, bei der Ergebnissicherung geführten Diskussionen, wird der Schwerpunkt auf die Deutung der erhaltenen Ergebnisse in der realen Situation gelegt. Lediglich die dritte Teilaufgabe der Aufgabe fünf des ersten Arbeitsblattes (vgl. Anhang A.1), in der die Schüler die PageRank-Berechnung als Lösung eines linearen Gleichungssystems erfassen, ist nicht an der realen Situation orientiert und lässt sich nicht bezogen auf diese diskutieren. Es ist fraglich inwieweit die Schüler diese Aufgabe als nützlich oder sinnvoll erachten, da sie nicht zur Entwicklung des PageRank-Algorithmus beiträgt und auch nachfolgend nicht mehr aufgegriffen wird.

Die Aufgaben auf den Arbeitsblättern sind kleinschrittig aufgebaut und bieten überwiegend nicht die Möglichkeit individuelle Lösungsstrategien zu verfolgen[37]. Dies ist zum Teil bereits der Natur der dem Workshop zugrundeliegenden Problemstellung geschuldet. So ist es das Ziel des Modellierungstages, dass die Schüler eine Antwort auf die Frage *Wie funktioniert eigentlich Google und was hat das mit Mathe zu tun?* erhalten. Das Herzstück von Google, den PageRank-Algorithmus, zu modellieren steht somit im Vordergrund. Dieser Algorithmus besteht jedoch in einer bestimmten Form und arbeitet mit definierten Schritten, sodass bei der Entwicklung der Aufgaben wenig Spielraum besteht, Lösungsansätze zu ermöglichen, die von diesen definierten Schritten abweichen.

Durch die in das Lernmodul integrierte *Einführung in Matrizen* ist das erfolgreiche absolvieren des Modellierungstages unabhängig davon, ob die Schüler sich in der Schule bereits mit Matrizen auseinander gesetzt haben. Dies wurde insbesondere durch den Vergleich der beiden Lerngruppen, der in Abschnitt 5.2.2 vollzogen wurde, unterstrichen. Die Durchführung dieses Lernmoduls kann somit je nach Vorwissen der Schüler zum einen als Einführung in neue Inhalte genutzt werden oder zum anderen der Vertiefung und Anwendung bereits erworbenen Wissens dienen. Das Lernmodul knüpft an folgende Kompetenzerwartungen und inhaltliche Schwerpunkte in der Qualifikationsphase aus dem Kernlehrplan des Landes Nordrhein-Westfalen an:
„Die Schülerinnen und Schüler

- verwenden die Matrizenmultiplikation zur Untersuchung stochastischer Prozesse (Vorhersage nachfolgender Zustände, numerisches Bestimmen sich stabilisierender Zustände).

[37] Lediglich die letzte Aufgabe des ersten Arbeitsblattes bietet den Schülern die Möglichkeit bezüglich des Lösungsweges kreativ und frei zu arbeiten.

- beschreiben stochastische Prozesse mithilfe von Zustandsvektoren und stochastischen Übergangsmatrizen.
- stellen lineare Gleichungssysteme in Matrix-Vektor-Schreibweise dar" (vgl. [17], Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, 2014, S. 28 ff.).

Mit Blick auf die von CAMMP erwünschte Weitergabe der Materialien an interessierte Lehrer, bietet das entwickelte Lernmodul somit die Möglichkeit mathematische Modellierung in den Unterricht einzubauen ohne dabei gleichermaßen *Zeit* für die Umsetzung der vom Lehrplan vorgeschriebenen Inhalte *zu verlieren*. Insbesondere durch die zahlreichen zusätzlichen Materialien für die Betreuer, wie beispielsweise das *Basic Paper* oder das *Methodische Konzept*, sollten interessierte Lehrer alle notwendigen Hintergrundinformationen und Anleitungen zur Durchführung des Lernmoduls in ihrem eigenen Unterricht haben.

Die entwickelte automatische Überprüfung der Schülerlösungen mithilfe der *Überprüfe-Funktion* (vgl. Abschnitt 4.3.7) ermöglicht ein individuelles Arbeitstempo während der Bearbeitung der jeweiligen Arbeitsblätter. Neben dem *Infoblatt* zum Satz von Markov gibt es jedoch keine Möglichkeiten zum individuellen Weiterarbeiten für schnellere Schüler. Insbesondere in der zweiten Phase des Modellierungstages, nach erfolgreicher Bearbeitung des zweiten Arbeitsblattes, wären zusätzliche Materialien für schneller arbeitende Schüler sinnvoll. Dem Anspruch, die Heterogenität innerhalb einer Lerngruppe zu fördern, wird das Lernmodul somit nur teilweise gerecht.

> „Die Inhaltsfelder Analysis, analytische Geometrie und lineare Algebra sowie Stochastik sind nicht isoliert nebeneinander zu betrachten, vielmehr werden sie konzeptionell vernetzt (z. B. durch übergreifende Konzepte wie funktionaler Zusammenhang, Mittelwert, Kumulation, Iteration, Grenzwert). Wo möglich sollten fächerverbindende Aspekte, insbesondere im Zusammenhang mit Naturwissenschaften und Technik, aber auch den Sozialwissenschaften Berücksichtigung finden. Im Mathematikunterricht stehen realitätsbezogene Anwendungen gleichgewichtig und gleichwertig neben innermathematischen Fragestellungen. Schülerinnen und Schüler sollen zum Ende der Qualifikationsphase Fachkompetenzen erworben haben, die es ihnen ermöglichen, sowohl die Gemeinsamkeiten als auch die Besonderheiten der Inhaltsfelder zu identifizieren und die ihnen zugrunde gelegten Konzepte flexibel zu nutzen" (vgl. [17], Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, 2014, S. 18).

Das entwickelte Lernmodul unterstützt eben diesen Vernetzungsgedanken, indem es unterschiedliche Konzepte der Inhaltsfelder aufgreift. So lässt sich das Thema sowohl der Linearen Algebra (Eigenwertproblem, Übergangsmatrizen), als auch der Stochastik (relative Häufigkeiten, Wahrscheinlichkeiten) und der Analysis (Grenzwert, stabile Grenzverteilung) zuordnen (vgl. [10], Humenberger, 2009, S. 663).

Die Schülerbefragung und auch meine eigenen Beobachtungen ergaben, dass die Ziele hinsichtlich des Zuwachses an Wissen über den Modellierungsprozess seitens der Schüler erreicht wurden. Auch wurde die Bedeutung und die Nutzbarkeit von mathematischer Modellierung für die Lösung realer Problemstellungen durch die Modellierungstage verdeutlicht.

7. Ausblick

Nachfolgend sollen mit Blick auf zukünftige Durchführungen des Lernmoduls im Zuge von Modellierungstagen weiterführende Anregungen zur Entwicklung und Optimierung des Lernmoduls festgehalten werden.

Der Schwerpunkt des entwickelten Lernmoduls zum Thema Google liegt insbesondere auf der mathematischen Modellierung des PageRank-Algorithmus. Dazu wurde der Algorithmus in seiner elementaren Form in kleineren Teilaufgaben entwickelt. Neben den im Workshop betrachteten Schritten vollführt die Suchmaschine Google jedoch noch zahlreiche weitere Arbeitsprozesse. Mit dem Ziel weiterführende Aufgaben zu entwickeln, die der Ausweitung des Modellierungstages auf mehrere Tage oder als Zusatzaufgaben für schneller arbeitende Schüler dienen könnten, ließe sich beispielsweise das Thema *Google-bombs*[38] und damit die Optimierung von Suchmaschinen vertiefter in das Modul einarbeiten.

Viele der Schritte, die die Suchmaschine Google durchführt sind weniger vom mathematischen Standpunkt, sondern insbesondere von der Seite der Informatik von Interesse. Möchte man das Thema tiefer gehend beleuchten, so würde sich ein kooperatives Projekt mit dem Schülerlabor der RWTH für Informatik[39] anbieten. Nachdem an einem Tag zunächst im Schülerlabor CAMMP die elementaren mathematischen Ideen der Suchmaschine erfasst und durchleuchtet werden, könnte man dann an einem nachfolgenden Tag im Schülerlabor *InfoSphere* die technische Umsetzung und Implementierung des Algorithmus thematisieren.

[38] *Google-bombing* meint das manipulative Hinzufügen von Links, um auf diese Weise die Erhöhung des PageRanks einer Seite zu bewirken.

[39] schuelerlabor.informatik.rwth-aachen.de, Stand: 02.08.2015

Anhang

A. Arbeitsblätter mit Lösungen

A.1. Arbeitsblatt 1

CAMMP day

Wie funktioniert eigentlich Google und was hat das mit Mathe zu tun?

Der PageRank - Modell zur Sortierung von Webseiten

Google ist eine Suchmaschine, die das WWW innerhalb von Sekundenbruchteilen durchsucht und uns die *relevanten* Seiten immer ganz oben auflistet. Sie wurde in den 90er Jahren von den beiden Wissenschaftlern Larry Page und Sergej Brin entwickelt und ist mittlerweile unangefochten die Nummer 1 unter den Suchmaschinen. Was Google so erfolgreich machte ist der PageRank-Algorithmus, der monatlich die relative Wichtigkeit einer Internetseite berechnet. Dies ist eine Zahl zwischen 0 und 1, die auch PageRank genannt wird und die bei Suchanfragen das Kriterium ist, nach dem die Suchergebnisse sortiert werden.

Problembeschreibung | Ein kleines Netzwerk

Öffne zunächst das Matlab-Skript *google.m* mit einem Doppelklick und gib in der 10. Zeile eine beliebige Internetseite ein. Starte dann das Skript durch einen Klick auf den *Run*-Button.
Es wird ein Teilnetz erstellt, dass die Verlinkung der einzelnen Seiten darstellt. In der Datei pagerank.txt werden die Seiten gemäß des berechneten PageRanks aufgelistet.

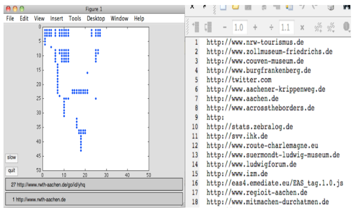


Bild 1: Vernetzungsstruktur und nach PageRank sortierte Seiten

Wir starten zunächst mit einem kleinen Netzwerk aus nur 4 Webseiten A, B, C und D (Bild 2) und versuchen deren relative Wichtigkeit, den PageRank, zu bestimmen.

Öffne das Matlab-Skript *kleines_netzwerk.m* und bearbeite die folgenden Schritte:

Schritt 1 | Bestimmung der relativen Anteile

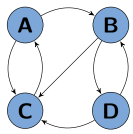


Bild 2: Kleines Netzwerk

Wir betrachten zu Beginn den Fall, dass alle User sich auf die Seiten A bis D gleichmäßig verteilen. Von der Gesamtmenge aller User befinden sich dann jeweils $\frac{1}{4}$ auf einer der Seiten A bis D. Wir schreiben:

$$(A_0, B_0, C_0, D_0) = \left(\begin{pmatrix} \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \end{pmatrix}\right)$$

Die Werte A_0 bis D_0 geben die relativen Anteile der User auf den Seiten A bis D an. Nun machen wir folgende Annahme:

1. Modellannahme: Der RandomSurfer Die User bewegen sich über Links fort. Alle Links einer Seite werden mit der gleichen Wahrscheinlichkeit benutzt.

a) Wie groß sind die *relativen Anteile der User* auf den einzelnen Seiten, wenn alle User genau einem Link gefolgt sind?
Stelle Formeln zur Berechnung von (A_1, B_1, C_1, D_1) auf und gib sie bei Matlab ein.
(Hinweis: $A_1 \widehat{=}$ relativer Anteil der User auf Seite A nach dem 1. Klick)

b) Stelle dann allgemeine Formeln für die Berechnung der relativen Anteile nach dem n-ten Klick (A_n, B_n, C_n, D_n) ausgehend davon auf, dass du $(A_{n-1}, B_{n-1}, C_{n-1}, D_{n-1})$ bereits kennst.
Diese 4 Formeln stellen zusammen ein lineares Gleichungssystem dar.

1/2

Schritt 2 | Umschreiben in Matrix-Vektor Schreibweise

Das Gleichungssystem aus Schritt 1 wird in der Mathematik oft in einer anderen Schreibweise zusammengefasst, der sogenannten Matrix-Vektor Schreibweise.

Lies die *Kleine Einführung in Matrizen*. Schreibe dann das Gleichungssystem aus Schritt 1 in Matrix-Vektor-Schreibweise auf.
Gib die Matrix, die beim Umschreiben des Gleichungssystems in Matrix-Vektor Schreibweise entsteht bei Matlab ein. Dort wird sie U genannt.

Schritt 3 | Aufstellen von Netzwerken

Nun kann man auch umgekehrt von einer Matrix ablesen, wie das zugehörige Netzwerk aussieht. Können die nachfolgenden Matrizen zu einem Netzwerk gehören? Falls ja, zeichne es auf.

$$E_1 = \begin{pmatrix} 0 & 1/3 & 0 & 1 \\ 1/2 & 0 & 0 & 0 \\ 1/2 & 1/3 & 0 & 0 \\ 0 & 1/3 & 1 & 0 \end{pmatrix}, \quad E_2 = \begin{pmatrix} 0 & 0 & 1/2 & 0 \\ 1/2 & 0 & 1 & 0 \\ 0 & 1/2 & 0 & 1 \\ 0 & 1/2 & 0 & 0 \end{pmatrix}.$$

Wenn du dir noch unsicher bist, wie du die Matrizen interpretieren musst, findest du auf *Hilfekarte 1* ein kleines Beispiel.

Schritt 4 | Bestimmung der Grenzverteilung

Wir haben eine Möglichkeit gefunden die relativen Anteile an Usern auf den einzelnen Webseiten nach einer bestimmten Anzahl von Klicks zu berechnen. Mithilfe des PageRank-Algorithmus legt Google die Wichtigkeit einer Seite über die *langfristige Verteilung* der User auf den Internetseiten fest. Die Wichtigkeit einer Webseite ist dann umso höher, je wahrscheinlicher es ist, dass ein User sie besucht. Ziel ist es also die langfristige Verteilung zu berechnen.

a) Stelle für unser kleines Netzwerk (Bild 2) eine Formel in Matrix-Vektor Schreibweise auf, mit der du $\vec{v}_4 = \begin{pmatrix} A_4 \\ B_4 \\ C_4 \\ D_4 \end{pmatrix}$ ausgehend von der Startverteilung $\vec{v}_0 = \begin{pmatrix} A_0 \\ B_0 \\ C_0 \\ D_0 \end{pmatrix}$ bestimmen kannst.

b) Bestimme nun eine Formel zur Berechnung von $\vec{v}_n$ ausgehend von der Startverteilung $\vec{v}_0$. Gib diese Formel bei Matlab ein.

c) Notiere was geschieht, wenn du für n sehr große Werte bei Matlab eingibst. Wie ändert sich der sogenannte Verteilungsvektor $\vec{v}_n$? Welches ist die Wichtigste der Seiten A, B, C und D?

Schritt 5 | Änderung der Startverteilung

Ausgehend von der Startverteilung $\vec{v}_0 = \begin{pmatrix} \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \end{pmatrix}$ hast du die Wichtigkeit der Seiten A bis D bestimmt.

a) Wie ändert sich der langfristige Verteilungsvektor $\vec{v}_n$ und somit die Wichtigkeiten der vier Seiten, wenn du die Startverteilung änderst?
(**Hinweis:** Überlege vorher welche Eigenschaften ein Startvektor erfüllen muss.)

b) Welches Ergebnis erhältst du, wenn du die Matrix U mit dem Verteilungsvektor $\vec{v}_n$ multiplizierst?

Schritt 6 | Manipulation der Wichtigkeit einer Seite

Angenommen du bist der Webmaster der Internetseite A. Du möchtest die Wichtigkeit deiner Internetseite maximal erhöhen. Dazu darfst du genau einen Link hinzufügen oder entfernen.
Ändere das Netzwerk so, dass du die maximale Wichtigkeit für deine Internetseite erreichst.

A.2. Zusatzmaterialien zum Arbeitsblatt 1

CAMMP day

Wie funktioniert eigentlich Google und was hat das mit Mathe zu tun?

Kleine Einführung in Matrizen

Matrizen und Vektoren

Eine Matrix ist ein rechteckiges Zahlenschema, das die Schreibweise von linearen Gleichungssystemen erheblich erleichtert. Hier sind zwei Beispiele einer rechteckigen und einer quadratischen Matrix:

$$A = \begin{pmatrix} -2 & 1 & 0 & 3 \\ -2 & 0 & 5 & -9 \\ 0 & 2 & -5 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & 1 \\ 0 & -2 & 4 \\ 0.5 & 0.25 & -0.4 \end{pmatrix}.$$

Ein Zahlenschema, das nur aus einer Spalte besteht, nennt man Vektor. Ein Beispiel ist der Vektor $\vec{v}$

$$\vec{v} = \begin{pmatrix} 1 \\ 4 \\ -5 \\ 1 \end{pmatrix}.$$

Rechenregeln für Matrizen

Um den PageRank-Algorithmus nachvollziehen zu können brauchen wir ein paar wichtige Rechenregeln für **quadratische Matrizen**. Quadratische Matrizen haben die gleiche Spalten- und Zeilenlänge. Die exakte Berechnung übernimmt Matlab für uns.

Addition:
Quadratische Matrizen gleicher Größe ergeben bei Addition wieder eine quadratische Matrix:

$$C + D = E$$

Multiplikation:
Quadratische Matrizen gleicher Größe können miteinander multipliziert werden:

$$C \cdot D = F \quad \text{und} \quad U^3 = U \cdot U \cdot U$$

Matrix-Vektor Multiplikation:
Quadratische Matrizen können mit Vektoren multipliziert werden, wenn die Länge des Vektors der Länge der Matrix entspricht. Das Ergebnis ist wieder ein Vektor.

$$A \cdot \vec{v} = \vec{w}$$

Lineare Gleichungssysteme

Jedes lineare Gleichungssystem kann in der folgenden Art und Weise formuliert werden:

$$A \cdot \vec{x} = \vec{b},$$

hierbei ist A eine Matrix, $\vec{x}$ ein Vektor, der die Unbekannten enthält, und $\vec{b}$ ein weiterer fester Vektor.

