Der PageRank von Google – eine aktuelle Anwendung im Mathematikunterricht

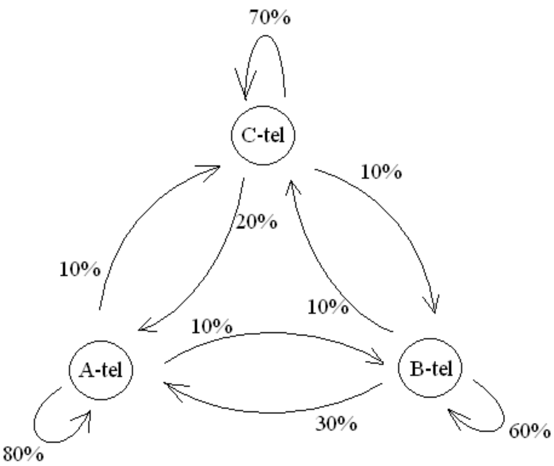
Hans HUMENBERGER , Universität Wien

1 Einleitung

Das Thema „Mehrstufige Prozesse“ (in einer elementaren Form!) gehört in manchen deutschen Bundesländern zum möglichen Lehrstoff in der Oberstufe. In der Tat ist es ja ein Gebiet, in dem der Vernetzungsgedanke sehr gut verwirklicht werden kann, und *Vernetzung* gehört sicher zu jenen Begriffen, die in den Präambeln von so gut wie allen Lehrplänen als positiv und wünschenswert hervorgehoben werden. Das Thema gehört einerseits zur Linearen Algebra („Übergangsmatrizen“), andererseits zur Stochastik (Wahrscheinlichkeiten, relative Häufigkeiten) bzw. zur Analysis (Grenzwerte), so dass in diesem Thema auch eine sehr gute Realisierung des so wichtigen Vernetzens gesehen werden kann.

Google ist mittlerweile die Suchmaschine im WWW schlechthin geworden mit einem Marktanteil von 62% (Juni 2009), jede(r) hat viel Erfahrung damit, und es erhebt sich die interessante Frage: Wie kommt Google eigentlich zu einer Reihung im Suchergebnis (in der „Liste“), wie schafft es Google, dass bzgl. des Suchbegriffs *relevante* Seiten am *Anfang* der Liste stehen?

Als **Einstiegsaufgabe** wäre denkbar:
Der Telefonmarkt in einem Land sei durch drei Telefongesellschaften bestimmt (A-tel, B-tel und C-tel). Diese schließen mit ihren Kunden jeweils Jahresverträge ab[1], wobei die Kunden zu einem bestimmten Prozentsatz bleiben bzw. zu anderen Betreibern wechseln. Man kann das am einfachsten mit einem so genannten *gerichteten Graphen* (vgl. nebenstehende Abb., auch *Übergangsgraph* genannt) beschreiben: Dies bedeutet für die Firma C-tel z. B.: 70% der C-tel-Kunden bleiben nach Ablauf des Jahresvertrages bei C-tel, 20% wechseln zu A-tel und 10% zu B-tel. Analog sind die anderen relativen Übergangshäufigkeiten zu interpretieren. Angenommen, diese Übergangsraten bleiben über 5 (10; 20) Jahre konstant; wie sieht die Verteilung der Kunden auf die einzelnen Firmen aus, wenn zu Beginn alle 1/3 der Kunden haben? Wie wären die entsprechenden Werte, wenn die „Anfangsverteilung“ nicht $(A_0, B_0, C_0) = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$, sondern $(A_0, B_0, C_0) = (30\%, 50\%, 20\%)$ wäre? Auch wenn die Schüler(innen) noch nichts von Markoff-Ketten und Übergangsmatrizen gehört haben, können sie dieses Problem relativ leicht lösen, am besten vielleicht mit einer Tabellenkalkulation (z. B. EXCEL). Die zugehörigen Rekursionen können aus dem *Übergangsgraphen* direkt



$$\begin{aligned}
0{,}8A_n + 0{,}3B_n + 0{,}2C_n &= A_{n+1}\\
0{,}1A_n + 0{,}6B_n + 0{,}1C_n &= B_{n+1}\\
0{,}1A_n + 0{,}1B_n + 0{,}7C_n &= C_{n+1}
\end{aligned}$$

[1] Der Einfachheit halber nehmen wir an, dass diese Verträge jeweils genau ein Kalenderjahr gelten mit Wechselmöglichkeiten jeweils zu Jahresende/Jahresbeginn.

abgelesen und dort als Formel eingegeben werden. Durch das bekannte „Herunterziehen" können dann die entsprechenden Werte auch nach vielen Jahren schnell abgelesen werden (allein mit einem Taschenrechner wäre das ungleich mühsamer). Man wird hier feststellen, dass sich die entsprechenden Werte schnell bei $(A_n, B_n, C_n) = (55\%,\ 20\%,\ 25\%)$ „einpendeln" und zwar unabhängig von der Anfangsverteilung. Wir haben hier also eine stabile „Grenzverteilung" unabhängig von der Anfangsverteilung.

Genau solche Grenzverteilungen werden im Folgenden im (mathematischen) Zentrum stehen. Für das experimentelle Ermitteln solcher Grenzverteilungen zwischen einigen wenigen möglichen „Stationen" (oben nur 3: A-tel, B-tel, C-tel) ist eine *Tabellenkalkulation* sehr gut geeignet. Man braucht dabei weder Matrizen noch weitere Theorie dazu, nur ganz elementare z. B. Kenntnisse bei Tabellenkalkulationen. Das Bestimmen dieser Grenzverteilung geschieht dabei *iterativ*. Beim PageRank-Algorithmus geschieht das Lösen des zugehörigen linearen Gleichungssystems auch iterativ (näherungsweise) und nicht durch eine geschlossene Formel oder Ähnliches – siehe unten. Dieses iterative Vorgehen mit Tabellenkalkulation ist also einerseits ein einfacher Einstieg, andererseits aber gar nicht so weit weg von der prinzipiellen Vorgehensweise, wie solche „Grenzverteilungen" bei riesigen Dimensionen bestimmt werden.

Für eine etwas weiter und tiefer gehende Auseinandersetzung mit „Grenzverteilungen" (insbesondere theoretische Aspekte) kommt man mit Tabellenkalkulation nicht mehr aus, da braucht man dann *Übergangsmatrizen* – siehe unten.

2 Google und seine Gründer

Es gibt sehr viele „Suchmaschinen", die das WWW in Bruchteilen von Sekunden gleichsam durchforsten können, wobei „wichtige