Beispiel
Das lineare Gleichungssystem

$$3x_1 + 5x_2 - x_3 = 2$$

$$-2x_1 + 2x_2 + 3x_3 = 0$$

$$-x_1 + 5x_2 + 2x_3 = 1$$

lässt sich schreiben als

$$A \cdot \vec{x} = \vec{b}$$

bzw.

$$\underbrace{\begin{pmatrix} 3 & 5 & -1 \\ -2 & 2 & 3 \\ -1 & 5 & 2 \end{pmatrix}}_{A} \cdot \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}}_{\vec{x}} = \underbrace{\begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}}_{\vec{b}} .$$

Infoblatt

Bestimmung der Grenzverteilung

In Schritt 4 haben wir festgestellt, dass sich die Einträge im Verteilungsvektors $\vec{v}_n$ für große Werte von n nicht mehr ändern. Man erhält also einen Grenzvektor $\vec{v}$, der auch Grenzverteilung genannt wird.

$$U^n \cdot \vec{v}_0 \xrightarrow[n\to\infty]{} \vec{v}$$

Die Grenzverteilung soll natürlich konstante Werte für die Wichtigkeiten der Seiten liefern.
In Schritt 5 haben wir festgestellt, dass der Verteilungsvektor für unser kleines Netzwerk unabhängig von dem gewählten Startvektor ist.
Aber ist dies auch bei anderen Netzwerken der Fall? Und kann man auf die gleiche Weise auch für jedes andere Netzwerk eine eindeutige Grenzverteilung berechnen?

In der Mathematik werden bei der Entwicklung eines mathematischen Modells zur Lösung eines realen Problems oft mathematische Aussagen und Sätze verwendet, die bereits bewiesen wurden. Auch wir können einen solchen Satz nutzen, um die Fragen oben zu beantworten:

Der Satz von Markov

Erfüllt die Matrix U die folgenden Eigenschaften:

- sie ist stochastisch (siehe Erklärung auf Rückseite)
- irgendeine Potenz der Matrix U (z.B. U^3 oder U^9) hat in einer Zeile nur positive Einträge

dann ist der Grenzwert

$$U^n \xrightarrow[n\to\infty]{} G$$

eine stochastische Matrix G, deren Spalten alle identisch sind.

Eine für uns wichtige Folgerung:
Multipliziert man die Grenzmatrix G mit einem beliebigen stochastischen Vektor dann ergibt sich ein konstanter Grenzvektor. Hier ist ein Beispiel mit einer kleinen quadratischen Matrix:

$$U^n \cdot \vec{v}_0 \xrightarrow[n\to\infty]{} G \cdot \vec{v}_0 = \begin{pmatrix} g_1 & g_1 & g_1 & g_1 \\ g_2 & g_2 & g_2 & g_2 \\ g_3 & g_3 & g_3 & g_3 \\ g_4 & g_4 & g_4 & g_4 \end{pmatrix} \cdot \begin{pmatrix} A_0 \\ B_0 \\ C_0 \\ D_0 \end{pmatrix} = \begin{pmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \end{pmatrix}.$$

Für Fortgeschrittene | Aufgabe 1

Überprüfe die Folgerung.
Das heißt zeige, dass sich bei Multiplikation irgendeiner Grenzmatrix G mit einem beliebigen stochastischen Vektor $\vec{v}_0$ tatsächlich immer der gleiche Grenzvektor ergibt.

Stochastische Matrizen und Vektoren

Es gibt Matrizen, die besondere Eigenschaften haben. Wichtig für uns sind sogenannte stochastische Matrizen.

Eine stochastische Matrix hat folgende Eigenschaften:

- sie ist quadratisch
- für jeden Eintrag x in der Matrix gilt: $0 \leq x \leq 1$
- die Summe der Einträge in jeder Spalte beträgt 1

Es gibt auch stochastische Vektoren mit den folgenden Eigenschaften:

- für jeden Eintrag x in dem Vektor gilt: $0 \leq x \leq 1$
- die Summe der Einträge beträgt 1

CAMMP day

Wie funktioniert eigentlich Google und was hat das mit Mathe zu tun?

Modellverbesserungen

1. Modellverbesserung

Für sehr kleine und einfache Netzwerke können wir nun eine Sortierung der Seiten nach deren **PageRank** vornehmen. Manchmal treten jedoch kompliziertere Fälle auf. . .

Schritt 1

Öffne zunächst das MATLAB-Skript *komplexeres_netzwerk.m*.
Nun betrachten wir ein etwas größeres Netzwerk aus 6 Internetseiten (Bild 1). Wie lautet die zugehörige Matrix U_1 zu diesem Netzwerk, mit der wir die Grenzverteilung bestimmen können? Stelle die Matrix auf und bestimme die Grenzverteilung, bzw. den PageRank, mit Matlab.

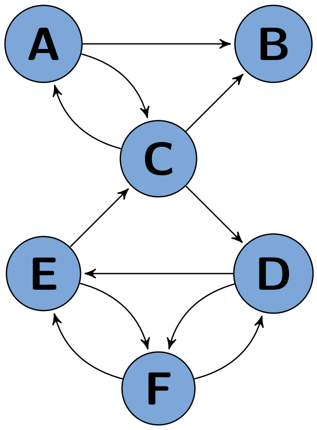


Bild 1: Komplexeres Netzwerk 1.

Schritt 2

Unser bisheriges Modell liefert kein sinnvolles Ergebnis für den PageRank der einzelnen Seiten. Woran liegt dieses Problem? Betrachte dazu das Netzwerk genauer.
Sammle Ideen, wie man dieses Problem lösen könnte und formuliere dazu eine 2. Modellannahme.

Hinweis: Eure Ideen werden wir gleich gemeinsam sammeln und dann einen Blick darauf werfen, für welche Idee Page & Brin sich damals bei der Problemlösung entschieden haben.

2. Modellannahme:
Wenn das folgende Problem auftritt...

dann...

Wie muss die Matrix U_1 geändert werden? Gib die geänderte Matrix U_1 bei Matlab ein. Erhältst du nun ein sinnvolles Ergebnis für den PageRank?

1/2

Schritt 3

Nun betrachten wird das folgende Netzwerk (Bild 2). Die Matrix zu diesem Netzwerk ist bereits in Matlab unter Schritt 3 hinterlegt.
Berechne den PageRank dieses Netzwerkes. Variiere dann mehrfach die Startverteilung und berechne den PageRank erneut. Versuche schließlich eine Erklärung für deine Beobachtungen zu finden. Betrachte dazu das Netzwerk genauer.

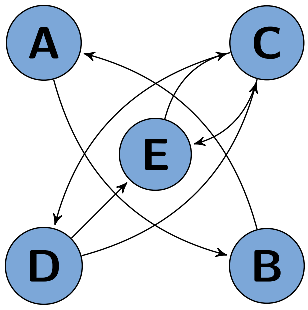


Bild 2: Komplexeres Netzwerk 2.

Die nächste und letzte Modellverbesserung löst nun auch das Problem aus Schritt 3:

2. Modellverbesserung

Mal realistisch: Klickt sich ein Internet User tatsächlich immer nur über Links durch das WWW? Natürlich ist es auch möglich, dass der User zufällig eine Internetseite eingibt und somit keinem Link folgt.

Um diese Überlegung in ihr Modell einzubauen, trafen Page & Brin damals folgende Annahme:

3. Modellannahme:
Der User folgt mit einer Wahrscheinlichkeit von $p = 0.85$ den Links und mit der Wahrscheinlichkeit $(1 - p) = 0.15$ springt er zufällig auf eine beliebige Internetseite.

Die 3. Modellannahme führt uns zu einer neuen Matrix. Diese lässt sich als Summe von 2 Matrizen zusammensetzen:
Der Matrix U_1 (siehe Schritt 2), bei der der User den Links folgt und der Matrix U_2 (siehe Schritt 4), bei der der User zufällig eine beliebige Internetseite eingibt.

Schritt 4

Stelle eine Matrix U_2 auf bei der **nie** den Links gefolgt wird, sondern stets eine beliebige der 6 Internetseiten des komplexeren Netzwerkes 1 (Bild 1) eingegeben wird. Überlege dir, wie der PageRank der einzelnen Seiten aussehen muss und überprüfe deine Vermutungen dann mit Matlab.

Schritt 5

Setze nun die endgültige Matrix T aus den beiden Matrizen U_1 (Schritt 2) und U_2 (Schritt 4) zusammen. Beachte dabei die Wahrscheinlichkeiten aus der 3. Modellannahme. Bestimme dann die Grenzverteilung.
Welche Seite hat den höchsten PageRank?

2/2

CAMMP day

Wie funktioniert eigentlich Google und was hat das mit Mathe zu tun?

Musterlösung zu den Modellschritten

Schritt 1 | Bestimmung der relativen Anteile

Die Formeln zur Berechnung der relativen Anteile nach genau einem Klick auf einen Link lauten:

$$\begin{aligned}A_1 &= 1 \cdot C_0\\ B_1 &= 0.5 \cdot A_0 + 0.5 \cdot D_0\\ C_1 &= 0.5 \cdot A_0 + 0.5 \cdot B_0 + 0.5 \cdot D_0\\ D_1 &= 0.5 \cdot B_0\end{aligned}$$

bzw. nach n Klicks:

$$\begin{aligned}A_n &= 1 \cdot C_{n-1}\\ B_n &= 0.5 \cdot A_{n-1} + 0.5 \cdot D_{n-1}\\ C_n &= 0.5 \cdot A_{n-1} + 0.5 \cdot B_{n-1} + 0.5 \cdot D_{n-1}\\ D_n &= 0.5 \cdot B_{n-1}\end{aligned}$$

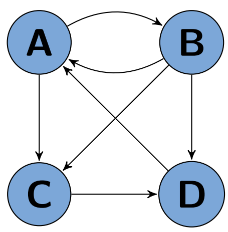
Schritt 2 | Umschreiben in Matrix-Vektor-Schreibweise

Das Gleichungssystem aus Schritt 2 sieht in Matrix-Vektor Schreibweise folgendermaßen aus:

$$\begin{pmatrix} A_n \\ B_n \\ C_n \\ D_n \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 1/2 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} A_{n-1} \\ B_{n-1} \\ C_{n-1} \\ D_{n-1} \end{pmatrix}.$$

Schritt 3 | Aufstellen von Netzwerken

Das Netzwerk zu der Übergangsmatrix E_1 sieht folgendermaßen aus:



Zu der Matrix E_2 kann kein Netzwerk gezeichnet werden, da die Spaltensumme der Spalten 1 und 3 nicht 1 ist.

Schritt 4 | Bestimmung der Grenzverteilung

a) Zunächst ergibt sich folgende Formel für $\vec{v}_4$:

$$\vec{v}_4 = U \cdot \vec{v}_3 = U \cdot U \cdot \vec{v}_2 = U \cdot U \cdot U \cdot \vec{v}_1 = U \cdot U \cdot U \cdot U \cdot \vec{v}_0 = U^4 \cdot \vec{v}_0$$

b) Für die Verteilung der relativen Anteile nach dem n-ten Klick $\vec{v}_n$ ergibt sich somit:

$$\vec{v}_n = U^n \cdot \vec{v}_0$$

1/3

c) Setzt man nun für n große Werte ein, so lässt sich beobachten, dass die Einträge der Matrix und des Verteilungsvektor $\vec{v}_n$ sich kaum noch ändern.
Ab $n = 15$ können die Schüler keine Änderungen mehr feststellen, die Matrix U und der Verteilungsvektor $\vec{v}_n$ sind konstant.
Der Verteilungsvektor lautet

$$\vec{v} = \begin{pmatrix} 3/9 \\ 2/9 \\ 3/9 \\ 1/9 \end{pmatrix}$$

Die Seiten A und C haben somit den höchsten PageRank und die höchste Wichtigkeit.

Schritt 5 | Änderung der Startverteilung

a) Der Verteilungsvektor ändert sich nicht, wenn die Schüler den Startvektor variieren. Wichtig dabei ist, dass die Schüler sich zuvor klar machen, welche Eigenschaft ein Startvektor erfüllen muss:
Er muss stochastisch sein, d. h. die relativen Anteile der User, die als Einträge im Startvektor stehen, müssen in Summe 1 ergeben.

b) Bei Multiplikation von der Matrix U mit $\vec{v}$ ergibt sich wieder der Verteilungsvektor $\vec{v}$. Es gilt also:

$$U \cdot \vec{v} = \vec{v}$$

Dies lässt sich auch recht schnell für die Schüler einsehen:
Die Schüler haben bereits erkannt, dass für n genügend groß gilt:

$$U^n \cdot \vec{v}_0 \approx \vec{v}$$

Multiplikation von U an $\vec{v}$ liefert somit:

$$U \cdot \vec{v} = U \cdot (\lim_{n\to\infty} U^n \cdot \vec{v}_0) = \lim_{n\to\infty} U^{n+1} \cdot \vec{v}_0 = \vec{v}$$

Schritt 6 | Manipulation der Wichtigkeit

Der maximale PageRank wird erreicht, wenn der Link von Seite A zu B entfernt wird. Dies führt dazu, dass sich langfristig alle User auf den Seiten A und C befinden werden.
Der Verteilungsvektor lautet dann

$$\vec{v} = \begin{pmatrix} 0.55 \\ 0 \\ 0.45 \\ 0 \end{pmatrix}$$

Andere Änderungen bewirken folgende Verteilungen:

- Ein Link von Seite B zu Seite A wird zugefügt:

$$\vec{v} = \begin{pmatrix} 0.3846 \\ 0.2308 \\ 0.3077 \\ 0.0769 \end{pmatrix}$$

- Ein Link von Seite D zu Seite A wird zugefügt:

$$\vec{v} = \begin{pmatrix} 0.3571 \\ 0.2143 \\ 0.3214 \\ 0.1071 \end{pmatrix}$$

2/3

- Ein Link von Seite A zu Seite C wird entfernt:

$$\vec{v} = \begin{pmatrix} 0.3152 \\ 0.3750 \\ 0.1875 \\ 0.1250 \end{pmatrix}$$

Musterlösung zum Infoblatt

Für Fortgeschrittene | Aufgabe 1

$$\begin{pmatrix} g_1 & g_1 & g_1 & g_1 \\ g_2 & g_2 & g_2 & g_2 \\ g_3 & g_3 & g_3 & g_3 \\ g_4 & g_4 & g_4 & g_4 \end{pmatrix} \cdot \begin{pmatrix} A_0 \\ B_0 \\ C_0 \\ D_0 \end{pmatrix} = \begin{pmatrix} g_1 \cdot A_0 + g_1 \cdot B_0 + g_1 \cdot C_0 + g_1 \cdot D_0 \\ g_2 \cdot A_0 + g_2 \cdot B_0 + g_2 \cdot C_0 + g_2 \cdot D_0 \\ g_3 \cdot A_0 + g_3 \cdot B_0 + g_3 \cdot C_0 + g_3 \cdot D_0 \\ g_4 \cdot A_0 + g_4 \cdot B_0 + g_4 \cdot C_0 + g_4 \cdot D_0 \end{pmatrix}$$

$$= \begin{pmatrix} g_1 \cdot (A_0 + B_0 + C_0 + D_0) \\ g_2 \cdot (A_0 + B_0 + C_0 + D_0) \\ g_3 \cdot (A_0 + B_0 + C_0 + D_0) \\ g_4 \cdot (A_0 + B_0 + C_0 + D_0) \end{pmatrix} = \begin{pmatrix} g_1 \cdot 1 \\ g_2 \cdot 1 \\ g_3 \cdot 1 \\ g_4 \cdot 1 \end{pmatrix} = \begin{pmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \end{pmatrix}$$

CAMMP day

Wie funktioniert eigentlich Google und was hat das mit Mathe zu tun?

Musterlösung zu den Modellverbesserungen

1. Modellverbesserung | Schritt 1

Die Übergangsmatrix zu diesem etwas komplizierteren Netzwerk sieht folgendermaßen aus:

$$U_1 = \begin{pmatrix} 0 & 0 & 1/3 & 0 & 0 & 0 \\ 1/2 & 0 & 1/3 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1/2 & 0 \end{pmatrix}$$

Der Grenzvektor lautet:

$$\vec{v} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Berechnen die Schüler die langfristige Verteilung (ab $n > 8000$ über $\vec{v} = U_1^n \cdot \vec{v}_0$) erhalten sie den Nullvektor. Dies ist kein sinnvolles Ergebnis für den PageRank-Vektor.

1. Modellverbesserung | Schritt 2

Die Matrix erfüllt die Bedingungen für den Satz von Markov nicht. Die zweite Spalte ist eine Nullspalte. Sie hat somit nicht Spaltensumme 1.
Das Netzwerk betrachtend liegt dieses Problem darin, dass von Seite B keine Links ausgehen. Navigiert der User auf diese Seite, so landet er also in einer Sackgasse. Diese wird auch Senke genannt. Setzen die Schüler große Werte für n ein, so erhalten sie einen Grenzvektor $\vec{v}$, der nicht stochastisch ist (Nullvektor). Dies ist kein sinnvolles Ergebnis für den Verteilungsvektor, bzw. den PageRank.

Mögliche Ideen um dieses Problem zu lösen könnten sein:

- Alle User die auf Seite B gelangen, beenden den Suchvorgang und bleiben bei Seite B. Dies beschreibt das Verhalten eines Users nicht sehr realistisch.
- Der User geht mit dem Browser eine Seite zurück und benutzt von dort aus einen anderen Link. Dies würde kompliziert werden, da unterschieden werden müsste, von welcher Seite der User auf Seite B gelangt ist.
- Die Erfinder Brin & Page hatten folgende Idee, um dieses Problem zu lösen:

2. Modellannahme:
Wenn das folgende Problem auftritt: Der User navigiert in eine Sackgasse,
dann gibt er mit der gleichen Wahrscheinlichkeit irgendeine der Internetseiten ein.

Hinweis: Die Tatsache, dass die Wahrscheinlichkeit geringer ist, dass der User erneut Seite B besuchen wird, wird bei dieser Annahme nicht beachtet. Bei einem großen Netzwerk macht dies jedoch auch keinen nennenswerten Unterschied.
Die geänderte Matrix U_1 sieht dann folgendermaßen aus:

$$U_1 = \begin{pmatrix} 0 & 1/6 & 1/3 & 0 & 0 & 0 \\ 1/2 & 1/6 & 1/3 & 0 & 0 & 0 \\ 1/2 & 1/6 & 0 & 0 & 0 & 0 \\ 0 & 1/6 & 1/3 & 0 & 0 & 1/2 \\ 0 & 1/6 & 0 & 1/2 & 0 & 1/2 \\ 0 & 1/6 & 0 & 1/2 & 1 & 0 \end{pmatrix}$$

Der PageRank lautet nun:

$$\vec{v} = \begin{pmatrix} 0.0755 \\ 0.1132 \\ 0.1698 \\ 0.1887 \\ 0.2264 \\ 0.2264 \end{pmatrix}$$

2. Modellverbesserung | Schritt 3

Geben die Schüler unterschiedliche Startvektoren ein, so erhalten sie verschiedene Ergebnisse für den PageRank-Vektor. Dieser ist somit nicht eindeutig. Dies liegt daran, dass das gegebene Netzwerk nicht zusammenhängend ist. Es gibt ein kleines Netzwerk der Seiten A und B und ein zweites Netzwerk aus den Seiten C, D und E.

Beispiele:
Mit dem Startvektor

$$\vec{v}_0 = \begin{pmatrix} 1/5 \\ 1/5 \\ 1/5 \\ 1/5 \\ 1/5 \end{pmatrix}$$

ergibt sich folgender Verteilungsvektor:

$$\vec{v} = \begin{pmatrix} 0.2 \\ 0.2 \\ 0.2667 \\ 0.1333 \\ 0.2 \end{pmatrix}$$

Mit dem Startvektor

$$\vec{v}_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

ergibt sich folgender Verteilungsvektor:

$$\vec{v} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Mit dem Startvektor

$$\vec{v}_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$

ergibt sich folgender Verteilungsvektor:

$$\vec{v} = \begin{pmatrix} 0 \\ 0 \\ 0.4444 \\ 0.2222 \\ 0.3333 \end{pmatrix}$$

Mit dem Startvektor

$$\vec{v}_0 = \begin{pmatrix} 0 \\ 1/2 \\ 0 \\ 0 \\ 1/2 \end{pmatrix}$$

ergibt sich folgender Verteilungsvektor:

$$\vec{v} = \begin{pmatrix} 0 \\ 0.5 \\ 0.2222 \\ 0.1111 \\ 0.1667 \end{pmatrix}$$

2. Modellverbesserung | Schritt 4

Es ist für alle Seiten gleich wahrscheinlich, dass der User sie besucht. Die Matrix U_2 lautet folgendermaßen:

$$U_2 = \begin{pmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{pmatrix}$$

Der PageRank-Vektor, der sich dann ergibt lautet:

$$\vec{v} = \begin{pmatrix} 1/6 \\ 1/6 \\ 1/6 \\ 1/6 \\ 1/6 \\ 1/6 \end{pmatrix}$$

2. Modellverbesserung | Schritt 5

Schließlich müssen die Schüler die Matrix T nur noch aus den beiden Matrizen aus Schritt 2 & 4 zusammensetzen und diese mit den jeweiligen Wahrscheinlichkeiten verknüpfen. Dabei ergibt sich notwendigerweise eine stochastische Matrix, die den Satz von Markov erfüllt (siehe Infoblatt).

$$T = p \cdot U_1 + (1 - p) \cdot U_2$$

Der berechnete PageRank-Vektor lautete dann:

$$\vec{v} = \begin{pmatrix} 0.0926 \\ 0.1319 \\ 0.1725 \\ 0.1820 \\ 0.2105 \\ 0.2105 \end{pmatrix}$$

4/4

B. Hilfekarten

CAMMP day

Wie funktioniert eigentlich Google und was hat das mit Mathe zu tun?

Hilfekarten

Schritt 1 | Hilfekarte 1

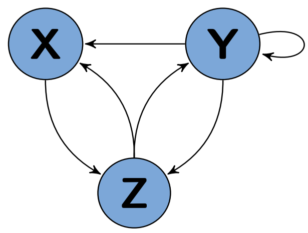


Bild 1: Webseiten X, Y und Z

Ein kleines Beispiel:
Wir betrachten ein Netzwerk aus drei Seiten X, Y und Z. Die Gesamtmenge aller User verteilt sich am Anfang **gleichmäßig** auf die 3 Webseiten. Die Anfangsverteilung der User lautet dann:

$$X_0 = 1/3$$
$$Y_0 = 1/3$$
$$Z_0 = 1/3$$

D.h. jeweils $1/3$ aller User befindet sich auf einer der Seiten X, Y und Z.
Die Pfeile (Links) zeigen an, zu welchen Seiten die User nach einem Klick auf einen Link wechseln. Dabei gilt, dass die User allen Links mit der gleichen Wahrscheinlichkeit folgen.
Wie groß ist der relative Anteil an Usern, die sich nach dem ersten Webseitenwechsel auf Seite X befinden?

Lösung: Von Y gehen 3 Links aus. Einer davon führt zu X. Genau $1/3$ von dem Anteil der User auf Seite Y wandert somit zu Seite X.
Von Z gehen nur 2 Links aus. Genau $1/2$ von dem Anteil der User auf Seite Z wandert somit zu Seite X.
Der Anteil der User auf Seite X nach dem ersten Webseitenwechsel setzt sich dann so zusammen:

$$X_1 = 1/3 \cdot 1/3 + 1/2 \cdot 1/3 = 5/18$$

Genau $5/18$ aller User befindet sich nach dem ersten Wechsel auf Seite X. Die Formel dazu lautet:

$$X_1 = 1/3 \cdot Y_0 + 1/2 \cdot Z_0$$

Schritt 3 | Hilfekarte 1

Die Matrix G gehört zu einem kleinen Netzwerk, das aus 4 Seiten A, B, C und D besteht.

$$G = \begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 1/2 \\ 1/2 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{array}.$$

Die rot markierte 4. Spalte der Matrix lässt sich so interpretieren:
Die User von Seite D gehen zur Hälfte auf Seite A und zur anderen Hälfte auf Seite B über.

Hilfekarten

Schritt 3 | Hilfekarte 1

Das Netzwerk aus Schritt 3 kannst du auch folgendermaßen aufzeichnen:

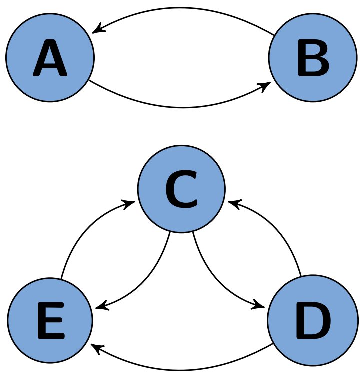
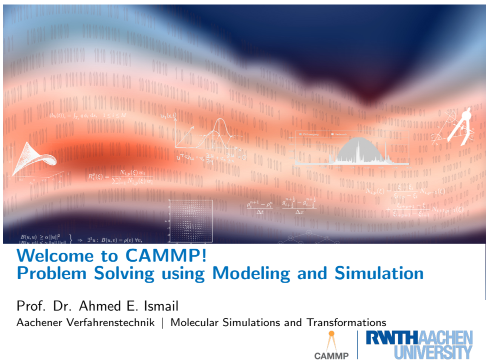
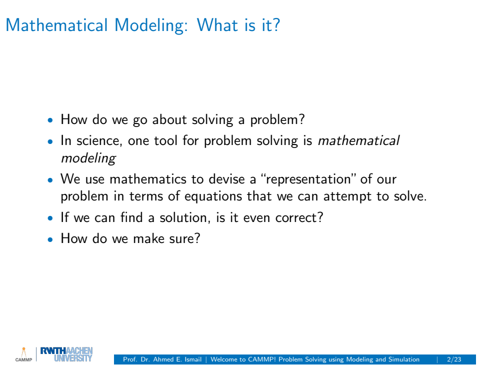


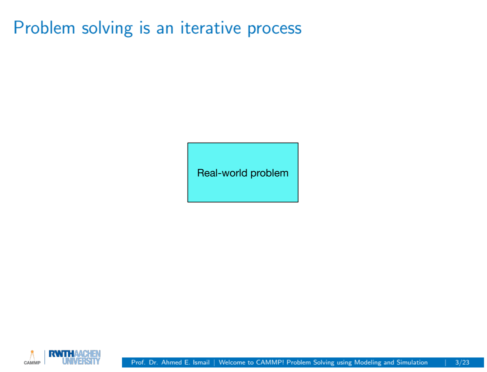
Bild 1: komplexeres Netzwerk

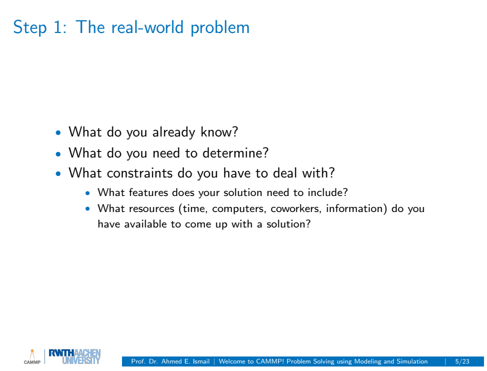
C. Präsentationen

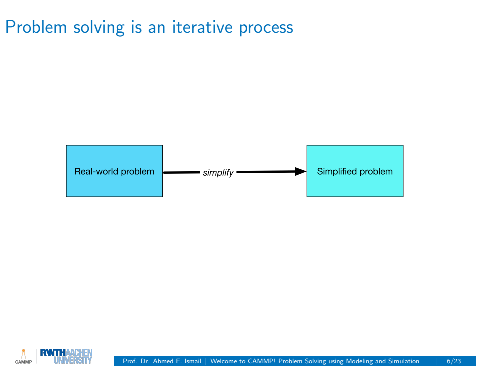
C.1. Modellierungsvortrag







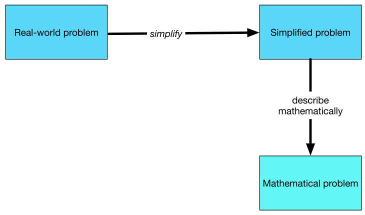




Step 2: Simplify the problem

- The full problem might be too hard: **Simplify!**
  - **Assumptions** introduce **new information**
    - Needed when you don't have enough data to enable a solution
    - *Example*: The car is traveling at constant velocity and does not accelerate.
  - **Simplifications** make the math simpler
    - We treat the image of the car as a rectangle or trapezoid instead of finding the complete outline.
- Einstein: "Models should be as simple as possible, **but no simpler!**"

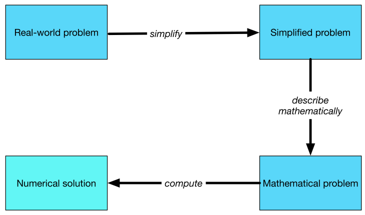
Problem solving is an iterative process



Step 3: Describe your model mathematically

- Convert your real-world system into a set of mathematical equations to be solved
- How do you *quantify* the problem?
- Don't forget to take **all** of your known data, assumptions, simplifications, and constraints into account!
- After you've built the system, you'll need to figure out how to solve it!

Problem solving is an iterative process

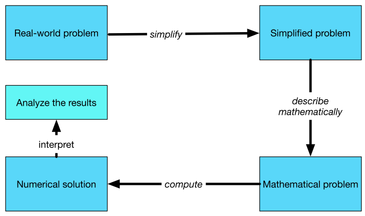


Step 4: Solve your model

Analytical versus numeric solutions

- For a very small number of problems, an exact analytical solution is available
- These special cases usually involve very "simple" problems
  - Symmetric geometries
  - No "special conditions" (no reactions, no heat loss, no time-dependence, etc.)
- When an analytical solution is unavailable, a numerical solution is required
- For CAMMP, you'll be using **Matlab** to sole your problems

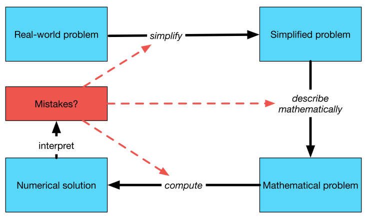
Problem solving is an iterative process



Step 5: Analyze your results

- Check the solution
  - Is the solution consistent with the given data and assumptions?
  - Does the solution satisfy all your constraints and requirements?
- Example: If the car is moving at 0.01 km/h or 1500 km/h, something's probably wrong!
- Can you test your method with a simple "test" problem before solving the bigger problem?

Problem solving is an iterative process



Step 6: Refine your model, if needed

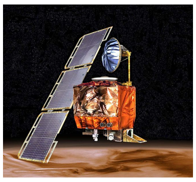
- If you didn't find a valid solution the first time, or are not satisfied with the solution, try again
  - Did you make an invalid assumption or simplification of the problem?
  - Did you make a mistake converting your problem into mathematics?
  - Did you solve the problem incorrectly?
  - Is there a bug in your code?
- **Don't be afraid of making a mistake!**

Who uses modeling?

- Every major branch of science and engineering now uses mathematical modeling
- Many businesses and industries use it, too
- One cool example: astrophysics!

Why testing and checking are important

Mars Surveyor '98

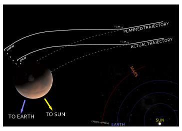


NASA

- Mars Climate Orbiter was supposed to measure the atmosphere and surface features of Mars
- Reported as "lost in transit"
- NASA personnel reported probe was at too low an altitude relative to the planet

Why testing and checking are important

Mars Surveyor '98

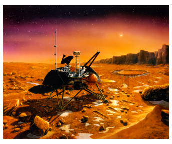


NASA

- Problem traced to an error in units handling
- Probe probably disintegrated due to stresses created flying through atmosphere instead of in exoplanet orbit

Why testing and checking are important

Mars Surveyor '98

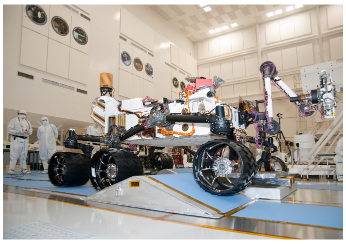


NASA

- Mars Polar Lander was supposed to explore the polar geology
- Never made contact after "landing"
- Most likely cause: vibrations due to deployment of landing gear was interpreted by software as "touchdown"
- Polar lander fell to surface from 40 m instead of 12 m
- Polar lander became "Polar smasher"

Getting it right

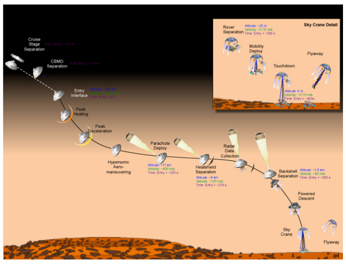
Mars Curiosity



Source: JPL

Getting it right

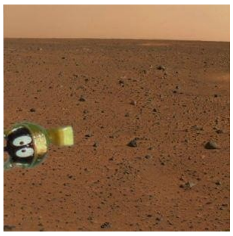
Mars Curiosity



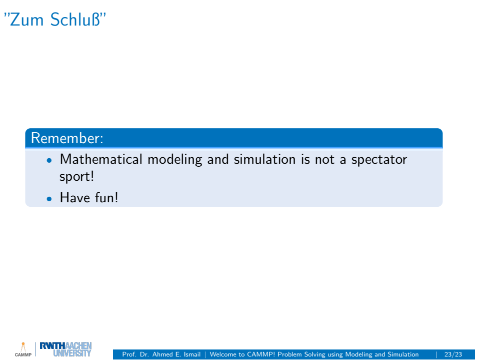
Source: NASA

Getting it right

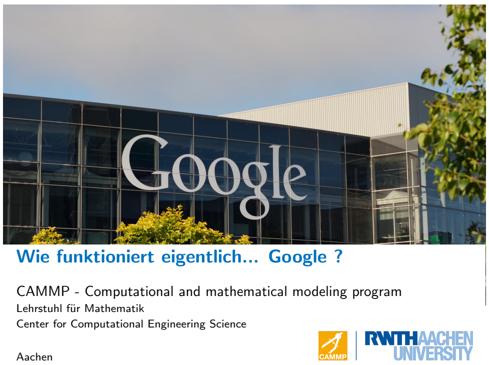
Mars Curiosity

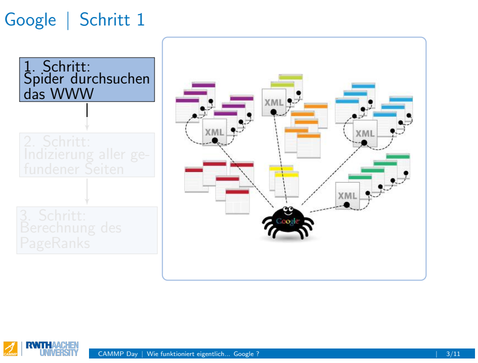


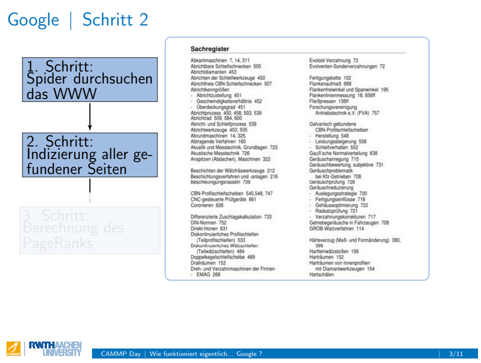
Source: NASA

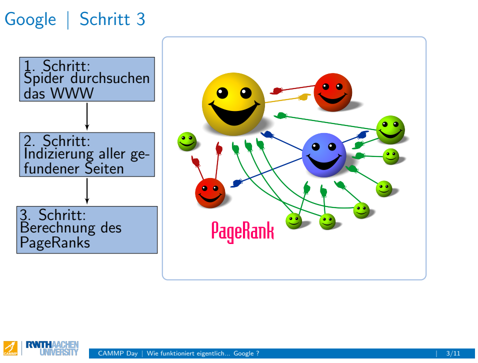


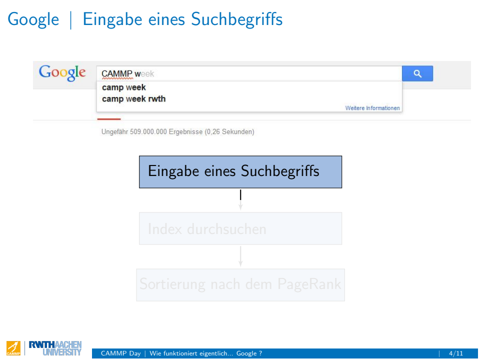
C.2. Problemeinführungsvortrag



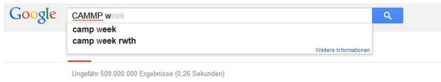






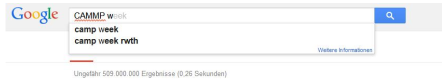


Google | Index durchsuchen



Eingabe eines Suchbegriffs
↓
Index durchsuchen
↓
Sortierung nach dem PageRank

Google | Sortierung nach dem PageRank



Eingabe eines Suchbegriffs
↓
Index durchsuchen
↓
Sortierung nach dem PageRank

Die Wichtigkeit einer Webseite | Der PageRank

Deine Webseite
↓
Nummer 1 im Ranking

Die Wichtigkeit einer Webseite | Fragen

Deine Webseite
↓
Nummer 1 im Ranking

Wie legt Google die Wichtigkeit einer Seite fest?

Wie können wir die Wichtigkeit mit einem mathematischen Modell beschreiben?

Die Wichtigkeit einer Webseite | Diskussion

Eure Ideen:
Wann haltet ihr eine Webseite für wichtig? Wann sollte sie bei Suchanfragen weit oben erscheinen?

Die Wichtigkeit einer Webseite | Diskussion

Eure Ideen:
Wann haltet ihr eine Webseite für wichtig? Wann sollte sie bei Suchanfragen weit oben erscheinen?

Page & Brins Idee:
Eine Webseite ist wichtig, wenn es wahrscheinlich ist, dass viele User diese besuchen.

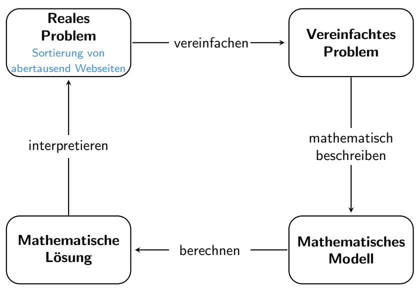
Die Wichtigkeit einer Webseite | Diskussion

Eure Ideen:
Wann haltet ihr eine Webseite für wichtig? Wann sollte sie bei Suchanfragen weit oben erscheinen?

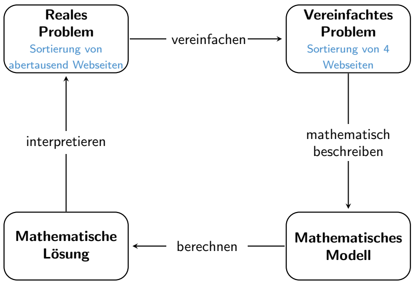
Page & Brins Idee:
Eine Webseite ist wichtig, wenn es wahrscheinlich ist, dass viele User diese besuchen.

Ziel:
Bestimmen wie sich die User auf alle Internetseiten verteilen.

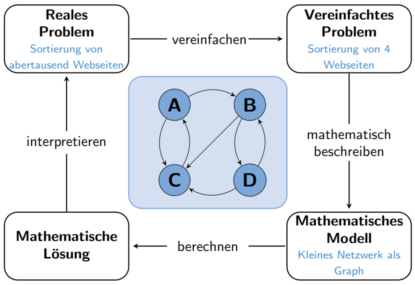
Modellierungskreislauf



Modellierungskreislauf

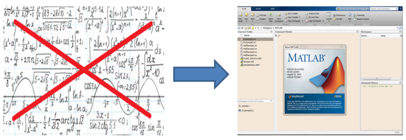


Modellierungskreislauf



Bearbeitung

Partnerarbeit an einem Laptop.



Präsentation

Präsentation der Ergebnisse vor der Mittagspause.

C.3. Hinweise zum Problemeinführungsvortrag

CAMMP day

Wie funktioniert eigentlich Google und was hat das mit Mathe zu tun?

Hinweise zum Eröffnungsvortrag

Folie 1 | Eröffnungsfolie

Nach dem einführenden Vortrag in die Welt der mathematischen Modellierung können wir nun auch schon mit dem Workshop zum Thema Google beginnen.

Folie 2 | Der Erfolg der Suchmaschine Google

- Google ist die erfolgreichste Suchmaschine weltweit
- Der Prototyp der Suchmaschine Google wurde in den 90er Jahren von Larry Page und Sergej Brin an der Stanford University entwickelt. Keines der großen Internetportale (z.B. Yahoo) interessierte sich damals dafür (...ein großer Fehler). Deswegen gründeten Page & Brin 1998 die eigene Firma *Google Inc.*
- Der Erfolg der Suchmaschine liegt im PageRank-Algorithmus. Dieser machte Google schneller als die Konkurrenz. Außerdem wurden bei Google die *besten* Internetseite immer als erstes angezeigt.
- Mittlerweile hat es das Verb *googlen* sogar schon in den Duden geschafft.
- Die Erfinder Page & Brin gehören laut der Forbes Liste (Stand: Mai 2015) zu den 20 reichsten Menschen weltweit.
- Das originale Paper zum PageRank, welches die Erfinder im Zuge einer Seminararbeit verfassten, ist tatsächlich online zu finden.
- Das Wort *Google* stammt von dem Begriff *Googol* ab. Dieser hatte sich in der ersten Hälfte des 20. Jahrhunderts in Amerika für die riesige Zahl 10^{100} etabliert.

Folie 3 | Google - 3 grundlegende Schritte

Hinweise:
Schritt 1 & 2 laufen gleichzeitig ab!

1. Schritt:
Google durchsucht das Internet mit Spidern, auch Webcrawler genannt. Dies sind Programme, welche nach und nach alle öffentlichen Internetseiten ermittelt. Dabei setzt Google eine gewaltige Anzahl an Computern ein um täglich Milliarden von Webseiten durchsuchen zu können. Bei jeder neuen Durchsuchung des Internets starten die Spider mit einer Liste von Webseiten, die sie aus der letzten Durchsuchung schon kennen. Bei dem Besuch dieser bekannten Webseiten erkennen die Spider alle Links dieser Seite. Diese Links werden einer Liste noch zu durchsuchender Webseiten hinzugefügt. Diese werden dann nach und nach besucht und so arbeiten sich die Spider durch das ganze Internet. Am Ende hat Google also eine umfassende Momentaufnahme des gesamten Internets mit all seinen Webseiten und deren Verlinkungsstruktur erstellt.

2. Schritt:
Alle von den Spidern gefundenen Webseiten werden dann in einer Index Datenbank gespeichert. Diesen Vorgang nennt man Indizierung. Dieser Index ist ganz ähnlich aufgebaut wie der Index (das Sachregister) hinten in einem Buch. Zu jedem Stichwort bzw. Suchbegriff steht dort genau aufgelistet auf welchen Seiten des Buches dieser Begriff auftaucht. Google speichert in seinem Index also

1/2

Informationen über einzelne Begriffe und deren *Position* auf den Webseiten im Internet.

3. Schritt: Nun wird die Wichtigkeit bzw. die Relevanz der gefundenen Internetseiten bestimmt. Das geschieht über den PageRank-Algorithmus. Dieser berechnet für jede Seite eine Zahl zwischen 0 und 1, die dann als Maß für die Wichtigkeit dient.
Ein PageRank nahe bei 0 bedeutet, dass die Seite unwichtig ist, ein PageRank nahe bei 1 heißt, dass die Seite relativ wichtig ist.
Die Berechnung des PageRanks beruht im wesentlichen der Verlinkungsstruktur unter den einzelnen Webseiten, wie wir nachher sehen werden.

Folie 4 | Eingabe eines Suchbegriffs

Wenn wir nun einen Suchbegriff bei Google eingeben, dann sucht Google in seinem *Index* nach allen Internetseiten, die den Begriff enthalten.
Die Sortierung der gefundenen Seiten erfolgt dann nach deren in Schritt 3 berechneten Wichtigkeit (PageRank).

Folie 5 | Die Wichtigkeit einer Webseite - der PageRank

Stellt euch vor, euer Sportverein, euer Schulprojekt, eure Jugendgruppe,... o.ä. hat eine eigene Webseite und ihr möchtet natürlich, dass diese Seite bei Google stets ganz oben aufgelistet wird. Eure Internetseite soll als einen hohe Wichtigkeit und somit einen hohen PageRank haben.
Dazu müssen wir verstehen wie Google die Wichtigkeit einer Internetseite festlegt.
Mit welchem mathematischen Modell berechnet Google also die Wichtigkeit einer Internetseite?

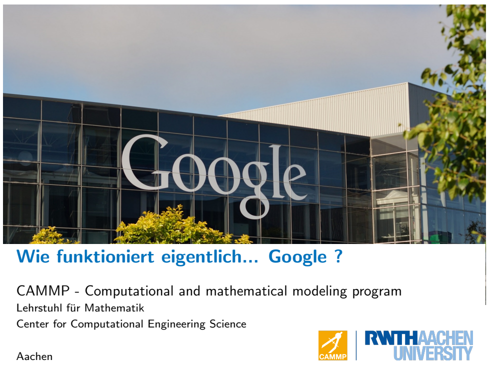
Folie 6 | Die Wichtigkeit einer Webseite: Diskussion

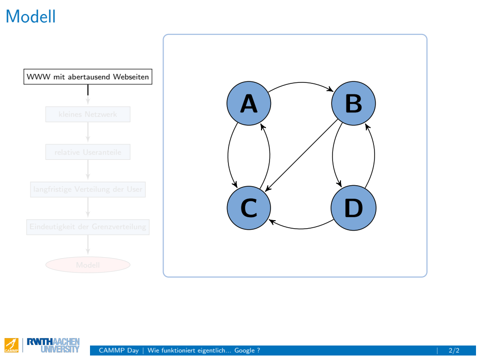
Eure Ideen: Dazu können wir zunächst einmal gemeinsam überlegen, wann wir selbst eine Webseite für wichtig halten und sie deswegen ganz oben in einem Ranking erscheinen sollte. Was meint ihr wann ist eine Seite wichtiger als eine andere?
Page & Brins Idee: Page & Brin haben sich damals überlegt, dass eine Seite dann besonders wichtig ist, wenn es wahrscheinlich ist, dass viele User sie besuchen werden.
Ziel: Das Ziel ist es nun also, ein mathematisches Modell zu entwickeln mit dem wir bestimmen können, wie sich die User beim Surfen durch das Internet auf die einzelnen Webseiten verteilen.

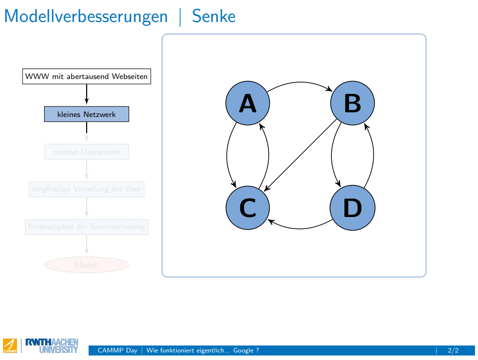
Folie 7, 8 & 9 | Der Modellierungskreislauf: Reales Problem

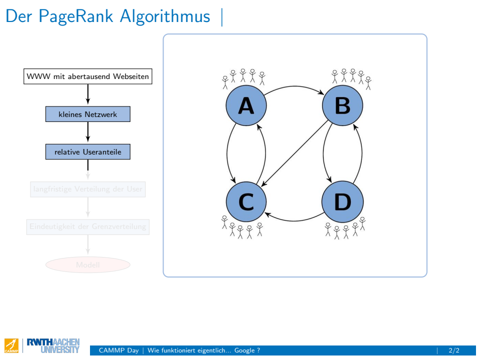
Im Modellierungsvortrag haben wir eben gehört, dass es meist sinnvoll oder notwendig ist die reale Situation erstmal zu vereinfachen.
Bei der Erarbeitung eines mathematischen Modells für die Berechnung der Wichtigkeit starten wir nicht direkt mit dem Internet, einem Netzwerk aus abertausend Webseiten, sondern mit einer vereinfachten Situation.
Wir versuchen zunächst ein mathematisches Modell zur Berechnung des PageRanks am Beispiel von einem kleinen Netzwerk zu entwickeln und überprüfen dann, ob es auch für komplizierte Netzwerke oder gar das ganze Internet sinnvoll ist.
Wir überlegen uns dazu, wie wir das kleine Netzwerk erstmal als reales Modell darstellen können.
Das machen wir indem wir das Internet als ein Netzwerk auffassen, bei dem die Seiten über Links, im Modell als Pfeile dargestellt, verbunden werden.
(Mathematischer ausgedrückt: Das Internet wird als Graph dargestellt. Ein Graph ist eine Menge aus Ecken und Kanten. Ecken sind in unserem Bespiele die Kreise, die für die Internetseiten stehen und die Kanten, die alle eine vorgegebene Richtung haben, sind die Pfeile (Links).

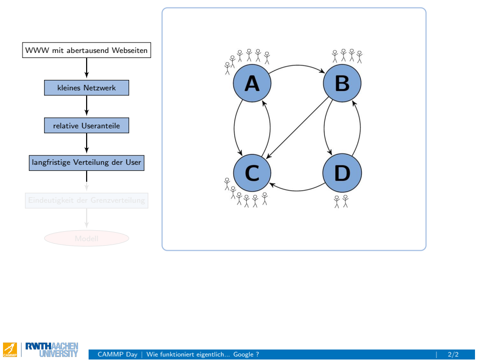
C.4. Vortrag zur Sicherung des ersten Modells

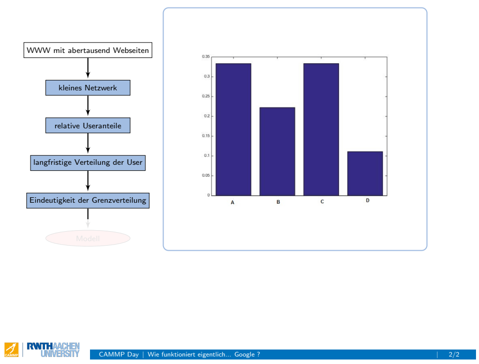


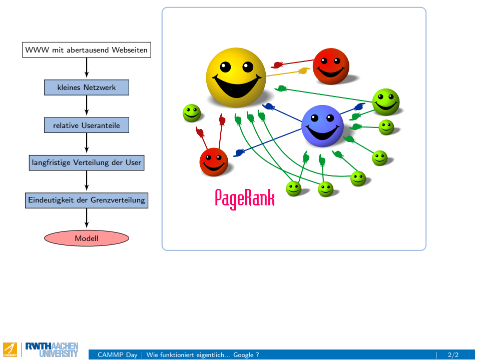












D. Basic Paper

CAMMP day

Wie funktioniert eigentlich... Google und was hat das mit Mathe zu tun?

CAMMP

RWTH AACHEN UNIVERSITY

Wie funktioniert eigentlich. . .
Google?

Lektion 1 Einleitung

Google ist eine Suchmaschine, die das WWW innerhalb von Sekundenbruchteilen durchsucht und dem Internetuser schließlich die *wichtigsten* Seiten zu seiner Suchanfrage ganz oben auflistet. Wie Google die Wichtigkeit einer Internetseite festlegt und welche elementaren mathematischen Ideen hinter der Berechnung dieser Wichtigkeit stecken wird in der nachfolgenden Lektion erläutert.

Aufbau des Lernmoduls

Zunächst wird anhand eines kleinen Beispielnetzwerkes ein erstes Modell zur Berechnung der Wichtigkeit von Internetseiten entwickelt. Dazu arbeiten die Schüler mit dem *Arbeitsblatt 1* und dem MATLAB-Skript *kleines_netzwerk.m*. Das entwickelte Modell wird dann im zweiten Teil des Lernmoduls an einem etwas größeren Netzwerk überprüft. Dabei wird sich herausstellen, dass das bisherige Modell noch nicht für alle beliebigen Netzwerke ein sinnvolles Ergebnis für die Sortierung der Internetseiten liefert, sodass Modellverbesserungen eingebaut werden müssen. Diesen Teil des Workshops bearbeiten die Schüler anhand des *Arbeitsblatts 2* unter Verwendung des MATLAB-Skripts *komplexes_netzwerk.m*. Der genau didaktisch-methodische Einsatz der Materialien im Rahmen des CAMMP days wird im *methodischen Konzept* detailliert beschrieben.

Ziel des Lernmoduls

Anhand der schrittweisen Erarbeitung des PageRank-Algorithmus sollen die Schüler den gesamten Modellierungskreislauf durchlaufen. Die einzelnen Modellierungsschritte sollen dabei weitestgehend eigenständig von den Schülern absolviert werden.

Benötigte Software

Das Lernmodul wurde im MATLAB erstellt, welches eine hochentwickelte Programmiersprache und interaktive Umgebung, die unter anderem für numerische Berechnungen und Datenanalyse bzw. -visualisierung und insbesondere zum disziplinübergreifenden arbeiten genutzt wird, darstellt. Steht die kostenpflichtige Software MATLAB nicht zur Verfügung so kann auch auf die freie Software GNU Octave zurückgegriffen werden, deren Skriptsprache von der Skriptsprache MATLABs abgeleitet wurde.

Lektion 2 Die Suchmaschine Google

Der Erfolg der Suchmaschine Google

Der Prototyp der Suchmaschine Google wurde Mitte der 90er Jahre von den beiden Informatikstudenten Lawrence Page und Sergej Brin an der Stanford University entwickelt. Nachdem Brin und Page im Jahr 1998 das Unternehmen Google Inc. gegründet hatten ging die Suchmaschine das erste mal online. Mit einem Marktanteil von über 70 %[1] ist Google mittlerweile die unangefochtene Nummer eins unter den Suchmaschinen und die Erfinder gehören laut Forbes-Liste[2] unterdessen zu den zwanzig reichsten Menschen weltweit. Google hob sich damals insbesondere dadurch von den anderen Suchmaschinen ab, dass die Ergebnisse bei Suchanfragen zum einen wesentlich schneller beantwortet wurden als bei konkurrierenden Suchmaschinen und zum anderen die für den Nutzer scheinbar interessantesten Internetseiten stets ganz oben im Ranking aufgelistet wurden.
Was Google so erfolgreich machte, ist der auf der Verlinkungsstruktur des Internets basierende PageRank-Algorithmus. Dieser berechnet *monatlich* die relative Wichtigkeit einer Internetseite. Diese Wichtigkeit wird durch den PageRank ausgedrückt, einer Zahl zwischen 0 und 1 nach der die Internetseiten bei Suchanfragen schließlich sortiert werden. Dabei gilt: Umso höher der PageRank, umso wichtiger die Seite.
Im nachfolgenden Abschnitt werden die grundlegenden Schritte Googles erläutert. Der Schwerpunkt wird dabei auf den mathematischen Hintergründen des PageRank - Algorithmus liegen.

Die grundlegenden Schritte der Suchmaschine Google

Google führt drei grundlegende Schritte aus. Im ersten Schritt durchsuchen sogenannten Webcrawler, auch Spider genannt, das Internet. Dies sind Softwareprogramme, die das Internet durchforsten und dabei alle öffentlich zugänglichen Internetseiten erfassen. Das Crawling beginnt dabei mit einer Liste von Internetadressen, die aus vorherigen Durchsuchungen des Internets bereits bekannt sind. Zunächst besuchen die Spider alle diese Internetseiten und folgen dann sämtlichen Links auf die diese Seiten verweisen, um dann wiederum den Links auf diesen Seiten zu folgen. So arbeiten sich die Spider durch das Internet. Im zweiten Schritt indiziert Google die Daten aus Schritt 1, um eine effiziente Durchsuchung nach relevanten Suchbegriffen, zusammen mit deren Position im Internet zu ermöglichen. Der dabei entstehenden Index ist ganz ähnlich aufgebaut wie der Index eines Buches, in dem zu jedem Stichwort aufgelistet steht auf welchen Seiten des Buches dieser Begriff auftaucht. Google speichert in seinem Index somit Informationen über einzelne Suchbegriffe sowie deren Position im Internet[3]. An dieser Stelle ist anzumerken, dass die Schritte 1 und 2 nicht nacheinander sondern bereits parallel zueinander ablaufen.

Im dritten Schritt wird schließlich mithilfe des PageRank-Algorithmus die Wichtigkeit einer jeden Internetseite berechnet. Nach dieser können die Seiten bei Suchanfragen dann sortiert werden.
Ziel des nachfolgenden Abschnitts ist es die grundlegenden mathematischen Ideen hinter dem PageRank-Algorithmus zu verdeutlichen.

An dieser Stelle soll jedoch festgehalten werden, dass dieser Algorithmus in der Praxis noch mit wesentlich weitreichenderen Nebenbedingungen verbunden und insbesondere technisch von weitaus höherer Komplexität ist als im Folgenden dargestellt. Die nachfolgende Darlegung des PageRank-Algorithmus ist strukturell stark an dem Aufbau des Lernmoduls orientiert.

[1] *www.netmarketshare.com*, Stand: 01.07.2015
[2] *www.forbes.com/forbes-400/list*, Stand: 26.06.2015
[3] www.google.com/insidesearch/howsearchworks/crawling-indexing.html, Stand: 02.07.2015

Lektion 3 Der PageRank-Algorithmus

Das Random Surfer Modell

Der PageRank-Algorithmus modelliert das Verhalten eines idealen, zufällig durch das Internet navigierenden Users. Dieser *Random Surfer* besucht zu Beginn irgendeine öffentlich zugängliche Internetseite. Dann folgt er zufällig einem beliebigen ausgehenden Link auf eine andere Seite. Von dieser Seite navigiert er dann wiederum über einen ausgehenden Link auf eine weitere Webseite. Auf diese Weise bewegt der Random Surfer sich durch das Internet fort. Von Zeit zu Zeit folgt er jedoch keinem Link, sondern springt zufällig auf eine beliebige Internetseite, indem er beispielsweise in der Adresszeile eine Internetadresse eingibt. Der Surfvorgang des Random Surfers ist nicht terminiert, d.h. er wird nicht müde immer und immer weiter durch das Internet zu surfen.

Diese Betrachtung des Internetnutzers als Random Surfer legt folgende Definition der relativen Wichtigkeit einer Internetseite nahe:

Definition der Wichtigkeit: Eine Internetseite ist wichtig, wenn es wahrscheinlich ist, dass der Random Surfer sie besucht.

Das Modell zu Berechnung des PageRanks einer Seite, dem das Verhalten des Random Surfers zugrunde liegt, soll in den nachfolgenden Abschnitten schrittweise erarbeitet werden. Dabei wird zunächst ein *erstes Modell* zu Berechnung des PageRank entwickelt, welches wie sich herausstellen wird, nicht für alle beliebigen Netzwerke anwendbar ist, sodass wichtige *Modellverbesserungen* eingebaut werden. Bei der Entwicklung des PageRank-Algorithmus werden die Modellierungsschritte und auftretenden Probleme stets an kleinen Beispielnetzwerken veranschaulicht. Dabei handelt es sich um eben die Netzwerke mit denen auch die Schüler im Rahmen des CAMMP days arbeiten.

Das Internet als gerichteter Graph

Zur Entwicklung eines ersten mathematischen Modells, das die Struktur des Internets möglichst genau darstellt, wird das Internet zunächst als gerichteter Graph dargestellt. Ein gerichteter Graph besteht aus Ecken (den Internetseiten) und Kanten (den Links) die alle eine Richtung aufweisen.
Angenommen das betrachtete Internet bestehe aus N Webseiten, die alle bekannt sind. Diese werden nun bezeichnet mit

$$G_i, i = 1, ...N$$

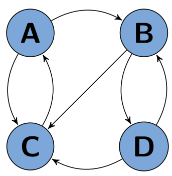


Abbildung 1: Kleines Beispielnetzwerk

Ein kleines Beispielnetzwerk aus 4 Internetseiten ist in Abbildung 1 dargestellt. Gemäß der obigen Bezeichnungen gilt für dieses:

$$G_1 = A,\ G_2 = B,\ G_3 = C,\ G_4 = D$$

Ein Pfeil von Seite C zu Seite D steht für einen Link von Seite C zu D usw.
Nun kann man folgende Modellannahme treffen:

1. Modellannahme: Die Internetnutzer bewegen sich über Links fort und alle Links einer Internetseite werden mit der gleichen Wahrscheinlichkeit genutzt.

Der gerichtete Graph lässt sich dann als Übergangsgraph darstellen, bei dem jede Kante ein Kantengewicht besitzt. Die Kantengewichte stellen die Übergangswahrscheinlichkeiten dar mit denen ein Internetnutzer von einer Internetseite zu einer anderen wechselt. Berechnen lassen sich diese indem man die Anzahl der ausgehenden Links einer Internetseite betrachtet und zudem die erste Modellannahme berücksichtigt. Bei einer Internetseite mit m ausgehenden Links betragen die Übergangswahrscheinlichkeiten von jedem dieser ausgehenden Links folglich $\frac{1}{m}$.

Nun wird die Gesamtmenge aller Internetnutzer und deren relative Anteile auf den einzelnen Seiten betrachtet. Wir bezeichnen die relativen Anteile zu Beginn mit $(G_1^0, G_2^0, ..., G_N^0)$ bzw. das Beispielnetzwerk betrachtend mit (A_0, B_0, C_0, D_0). Die relativen Anteile der Internetnutzer auf den jeweiligen Seiten nach dem n-ten Klick auf einen Link werden dann durch $(G_1^n, G_2^n, ..., G_N^n)$ bzw. (A_n, B_n, C_n, D_n) beschrieben. Mit diesen Bezeichnungen lassen sich rekursive Formeln aus dem Übergangsgraphen ablesen, welche die Verteilung der Internetnutzer auf die Internetseiten angeben.
Für das obige kleine Beispielnetzwerk findet man folgende Formeln, die die Verteilung der Internetnutzer nach n Zeitschritten, d.h. nach n Klicks auf einen Link, beschreiben:

$$\begin{array}{rllll} A_n = & & & 1\ C_{n-1} & \\ B_n = & 0.5\ A_{n-1} & & & +0.5\ D_{n-1} \\ C_n = & 0.5\ A_{n-1} & +0.5\ B_{n-1} & & +0.5\ D_{n-1} \\ D_n = & & 0.5\ B_{n-1} & & \end{array}$$

Das entstehende lineare Gleichungssystem lässt sich in Matrix-Vektor Schreibweise wie folgt darstellen:

$$\underbrace{\begin{pmatrix} A_n \\ B_n \\ C_n \\ D_n \end{pmatrix}}_{=:v_n} = \underbrace{\begin{pmatrix} 0 & 0 & 1 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 1/2 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 & 0 \end{pmatrix}}_{=:U} \cdot \underbrace{\begin{pmatrix} A_{n-1} \\ B_{n-1} \\ C_{n-1} \\ D_{n-1} \end{pmatrix}}_{=:v_{n-1}}.$$

bzw. kurz:

$$v_n = U \cdot v_{n-1}.$$

Allgemeiner lässt sich die Matrix U zum Übergangsgraphen aus seiner Adjazenzmatrix A entwickeln. Die Adjazenzmatrix A ist eine N x N-Matrix mit den Einträgen 0 und 1, wobei N die Anzahl der Ecken (Internetseiten) angibt. Es gilt:
Gibt es eine Verbindung von Ecke j zu Ecke i, dann ist der Eintrag (i, j) der Matrix A gleich 1, sonst 0. Kurz:

$$A_{ij} = \begin{cases} 1, & \text{es existiert eine Kante von } j \text{ nach } i \\ 0, & \text{sonst.} \end{cases}$$

Die Matrix U zu einem beliebigen Übergangsgraphen ergibt sich dann, indem die Einträge der Adjazenzmatrix durch die Einträge des Spaltensummenvektors geteilt werden. Die Einträge des Spaltensummenvektors lauten:

$$s_j = \sum_{i=1}^{N} A_{ij}.$$

Die Matrix U kann dann wie folgt definiert werden:

$$U_{ij} = \begin{cases} \frac{A_{ij}}{s_j}, & s_j \neq 0 \\ A_{ij}, & s_j = 0. \end{cases}$$

Die Einträge der Matrix U lassen sich folgendermaßen interpretieren:
Der Eintrag U_{ij} gibt die Wahrscheinlichkeit an mit der ein Internetnutzer, der sich gerade auf Seite G_j befindet im nachfolgenden Schritt über einen Link auf Seite G_i navigiert.

Betrachtet man die Übergänge der Reihe nach, so lässt sich mithilfe der Matrixschreibweise eine einfache Beschreibung der Verteilung nach dem n-ten Zeitschritt, d.h. nach dem n-ten Klick aller Internetnutzer auf einen Link, finden:

$$v_1 = U \cdot v_0, \quad v_2 = U \cdot v_1 = U \cdot (U \cdot v_0) = U^2 \cdot v_0 \,, \ldots, \quad v_n = U^n \cdot v_0$$

Um nun die Wichtigkeit einer Internetseite festlegen zu können, betrachtet man die langfristige Verteilung der Internetnutzer auf die einzelnen Seiten.
Zu betrachten ist somit der folgende Grenzwert:

$$\lim_{n\to\infty} v_n = \lim_{n\to\infty} U^n \cdot v_0.$$

Um einen eindeutigen PageRank für jede Internetseite zu erhalten, sollte dieser Grenzwert zu jeder beliebigen Startverteilung bzw. zu jedem stochastischen Vektor[4] eine eindeutige Grenzverteilung, d.h. einen eindeutigen stochastischen Vektor v liefern, dessen Einträge schließlich die Wichtigkeiten der einzelnen Internetseiten widerspiegeln. Wann tatsächliche zu jeder beliebigen Startverteilung eine eindeutige Grenzverteilung existiert, liefert der folgende Satz:

Satz von Markov

Sei $A \in \mathbb{R}^{nxn}$ eine stochastische Matrix[5]. Wenn irgendeine Matrixpotenz A^n für $n \geq 1$ nur positive Einträge hat, dann existiert die Grenzmatrix

$$A_\infty = \lim_{n\to\infty} A^n$$

mit identischen Spalten.
Schnell lässt sich dann einsehen, dass die Grenzverteilung v mit

$$v = A_\infty \cdot v_0$$

dann unabhängig von einem beliebigen stochastischen Startvektor v_0 durch diese Spalte gegeben ist. Es gilt also

$$A_\infty = \begin{pmatrix} a_1 & a_1 & . & . & . & a_1 \\ a_2 & a_2 & . & . & . & a_2 \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ a_n & a_n & . & . & . & a_n \end{pmatrix} \text{ und } v = \begin{pmatrix} a_1 \\ a_2 \\ . \\ . \\ . \\ a_n \end{pmatrix}.$$

Im Falle des kleinen Beispielnetzwerkes (vgl. Abb. 1) hat bereits die 5-te Potenz der Übergangsmatrix U nur positive Einträge. Für dieses Netzwerk existiert somit nach dem Satz von Markov unabhängig vom Startvektor die eindeutige Grenzverteilung v:

$$v = \left(\frac{3}{9}, \frac{2}{9}, \frac{3}{9}, \frac{1}{9}\right)^T.$$

Die Grenzverteilung v stellt selbst stets einen stochastischen Vektor dar, für den insbesondere gilt:

$$A \cdot v = A \cdot (A_\infty \cdot v_0) = A_\infty \cdot v_0 = v.$$

Die Grenzverteilung v stellt also gleichermaßen die eindeutig bestimmte stationäre Verteilung zum Eigenwert 1 von A dar und die Berechnung des PageRanks lässt sich demgemäß als Eigenwertproblem auffassen.

Es lässt sich festhalten, dass das bisherige Modell für das oben gewählte Beispielnetzwerk (vgl. Abb. 1) eine eindeutige Grenzverteilung und somit einen eindeutigen PageRank liefert. Dies ist jedoch nicht für alle beliebigen Netzwerke der Fall, wie das nachfolgende Beispiel verdeutlicht.

[4] Ein stochastischer Vektor hat reelle nicht-negative Einträge die in Summe 1 ergeben

[5] Eine quadratische Matrix heißt spalten-/zeilenstochastisch, wenn alle Einträge zwischen 0 und 1 liegen und die Spalten-/Zeilensummen 1 betragen

Modellverbesserung 1: Senke

Betrachtet man das in Abbildung 2 dargestellte Netzwerk, so liefert die Berechnung des PageRanks mit dem bisherige Algorithmus über die zu dem Netzwerk gehörende Matrix U_1

$$U_1 = \begin{pmatrix} 0 & 0 & 1/3 & 0 & 0 & 0 \\ 1/2 & 0 & 1/3 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1/2 & 0 \end{pmatrix}$$

als Ergebnis für die Grenzverteilung den Nullvektor. Dies ist weder ein sinnvolles, noch ein brauchbares Ergebnis für den PageRank-Vektor.
Das Problem besteht darin, dass von Seite B keine Links ausgehen. Das Netzwerk enthält somit eine Senke, die auch Sackgasse genannt wird. Dies äußert sich in der Matrix durch eine Nullspalte. Der Satz von Markov ist nicht erfüllt und der Algorithmus liefert keinen stochastischen Grenzvektor.

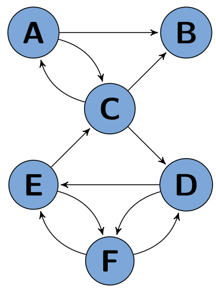


Abbildung 2: Beispielnetzwerk zur Modellverbesserung Senke

Um dieses Problem zu lösen wird folgende Modellannahme getroffen:

2. Modellannahme:
Navigiert der User beim Surfen durch das Internet in eine Senke, so kehrt er zu der Liste aller Internetseiten zurück und klickt nun eine beliebige der möglichen N Internetseiten an und zwar jede Seite mit der selben Wahrscheinlichkeit.

Dies führt dazu, dass die Nullspalte nun durch die Spalte

$$(\frac{1}{N}, \frac{1}{N}, ., ., ., \frac{1}{N})^T$$

ersetzt wird. Für das 2. Beispielnetzwerk ergibt sich damit die stochastische Übergangsmatrix U_1

$$U_1 = \begin{pmatrix} 0 & 1/6 & 1/3 & 0 & 0 & 0 \\ 1/2 & 1/6 & 1/3 & 0 & 0 & 0 \\ 1/2 & 1/6 & 0 & 0 & 0 & 0 \\ 0 & 1/6 & 1/3 & 0 & 0 & 1/2 \\ 0 & 1/6 & 0 & 1/2 & 0 & 1/2 \\ 0 & 1/6 & 0 & 1/2 & 1 & 0 \end{pmatrix}.$$

Erneutes Berechnen der Grenzverteilung liefert nun ein sinnvolles Ergebnis:

$$v = (0.0755, 0.1132, 0.1698, 0.1887, 0.2264, 0.2264)^T.$$

Die Modellannahme lässt außer Acht, dass es unwahrscheinlicher ist wieder auf die selbe Seite zu klicken. Dies fällt jedoch bei der Berechnung des PageRanks mit der zum realen Internet gehörende Matrix, die aus mehreren Billionen Zeilen besteht, kaum ins Gewicht.

Ein weiteres Problem, das auftreten kann und zu dem der Algorithmus in der bisherigen Version kein brauchbares Ergebnis liefert, wird am Beispiel des in Abbildung 3 dargestellten Netzwerkes deutlich.

Modellverbesserung 2: Unzusammenhängende Netzwerke

Das Ausführen des bisherigen Algorithmus mit der zu dem Netzwerk gehörenden Übergangsmatrix

$$U_2 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 1 \\ 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

liefert für unterschiedliche Startvektoren unterschiedliche Grenzverteilungen. So ergibt sich beispielsweise mit dem Startvektor

$$v_0 = (1/5, 1/5, 1/5, 1/5, 1/5)^T$$

die Grenzverteilung

$$v = (0.2, 0.2, 0.2667, 0.1333, 0.2)^T$$

mit dem Startvektor

$$v_0 = (1, 0, 0, 0, 0)^T$$

jedoch folgende Grenzverteilung:

$$v = (1, 0, 0, 0, 0)^T.$$

Dies hat zur Folge, dass den einzelnen Internetseiten kein eindeutiger PageRank zugeordnet werden kann.

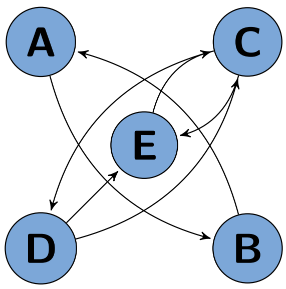


Abbildung 3: Beispielnetzwerk zu der Modellverbesserung Unzusammenhängende Netzwerke

Das an dieser Stelle auftretende Problem ist, dass der Graph zu diesem Netzwerk nicht zusammenhängend ist[6]. Dieses Problem wird durch die folgenden Modellannahme aufgegriffen:

3. Modellannahme: Der User folgt mit einer Wahrscheinlichkeit von $0 < p < 1$ den Links und mit der Wahrscheinlichkeit $(1 - p)$ springt er zufällig auf eine beliebige Internetseite.

Diese Modellannahme lässt sich folgendermaßen mathematisch beschreiben:
Wird den Links auf den Internetseiten gefolgt, so ist die Übergangsmatrix gemäß des bis hierhin entwickelten Algorithmus, d.h. gemäß der Übergangsmatrix U_1 gegeben. Betrachtet man nur das zufällige Springen des Internetnutzers auf beliebige Internetseiten, z.B. indem er oben in der Adresszeile eine Internetadresse eingibt, so lässt sich dies durch eine weitere Übergangsmatrix U_3 beschreiben.

[6] Genau genommen tritt neben dem Problem des unzusammenhängenden Netzwerkes zudem das Problem eines kreisförmigen (Teil-)Netzes zwischen den Seiten A und B auf. Dieses wird jedoch ebenfalls durch die nachfolgende Modellannahme aufgegriffen, weshalb darauf nicht näher eingegangen wird.

Jeder Eintrag dieser Matrix beträgt $\frac{1}{N}$, wobei N der Gesamtzahl aller Internetseiten, bzw. allgemeiner der Anzahl der Ecken des gerichteten Graphen, entspricht. Verknüpft man diese beiden Matrizen unter Berücksichtigung der angegebenen Wahrscheinlichkeiten, so ergibt sich eine neue Übergangsmatrix T gemäß:

$$T = p \cdot U_1 + (1 - p) \cdot U_3$$

Da p zwischen 0 und 1 liegt und zudem sowohl die Matrix U_1, als auch die Matrix U_3 stochastisch sind, ist auch T stochastisch. Des Weiteren hat die Matrix T nur positive Einträge. Der Satz von Markov ist somit erfüllt und der Algorithmus liefert stets eine von der Startverteilung unabhängige stochastische Grenzverteilung.
Die Erfinder Page und Brin verwendeten bei der Berechnung des PageRanks anfänglich den Wert $p = 0.85$. Mit welcher Wahrscheinlichkeit die Suchmaschine jedoch heutzutage arbeitet, ist ein gut gehütetes Geheimnis des Unternehmens Google Inc.

Lektion 4 Berechnung des PageRanks in der Praxis

Iterative Berechnung des PageRanks in der Praxis

Sowohl die Berechnung des PageRanks über das Lösen des Eigenwertproblems

$$v = T \cdot v$$

als auch die Berechnung über die Bestimmung riesiger Matrixpotenzen gemäß

$$\lim_{n\to\infty} T^n$$

ist für eine gigantische Übergangsmatrix mit mehr als 25 Billionen von Zeilen und Spalten, wie es im Falle des realen Internets zutrifft, ungeeignet.
In der Praxis wird die Grenzverteilung deswegen iterativ angenähert. Dazu startet der Algorithmus mit einer beliebigen Startverteilung v_0 und berechnet dann schrittweise

$$v_{r+1} = T \cdot v_r$$

solange bis der Fehler

$$\|v_{r+1} - v_r\|$$

genügend klein geworden ist.

Lektion 5 Weiterführende Literatur

Nachfolgend ist eine kleine Zusammenstellung der dem Lernmodul zugrundeliegender sowie weiterführender zu finden. Darunter befinden sich sowohl wissenschaftliche Zeitschriftenartikel zu den mathematischen Hintergründen, als auch zu der didaktisch-methodisches Umsetzung.

- The PageRank Citation Ranking: Dies ist das original Paper von Sergey Brin und Lawrence Page aus dem Jahr 1999, in dem sie erstmals Google und seine Funktionsweise, damals noch als wissenschaftliches Projekt, öffentlich präsentierten.
- The $25,000,000,000 eigenvector: the linear algebra behind Google: In diesem Paper kann man noch einmal ganz detailliert den Beweis dafür, dass das Iterationsverfahren immer eine eindeutige Lösung hat, nachlesen. Weiterhin findet ihr hier auch weitergehende Aufgaben und die Autoren bieten auf ihrer Webseite Material für die Programme Maple und Mathematica an.
- The anatomy of a large-scale hypertextual Web search engine: In diesem Paper beschreiben die beiden Google Gründer die technische Umsetzung ihrer Suchmaschine. Hier werden vor allem Details behandelt, die über den PageRank-Algorithmus hinausgehen.
- Google-bombing - Manipulating the PageRank Algorithm: Gibt einen allgemeinen Überblick über Möglichkeiten den PageRank zu manipulieren und welche Gegenmaßnahmen getroffen werden können, um solche Manipulationen zu verhindern.
- Inside PageRank: Dies ist ein sehr ausfürlicher Übersichtsartikel über den PageRank-Algorithmus, in dem auch weiterführende Probleme und Details sehr ausführlich behandelt werden. Zum Beispiel kann man hier eine genaue stochastische Analyse des Random Surfers nachlesen.
- Deeper Inside PageRank: Dies ist nocheinmal eine Erweiterung des obigen Papers. Hier werden auch technische Details wie die effiziente Speicherung der Google-Matrix, Beschleunigungsverfahren für die Berechnung des PageRank-Vektors und geschickte Verfahren für das Update des PageRank behandelt. Ebenfalls nicht leicht zu lesen, man sieht hier aber gut, wieviele verschiedene Fragestellungen mit diesem Problem verbunden sind.
- Efficient crawling through URL ordering: Auch wenn der PageRank-Algorithmus das Kernstück von Google ist, gibt es noch andere wichtige Komponenten. Zum Beispiel muss der Google-Index mithilfe eines sogenannten Webcrawlers regelmäßig aktualisiert werden. Wie das funktioniert, kann man hier nachlesen. Auch dies ist ein Paper aus der Zeit der Google Gründung.
- Der PageRank-Algorithmus von Google - eine aktuelle Anwendung im Mathematikunterricht. Von dem Didaktiker Hans Humenberger aus den Beiträgen zum Mathematikunterricht 2009.

E. Methodisches Konzept

CAMMP day

Wie funktioniert eigentlich Google und was hat das mit Mathe zu tun?

Methodisches Konzept zur Durchführung eines CAMMP days

CAMMP day regulär

Dauer: 0,5 h Einführung + 2,5 h Bearbeitungsdauer + 0,5 h Nachbesprechung + Pausen

Methodisches Konzept: Selbstständige Erarbeitung des Modells anhand des ersten Arbeitsblattes. Einbau von Modellverbesserungen anhand des zweiten Arbeitsblattes. Partnerarbeit in eigenem Tempo (Heterogenität). Sicherung durch Hilfekarten und Diskussion der Ergebnisse.

Ziel: Am Ende des CAMMP days, sollten alle SuS das Modell des Random Surfers erfasst haben. Sie sollten verstanden haben, wie sie die Wichtigkeit einer Internetseite zumindest theoretisch erhöhen könnten.

Ablauf:

1. Begrüßung, Einführungsvortrag CAMMP day allgemein (*presentation course*), *Modellierungvortrag*, Unterschriftenliste.

2. Problemeinführungsvortrag (*google presentation open*) halten. Bei der Vorbereitung kann sich an den Hinweisen zum Vortrag orientiert werden (*google presentation open notes*). Die SuS erhalten einen groben Eindruck von der Geschichte und dem Erfolg der Suchmaschine Google. Die grundlegenden Schritte der Suchmaschine werden erläutert. Datei *google.m* kurz erklären. MATLAB Einführung (*einfuehrung_matlab.m*).

3. Bearbeitung der folgenden sechs Modellierungsschritte
   Schritt 1: Bestimmung der relativen Anteile
   Schritt 2: Umschreiben in Matrix-Vektor-Schreibweise
   Schritt 3: Aufstellen von Netzwerken
   Schritt 4: Bestimmung der Grenzverteilung
   Schritt 5: Unabhängigkeit von der Startverteilung
   Schritt 6: Manipulation der Wichtigkeit

   Verwendung des Arbeitsblattes *google lesson1 problems* für die SuS bzw. *google lesson1 solutions* für die Betreuer. Als Einführung in Matrizen zu Schritt 2 den Zettel *google introduction matrix* verwenden. Als Hilfestellung zu Schritt 1 und Schritt 3 können die SuS die Hilfekarten *google lesson1 help* verwenden. Zur Bearbeitung verwenden die SuS die MATLAB Dateien *google.m, surfer.m* und *kleinesnetzwerk.m*, sowie die Datei *pagerank.txt*. Zum automatischen Lösungsabgleich wird zudem die Datei *ueberpruefe_Loesung.p* benötigt. Das Infoblatt *google information markov* erhalten schnellere Schüler nachdem sie das 1. Arbeitsblatt vollständig bearbeitet haben.

4. Kurz vor der Mittagspause: Sicherung der Ergebnisse. Einzelne SuS stellen die Modellierungsschritte an der Tafel vor. Dann wird die anfängliche Fragestellung aufgegriffen und diskutiert, wie die Wichtigkeit einer eigenen Seite erhöht werden kann. Es wird zur Zusammenfassung der Vortrag *presentation model* gehalten.

5. Mittagspause

6. Die SuS arbeiten mit einem komplizierteren Netzwerk und bauen Modellverbesserungen ein. Verwendung des Arbeitsblatt *google lesson2 problems* bzw. *google lesson2 solutions* für die Betreuer. Es werden folgende 5 Schritte bearbeitet:

   Schritt 1 & 2: Erarbeitung der Modellverbesserung *Senke*.
   In Schritt 2 sammeln die SuS Ideen zur Lösung des Problems *Senke*. Die Ideen der Schüler werden in kurzer frontaler Phase gesammelt. Der Betreuer stellt dann die Modellannahme der Google-Erfinder vor.
   Schritt 3: Erfassung des Problems *unzusammenhängende Netzwerke*.
   Schritt 4 & 5: Erarbeitung des finalen Modells.

   Zur Bearbeitung verwenden die SuS die MATLAB Datei *komplexesnetzwerk.m*. Als Hilfestellung zu Schritt 3 können die SuS die Hilfekarte *google lesson2 help* verwenden.

7. Kurz vor Ende des CAMMP days findet eine erneute Sicherung statt. SuS stellen das endgültige Modell vor. Abschlussvortrag CAMMP day allgemeine (*presentation close*).

8. Evaluation

F. Schülerbefragung

F.1. Evaluationsbogen

CAMMP Day

Wie funktioniert eigentlich ...,
und was hat das mit Mathe zu tun?

CAMMP | RWTH AACHEN UNIVERSITY

Evaluation

Rhein-Maas Gymnasium – 29. Mai 2015

Es besteht immer die Möglichkeit unsere Programme zu verbessern und wir würden gerne deine Meinung erfahren. Vielen Dank für deine Rückmeldung.

1. Persönliche Angaben:

Jahrgangsstufe: _____ Geschlecht: ☐ weiblich ☐ männlich

Leistungskurse: ____

2. Bewertung des Workshops

Brücken ☐ Google ☐ GPS ☐ JPEG ☐ mp3 ☐ Shazam ☐ Solarkraftwerk ☐

	Trifft gar nicht zu (--)	Trifft eher nicht zu (-)	Trifft zum Teil zu (+)	Trifft voll zu (++)
Durch den Workshop habe ich mathematisches Modellieren besser begriffen.	◯	◯	◯	◯
Der Vortrag über Modellierung war hilfreich.	◯	◯	◯	◯
Der einführende Kurzfilm war hilfreich.	◯	◯	◯	◯
Die Einführung in MATLAB war hilfreich.	◯	◯	◯	◯
Der Umgang mit MATLAB fiel mir schwer.	◯	◯	◯	◯
Die Aufgaben waren zu einfach.	◯	◯	◯	◯
Die Aufgaben waren zu schwierig.	◯	◯	◯	◯
Die Hilfekarten waren hilfreich.	◯	◯	◯	◯
Das Hilfesystem in MATLAB war hilfreich.	◯	◯	◯	◯

Kommentar:

Was würdest du am Workshop verändern bzw. verbessern wollen?

____

____

____

____

____

Bitte wenden...

1/2

Was hast du für dich persönlich durch die Teilnahme am Workshop gelernt?

3. Weiterführende Fragen:

	Trifft gar nicht zu (--)	Trifft eher nicht zu (-)	Trifft zum Teil zu (+)	Trifft voll zu (++)
1. Der Workshop hat mein Interesse an den Naturwissenschaften und Technik gesteigert.	○	○	○	○
2. Die Anleitungen zu den Aufgaben waren zufriedenstellend und gut verständlich.	○	○	○	○
3. Die Dozenten haben die Inhalte plausibel und klar dargestellt.	○	○	○	○
4. Die Betreuer haben das selbstständige Arbeiten gefördert.	○	○	○	○
5. Es wurden interessante Berufe und Studiengänge vorgestellt.	○	○	○	○
6. Ich kann mir gut vorstellen, später ein Studium oder eine Ausbildung im Bereich Technik oder Naturwissenschaften aufzunehmen.	○	○	○	○
7. Ich würde einen solchen Workshop gerne noch einmal besuchen.	○	○	○	○

Kommentar:

Abschließender persönlicher Kommentar (z.B. Lob, Kritik, Verbesserungsvorschläge):

2/2

CAMMP day

Wie funktioniert eigentlich Google, und was hat das mit Mathe zu tun?

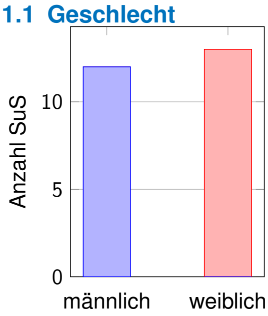
Evaluation

CAMMP day Gymnasium Herzogenrath

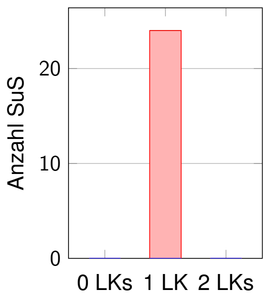
18 Juni 2015

1 Persönliche Angaben

1.1 Geschlecht

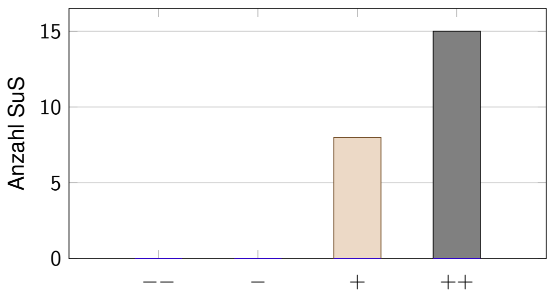


1.2 Anzahl LKs in Mathe, Info und Physik



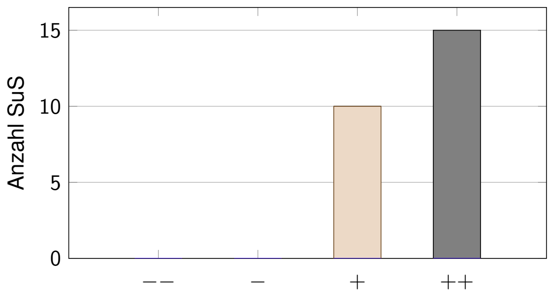
2 Bewertung des Workshops

2.1 Durch den Workshop habe ich mathematisches Modellieren besser begriffen.

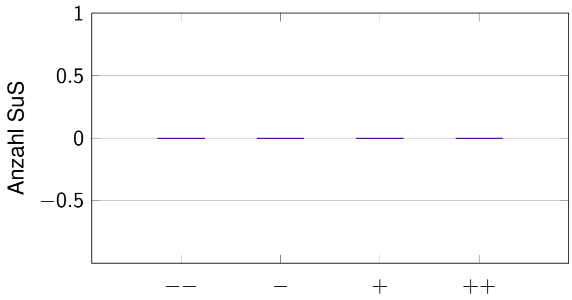


1/8

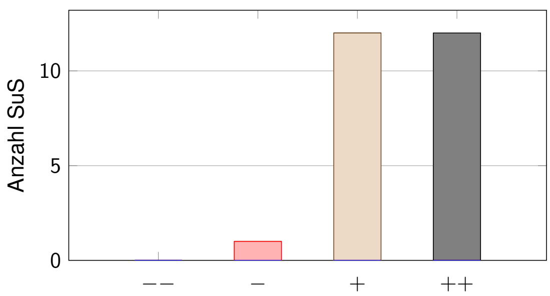
2.2 Der Vortrag über Modellierung war hilfreich.



2.3 Der einführende Kurzfilm war hilfreich.

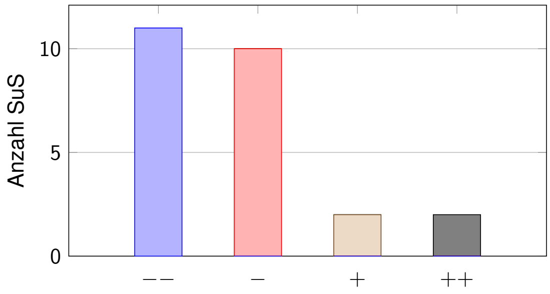


2.4 Die Einführung in MATLAB war hilfreich.

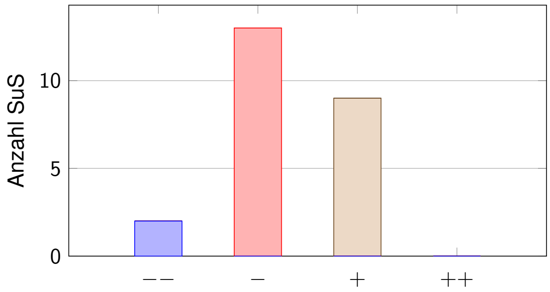


2/8

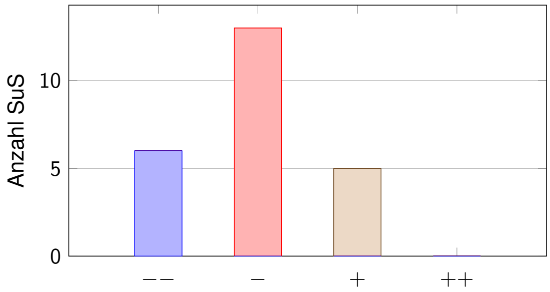
2.5 Der Umgang mit MATLAB fiel mir schwer.



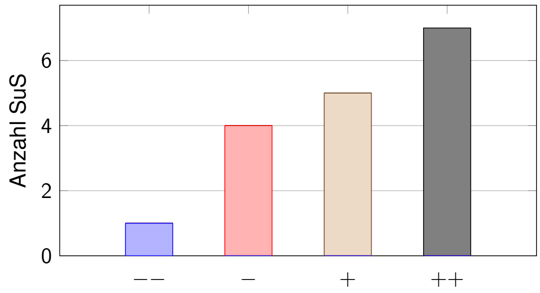
2.6 Die Aufgaben waren zu einfach.



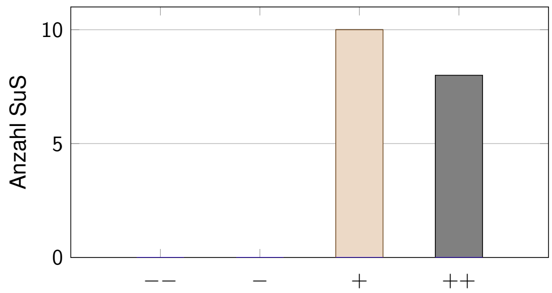
2.7 Die Aufgaben waren zu schwierig.



2.8 Die Hilfekarten waren hilfreich.



2.9 Das Hilfesystem in MATLAB war hilfreich.



2.10 Was würdest du am Workshop verändern bzw. verbessern wollen?

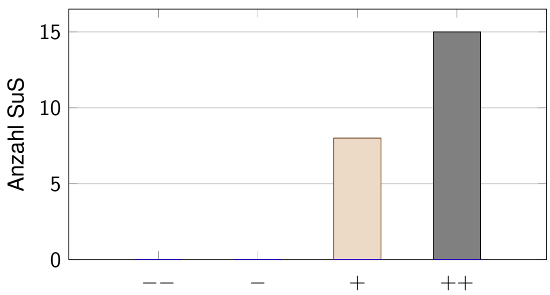
S.1	Gut.
S.2	Ziemlich interessant, wäre hilfreich gewesen, sich etwas länger mit MATLAB auseinanderzusetzen (vor Bearbeitung der Aufgaben).
S.3	Mehr offene Arbeit, weniger Aufgaben.
S.4	Mehr offene Arbeit, weniger Aufgaben.
S.5	Man sollte die Schüler entscheiden lassen, wann sie Pause machen.
S.6	Vorträge am Anfang kürzer.
S.7	Alles super.
S.8	Alles gut.
S.9	Alles super.
S.10	Genauere Einführung, mehr Hilfekarten.
S.11	Sehr gut organisiert, Zertifikat für Teilnahme..
S.12	Teilweise umfangreichere Erklärungen..

2.11 Was hast du für dich persönlich durch die Teilnahme am Workshop gelernt?

S.1	Ich bin dem Thema viel näher gekommen und habe einen Einblick erhalten, wie das Arbeiten in solchen Bereichen aussieht
S.2	Umgang mit MATLAB, reale Probleme durch Mathe und Informatik lösen, Informatik als Versuchsmodellierung, Vereinfachung von Problemen.
S.3	Mehr über Google und sein System erfahren.
S.4	Man kann viele reale Probleme mathematisch lösen.
S.5	Wie Google seine Seiten sucht und in welcher Reihe diese aufgelistet werden können, viele Probleme sind mathematisch lösbar.
S.6	Ich habe einen Einblick in MATLAB gewonnen und was mir sehr gut gefallen hat, ich habe etwas über mathematische Modellierung erfahren.
S.7	Modellierung, Umgang mit MATLAB.
S.8	Umgang mit MATLAB, Teilweise Funktion von Google.
S.9	Umgang MATLAB.
S.10	Umgang MATLAB.
S.11	Wie Probleme effizient gelöst werden, wie man Probleme modelliert.
S.12	Modellierung, Umgang mit mathematischen Problemen und MATLAB und Matrizen, Englisch.
S.13	Powerwall, Umgang MATLAB.
S.14	Mathematische Modellierung, Astrophysik.
S.15	Mathematisches Modellieren, Umgang MATLAB.
S.16	Wie man mit MATLABumgeht.
S.17	Wie man kopiert und einfügt.
S.18	Problemlösung Schritt für Schritt.
S.19	MATLAB, Google, bisschen Programmierung.
S.20	wie Google funktioniert.

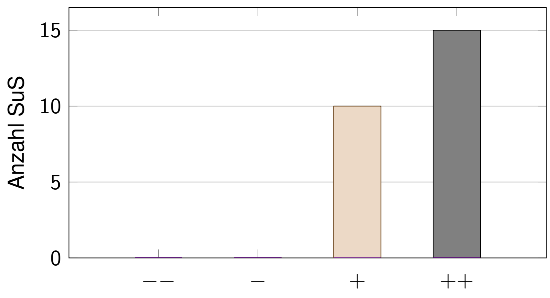
3 Weiterführende Fragen

3.1 Der Workshop hat mein Interesse an den Naturwissenschaften und Technik gesteigert.

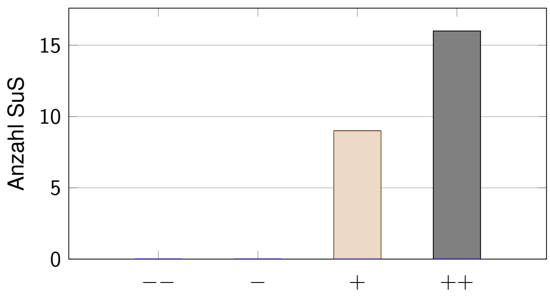


5/8

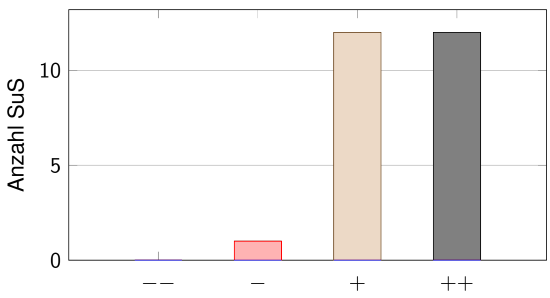
3.2 Die Anleitungen zu den Aufgaben waren zufriedenstellend und gut verständlich.



3.3 Die Dozenten haben die Inhalte plausibel und klar dargestellt.

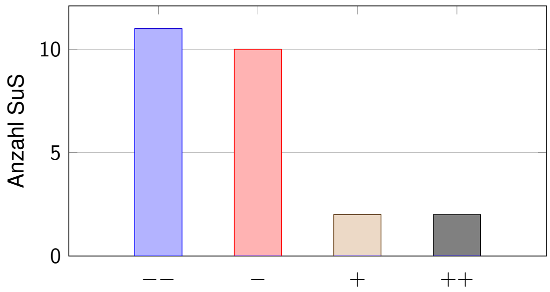


3.4 Die Betreuer haben das selbstständige Arbeiten gefördert.

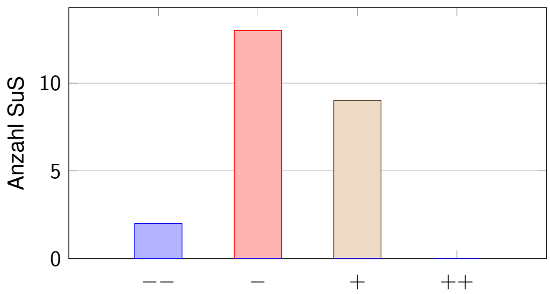


6/8

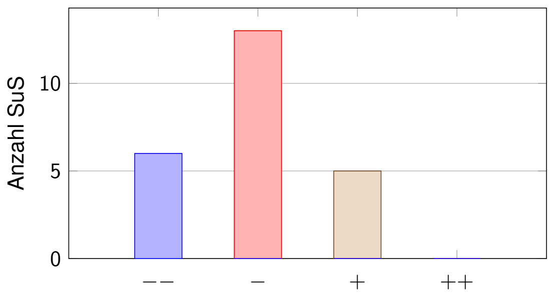
3.5 Es wurden interessante Berufe und Studiengänge vorgestellt.



3.6 Ich kann mir gut vorstellen, später ein Studium oder eine Ausbildung im Bereich Technik oder Naturwissenschaften aufzunehmen.



3.7 Ich würde einen solchen Workshop gerne noch einmal besuchen.



3.8 Abschließender persönlicher Kommentar

S.1	Interessant.
S.2	Informative Vorträge, über das Thema (Google) hinausgehende Informationen.
S.3	Sie waren sehr kompetent und freundlich und sind offen auf Fragen eingegangen.
S.4	Besprechung der Aufgaben war positiv, Kaum was zu verbessern.
S.5	Interessant, abwechslungsreich, Einblick in mathematisches Modellieren.
S.6	Interssanter Tag.
S.7	Hat mir gut gefallen.
S.8	Alles super.
S.9	Alles sehr gut.
S.10	Betreuerinnen waren sehr nett.
S.11	Hilfekarten könnten besser sein.
S.12	Dieses Thema hat mir besonders gefallen, Matrizen waren schwer zu verstehen.

CAMMP day

Wie funktioniert eigentlich Google, und was hat das mit Mathe zu tun?

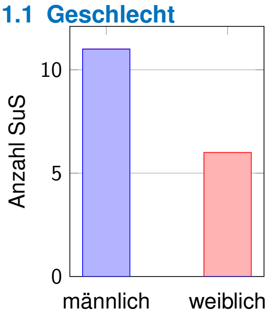
Evaluation

CAMMP day Rhein-Maas Gymnasium

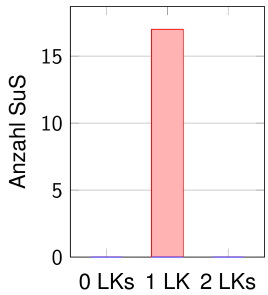
29 Mai 2015

1 Persönliche Angaben

1.1 Geschlecht

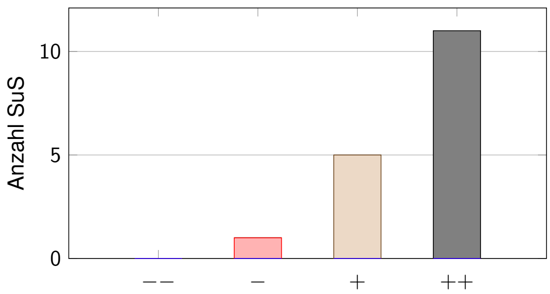


1.2 Anzahl LKs in Mathe, Info und Physik



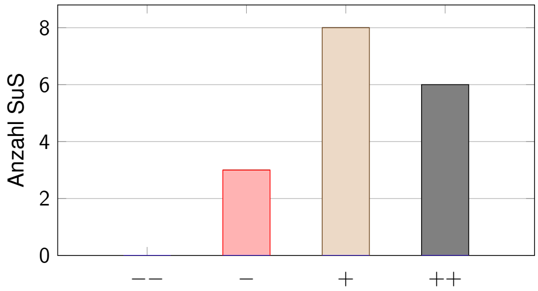
2 Bewertung des Workshops

2.1 Durch den Workshop habe ich mathematisches Modellieren besser begriffen.

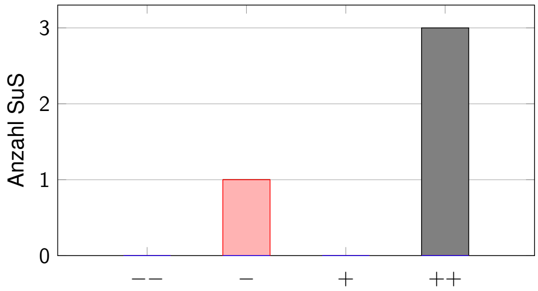


1/7

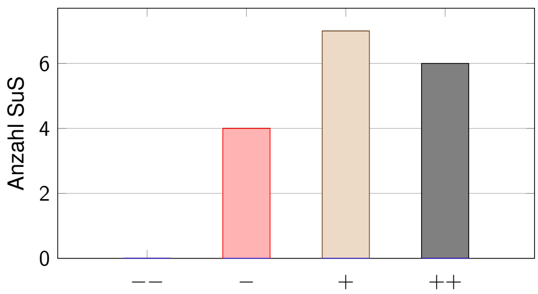
2.2 Der Vortrag über Modellierung war hilfreich.



2.3 Der einführende Kurzfilm war hilfreich.

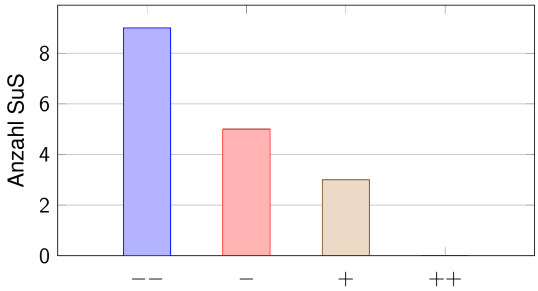


2.4 Die Einführung in MATLAB war hilfreich.

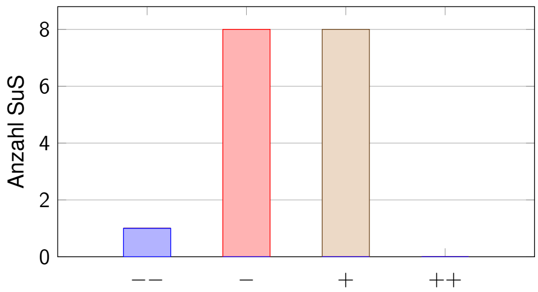


2/7

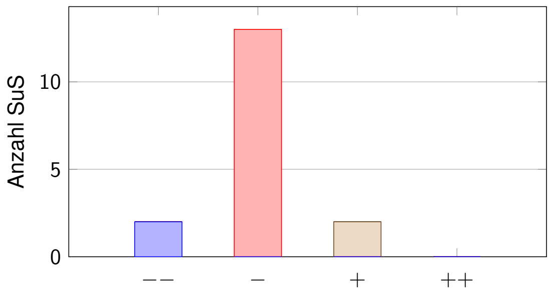
2.5 Der Umgang mit MATLAB fiel mir schwer.



2.6 Die Aufgaben waren zu einfach.

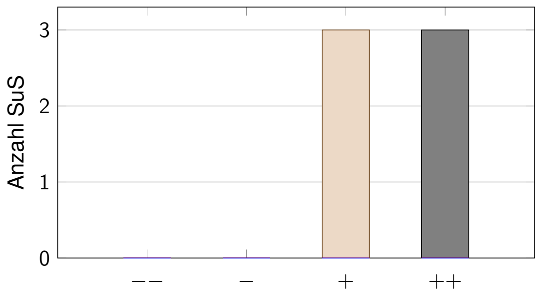


2.7 Die Aufgaben waren zu schwierig.

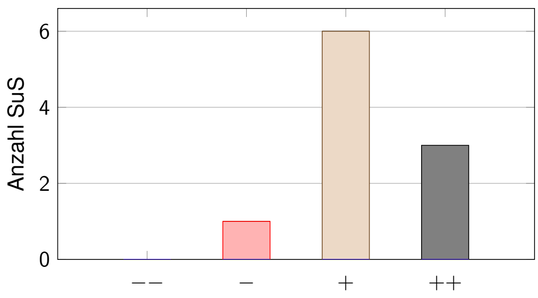


3/7

2.8 Die Hilfekarten waren hilfreich.



2.9 Das Hilfesystem in MATLAB war hilfreich.



2.10 Was würdest du am Workshop verändern bzw. verbessern wollen?

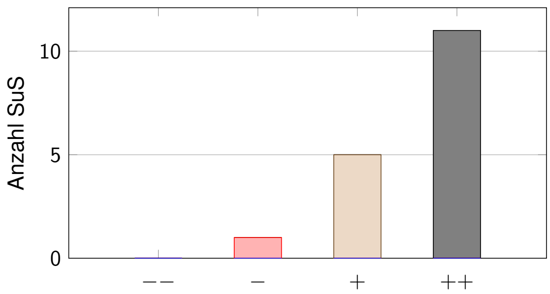
S.1	Längere Einführung in das Programm.
S.2	Einige Fragen könnten besser gestellt werden.
S.3	War gut vorgestellt.
S.4	Sehr gut und informativ.
S.5	Sehr hilfreich und interessant.
S.6	Mittagspause einhalten.
S.7	Alles gut.

2.11 Was hast du für dich persönlich durch die Teilnahme am Workshop gelernt?

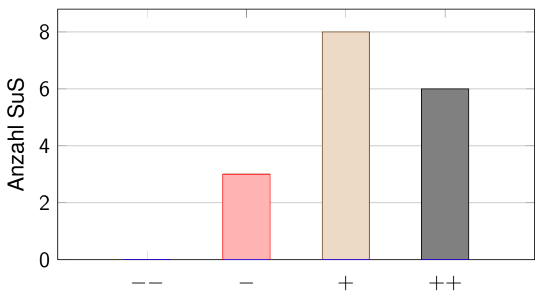
S.1	Umgang mit MATLAB, Ablauf von mathematischen Modellierungen.
S.2	Mit Problemen besser umzugehen und Fehler zu finden.
S.3	Wie Google seine Webseiten anzeigt, etwas über Modellierung, Geschichte der NASA.
S.4	Teilweise Probleme mit MATLAB, meine multimedialen Kenntnisse haben sich durchaus weitergebildet, Aufbau der Aufgaben war gut.
S.5	An Probleme besser dranzugehen und nach Fehlern zu suchen.
S.6	Bedeutung von mathematischen Modellen.
S.7	Umgang mit MATLAB.
S.8	Umgang mit MATLAB, Umgang mit Mac und Powerwall.
S.9	Umgang mit Mac, MATLAB, Powerwall.

3 Weiterführende Fragen

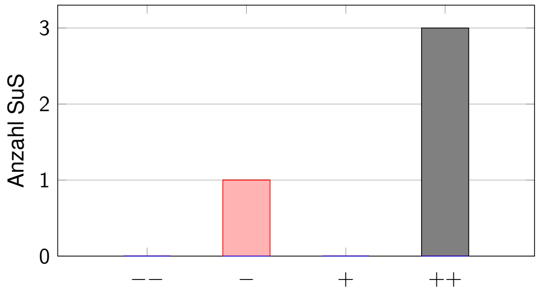
3.1 Der Workshop hat mein Interesse an den Naturwissenschaften und Technik gesteigert.



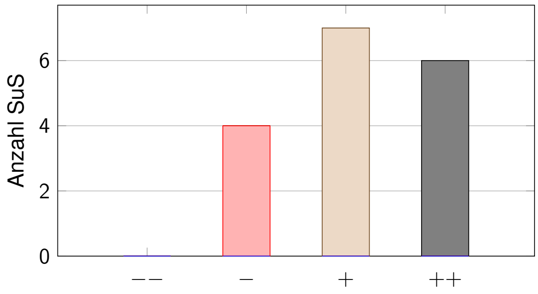
3.2 Die Anleitungen zu den Aufgaben waren zufriedenstellend und gut verständlich.



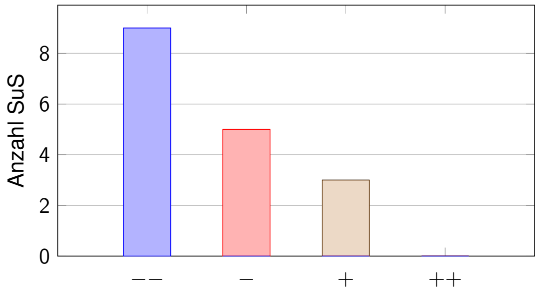
3.3 Die Dozenten haben die Inhalte plausibel und klar dargestellt.



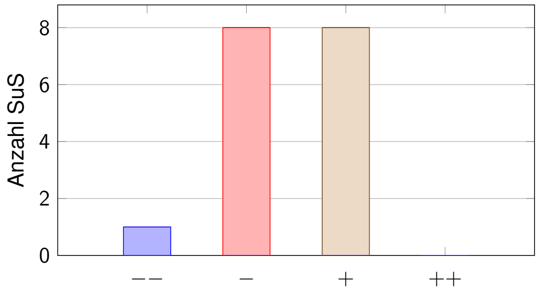
3.4 Die Betreuer haben das selbstständige Arbeiten gefördert.



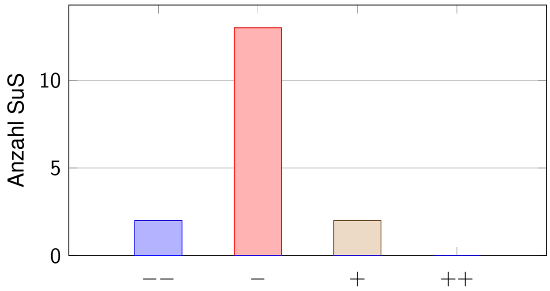
3.5 Es wurden interessante Berufe und Studiengänge vorgestellt.



3.6 Ich kann mir gut vorstellen, später ein Studium oder eine Ausbildung im Bereich Technik oder Naturwissenschaften aufzunehmen.



3.7 Ich würde einen solchen Workshop gerne noch einmal besuchen.



3.8 Abschließender persönlicher Kommentar

S.1	Sehr gut, dass wir selbstständig arbeiten durften.
S.2	Workshop war gut.
S.3	Interessant und klar vorgestellt, jedoch einige Aufgaben unklar.
S.4	Interessant gestaltet.
S.5	Sehr gut.
S.6	Aufgaben waren gut verständlich, Informationstexte haben gut gezeigt, womit wir uns beschäftigen, Dozenten haben gut weitergeholfen.
S.7	War gut.
S.8	Guter Workshop.
S.9	Hat mir gut gefallen, sehr seriös und informativ.
S.10	Alles top.

G. MATLAB Codes

G.1. MATLAB Einführung

```
%% Matlab Einfuehrung Google
% Live-Vorfuehrung mit kurzer Erklaerung
% Ein Wuerfel mit der Kantenlaenge a
% Variablen speichern
a=2;

% Volumen des Wuerfels
V=NaN

% Automatische Ueberpruefung der Loesung
ueberpruefe_Loesung('Einfuehrung',a,V);
```

G.2. MATLAB Code: kleines_netzwerk.m

```
% Schliesse alle Fenster und raeume den Speicher auf
close all
clear all
clc

%% Schritt 1
% Zunaechst speichern wie die Anfangsverteilung
A_0 = 0.25;
B_0 = 0.25;
C_0 = 0.25;
D_0 = 0.25;

%Gib hier deine Formeln fuer A_1 bis D_1 ein
A_1 = NaN
B_1 = NaN
C_1 = NaN
D_1 = NaN

% Um deine Loesung zu ueberpruefen druecke auf "Run Section"
ueberpruefe_Loesung('Aufgabe 1',A_1,B_1,C_1,D_1);

%% Schritt 2
%Eingeben der Matrix U, die beim aufstellen des linearen Gleichungssystems
%in Matrix-Vektor Schreibweise entsteht.
U = [NaN NaN NaN NaN;
     NaN NaN NaN NaN;
     NaN NaN NaN NaN;
     NaN NaN NaN NaN];

display(full(U));
% Um deine Loesung zu ueberpruefen druecke auf "Run Section"
ueberpruefe_Loesung('Aufgabe 2',U);

%% Schritt 3
% Deine Antwort brauchst du nicht bei Matlab eingeben. Wir werden sie
% nachher besprechen.

%% Schritt 4
% Zunaechst speichern wir die Anfangsverteilung in einem Vektor
v_0 = [A_0; B_0; C_0; D_0]
%
n =100;
v_n1 =  NaN


% Welche Seite ist die wichtigste in deinem Netzwerk?
% Um deine Loesung zu ueberpruefen druecke auf "Run"
ueberpruefe_Loesung('Aufgabe 4',v_n1,n);

%% Schritt 5
% Wie aendert sich der Verteilungsvektor fuer unterschiedliche
% Startverteilungen?

% a)
v_0=[ NaN; NaN; NaN; NaN ]
n=  NaN;
v_n2= NaN


% b)
% Welches Ergebnis erhaelst du bei Multiplikation von Matrix U mit dem
% Verteilungsvektor v
w= NaN
```

```
% Um deine Loesung zu ueberpruefen druecke auf "Run"
ueberpruefe_Loesung('Aufgabe 5',v_0,n,v_n2,w);
%% Schritt 6
%Manipuliere das Netzwerk und berechne erneut die Wichtigkeit der Seiten.
%Wann ist der PageRank der Seite A maximal?
v0 = [1/4; 1/4; 1/4; 1/4];

% neue Matrix eingeben
Uneu = [ NaN NaN NaN NaN;
         NaN NaN NaN NaN;
         NaN NaN NaN NaN;
         NaN NaN NaN NaN ];

n=1000;
vneu=Uneu^n * v0

% Deine Loesung werden wir spaeter zusammen besprechen
```

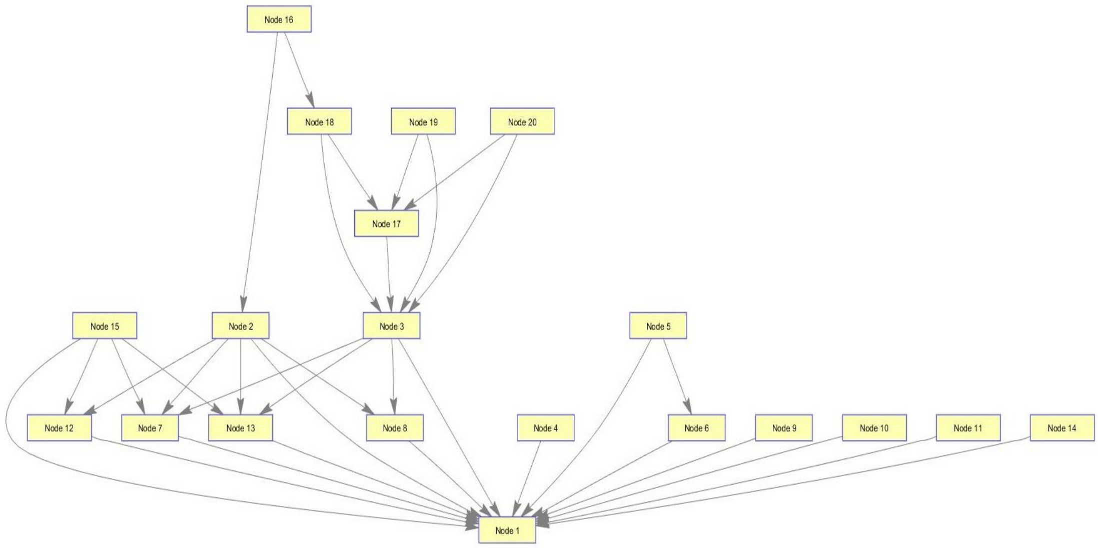
G.3. MATLAB Code: komplexes_netzwerk.m

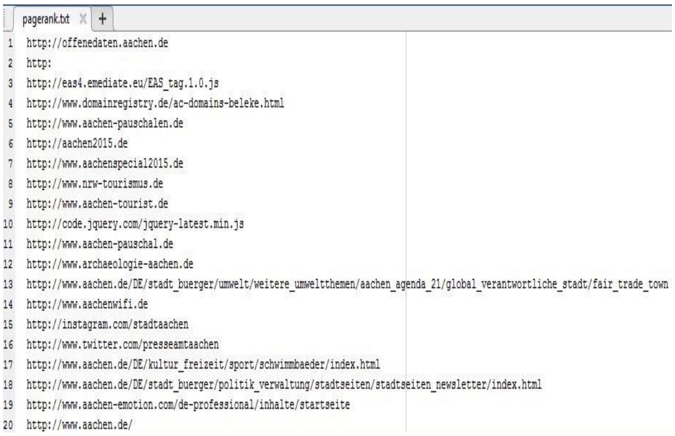
```
% Nun berechnen wird den PageRank fuer ein etwas groesseres Netzwerk

% Schliesse alle Fenster und raeume den Speicher auf
close all
clear all
clc

%% Schritt 1
% Gib zunaechst die Matrix U ein:
U = [ NaN NaN NaN NaN NaN NaN;
       NaN NaN NaN NaN NaN NaN;
       NaN NaN NaN NaN NaN NaN;
       NaN NaN NaN NaN NaN NaN;
       NaN NaN NaN NaN NaN NaN;
       NaN NaN NaN NaN NaN NaN]

% Berechne nun den Verteilungsvektor v und somit den Page Rank.
% Schaue dir den Verteilungsvektor fuer verschiedene n an
v_0 = [1/6; 1/6; 1/6; 1/6; 1/6; 1/6];
n=10000;
v = U^n*v_0

% Um deine Loesung zu ueberpruefen druecke auf "Run Section"
ueberpruefe_Loesung('Schritt 1 Modellverbesserung 1',U);

%% Schritt 2
%Eingeben der neuen Matrix U1
U1 = [ NaN NaN NaN NaN NaN NaN;
       NaN NaN NaN NaN NaN NaN;
       NaN NaN NaN NaN NaN NaN;
       NaN NaN NaN NaN NaN NaN;
       NaN NaN NaN NaN NaN NaN;
       NaN NaN NaN NaN NaN NaN]

% Berechne nun erneut den Verteilungsvektor v und somit den Page Rank.
v_0 = [1/6; 1/6; 1/6; 1/6; 1/6; 1/6];
v = U1^10000 * v_0

% Um deine Loesung zu ueberpruefen druecke auf "Run Section"
ueberpruefe_Loesung('Schritt 2 Modellverbesserung 1',U1);
%% Schritt 3
% Die Matrix G zu Schritt 3 ist bereits eingegeben worden.

G=[0 1 0 0 0;
    1 0 0 0 0;
    0 0 0 1/2 1;
    0 0 1/2 0 0;
    0 0 1/2 1/2 0];
% Berechne nun den Verteilungsvektor v und somit den Page Rank mit
% verschiedenen Startvektoren
v_0 = [0; 1/2; 0; 0; 1/2];
v = G^10001 * v_0

%% Schritt 4
% Gib hier die Matrix U2 ein:
U2  = [ NaN NaN NaN NaN NaN NaN;
       NaN NaN NaN NaN NaN NaN;
       NaN NaN NaN NaN NaN NaN;
       NaN NaN NaN NaN NaN NaN;
       NaN NaN NaN NaN NaN NaN;
       NaN NaN NaN NaN NaN NaN];

% Berechne nun erneut den Verteilungsvektor v und somit den Page Rank.
```

```
v_0 = [1/6; 1/6; 1/6; 1/6; 1/6; 1/6];
v = U2^1000 * v_0

% Um deine Loesung zu ueberpruefen druecke auf "Run Section"
ueberpruefe_Loesung('Schritt 4 Modellverbesserung 2',U2);
 %% Schritt 5
 % Gib hier die Formel zur Berechung der zusammengesetzten Matrix T ein und
 % bestimme den PageRank
 p=0.85;
 T = NaN

 % Um deine Loesung zu ueberpruefen druecke auf "Run"
 ueberpruefe_Loesung('Schritt 5 Modellverbesserung 2',T);
 %% Berechne nun erneut den Verteilungsvektor v und somit den Page Rank.
v_0 = [1/6; 1/6; 1/6; 1/6; 1/6; 1/6];
v = T^10000 * v_0

% Plotte den PageRank - Vektor
% Welche Seite ist die wichtigste in deinem Netzwerk?
figure; bar([1:6],v);
```

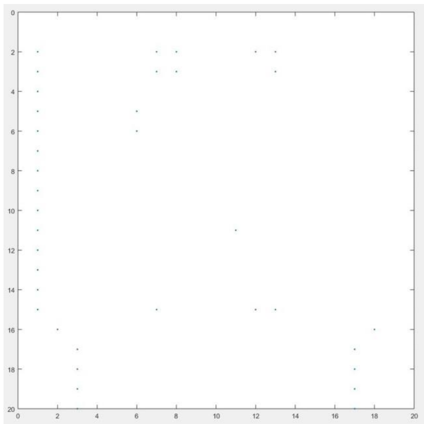
G.4. Grafiken und MATLAB Code: google.m

```
function [] = google()

% Sortiert ausgehend von einer Start-Webseite alle Webseiten nach Relevanz
% Ergebnis in Datei pagerank.txt
% Suchtiefe fuer Adjazenzgraph
m = 20;

% Startwebseite
site = 'http://www.aachen.de/';

% Korrekturfaktor fuer Google-Matrix
p = 0.85;

% Erstellung des Adjazenzgraphen mit Namen
[U,L] = surfer(site,m);

% Groesse
n = size(L,1);

% Spaltensummen
s = sum(L);

% Google-Matrix
for j=1:n
    if s(j)~=0
        G(:,j) = L(:,j)/s(j);
    else
        G(:,j) = 1/n*ones(n,1);
    end
end

% Mit Korrekturfaktor wird die Google-Matrix berechnet
G = p*G + 1/n*(1-p)*ones(n,n);

% Berechne die Eigenwerte und Eigenvektoren
[V,D] = eig(G);

% Die Eigenwerte sind geordnet, mit 1 als groesstetem, also ist der
% Relevanz-Vektor der erste Eigenvektor
r = abs(V(:,1));

% Sortieren nach PageRank
[~,index] = sort(r);

% Liste der sortierten Webseiten
list = U(index);
writeresult(list);

function [] = writeresult(list)
fid = fopen('pagerank.txt','w');
list1=cell(size(list,1),1);
for k=1:size(list,1)          % Ordnung umkehren
   list1{k} = list{size(list,1)-k+1};
end
fmtString = [repmat('%s\t',1,size(list,2)-1),'%s\n'];
fprintf(fid,fmtString,list1{:});
fclose(fid);
```




pagerank.txt

1 http://offenedaten.aachen.de
2 http:
3 http://eas4.emediate.eu/EAS_tag.1.0.js
4 http://www.domainregistry.de/ac-domains-beleke.html
5 http://www.aachen-pauschalen.de
6 http://aachen2015.de
7 http://www.aachenspecial2015.de
8 http://www.nrw-tourismus.de
9 http://www.aachen-tourist.de
10 http://code.jquery.com/jquery-latest.min.js
11 http://www.aachen-pauschal.de
12 http://www.archaeologie-aachen.de
13 http://www.aachen.de/DE/stadt_buerger/umwelt/weitere_umweltthemen/aachen_agenda_21/global_verantwortliche_stadt/fair_trade_town
14 http://www.aachenwifi.de
15 http://instagram.com/stadtaachen
16 http://www.twitter.com/presseamtaachen
17 http://www.aachen.de/DE/kultur_freizeit/sport/schwimmbaeder/index.html
18 http://www.aachen.de/DE/stadt_buerger/politik_verwaltung/stadtseiten/stadtseiten_newsletter/index.html
19 http://www.aachen-emotion.com/de-professional/inhalte/startseite
20 http://www.aachen.de/



Eigenständigkeitserklärung

Hiermit versichere ich, dass ich diese Bachelorarbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Die Stellen meiner Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen sind, habe ich in jedem Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht. Dasselbe gilt sinngemäß für Tabellen und Abbildungen. Diese Arbeit hat in dieser oder einer ähnlichen Form noch nicht im Rahmen einer anderen Prüfung vorgelegen.

Aachen, den 13. August 2015

Sarah Schönbrodt

Literaturverzeichnis

[1] Basten, E. : *Praxisbericht: Förderung inhalts- und prozessbezogener Kompetenzen.* http://www.finken.de/praxisbericht_kompetenzen_mathematik_basten. Version: 2011, letzter Aufruf: 12.07.2015

[2] Büchter, A. ; Henn, H.-W. : *Elementare Stochastik.* Springer, 2007

[3] Besser, M. ; Hagena, M. ; Leiss, D. : Lehrerlösungsprozesse beim Mathematischen Modellieren. In: Henn, H.-W. (Hrsg.) ; Kaiser, G. (Hrsg.): *Werner Blum und seine Beiträge zum Modellieren im Mathematikunterricht - Festschrift zum 70. Geburtstag von Werner Blum.* Wiesbaden: Springer-Spektrum, 2015

[4] Blum, W. : Modellierungsaufgaben im Mathematikunterricht – Herausforderung für Schüler und Lehrer. In: Büchter, A. (Hrsg.) ; Humenberger, H. (Hrsg.) ; Hussmann, S. (Hrsg.) ; Prediger, S. (Hrsg.): *Realitätsnaher Mathematikunterricht – vom Fach aus und für die Praxis. Festband für Hans-Wolfgang Henn zum 60. Geburtstag.* Hildesheim: Verlag Franzbecker, 2006

[5] Blum, W. : Mathematisches Modellieren. In: *Beiträge zum Mathematikunterricht 2007.* Hildesheim, Berlin: Verlag Franzbecker, 2007

[6] Bryan, K. ; Leise, T. : *The $25,000,000,000 Eigenvector: The Linear Algebra behind Google.* http://dx.doi.org/10.1137/050623280. Version: 2006, letzter Aufruf: 12.07.2015

[7] Cho, J. ; Garcia-Molina, H. ; Page, L. : *Efficient Crawling Through URL Ordering.* http://dx.doi.org/10.1016/S0169-7552(98)00108-1. Version: 1998, letzter Aufruf: 10.07.2015

[8] Greefrath, G. ; Kaiser, G. ; Blum, W. ; Borromeo Ferri, R. : Mathematisches Modellieren – Eine Einführung in theoretische und didaktische Hintergründe. In: Borromeo Ferri, R. (Hrsg.) ; Greefrath, G. (Hrsg.) ; Kaiser, G. (Hrsg.): *Mathematisches Modellieren für Schule und Hochschule.* Wiesbaden: Springer-Verlag, 2013

[9] Humenberger, H. : Das Google-PageRank-System – Mit Markoff-Ketten und linearen Gleichungssystemen Ranglisten erstellen. In: *mathematik lehren* (2009)

[10] Humenberger, H. : Der PageRank von Google – eine aktuelle Anwendung im Mathematikunterricht. In: Neubrand, M. H. (Hrsg.): *Beiträge zum Mathematikunterricht 2009.* Münster: WTM-Verlag, 2009

[11] Humenberger, H. : Google's PageRank: A Present-Day Application of Mathematics in the Classroom. In: *Trends in Teaching and Learning Mathematical Modeling.* Heidelberg, London, New York: Springer-Verlag, 2011, S. 579–589

[12] IQB (Hrsg.): *Kompetenzentwicklung im Mathematik-Unterricht: Modellieren. Handreichungen zu VERA 8 Mathematik 2009.* "http://www.standardsicherung.schulministerium.nrw.de/lernstand8/upload/download/mat_mathematik/Kompetenzentwicklung_Modellieren.pdf". Version: 2009, letzter Aufruf: 02.07.2015

[13] Kultusministerkonferenz: *Bilungsstandards im Fach Mathematik für den Mittleren Schulabschluss.* "http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2003/2003_12_04-Bildungsstandards-Mathe-Mittleren-SA.pdf". Version: 2003, letzter Aufruf: 09.07.2015

[14] Langville, A. N. ; Meyer, C. D.: Deeper Inside Pagerank. In: *Internet Mathematics* (2004)

[15] Leuders, T. : Chancen und Risiken des Computereinsatzes im Mathematikunterricht. In: Leuders, T. (Hrsg.): *Mathematikdidaktik Praxishandbuch für die Sekundarstufe I und II.* Berlin: Cornelsen Verlag, 2011

[16] Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen: *Kernlehrplan für das Gymnasium – Sekundarstufe I (G8) in Nordrhein-Westfalen – Mathematik.* "http://www.schulentwicklung.nrw.de/lehrplaene/upload/lehrplaene_download/gymnasium_g8/gym8_mathematik.pdf". Version: 2007, letzter Aufruf: 09.07.2015

[17] Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen: *Kernlehrplan für die Sekundarstufe II Gymnasium / Gesamtschule in Nordrhein-Westfalen – Mathematik.* "http://www.schulentwicklung.nrw.de/lehrplaene/upload/klp_SII/m/KLP_GOSt_Mathematik.pdf". Version: 2014, letzter Aufruf: 09.07.2015

[18] Page, L. ; Brin, S. ; Motwani, R. ; Winograd, T. : *The PageRank Citation Ranking: Bringing Order to the Web.* http://ilpubs.stanford.edu:8090/422/. Version: 1998, letzter Aufruf: 29.06.2015

[19] Westermann, B. : Anwendungen und Modellbildung. In: Leuders, T. (Hrsg.): *Mathematikdidaktik Praxishandbuch für die Sekundarstufe I und II.* Berlin: Cornelsen Verlag, 2011

[20] Wills, R. S.: Google's PageRank: The Math Behind the Search Engine. In: *Math. Intelligencer* (2006)

Abbildungsverzeichnis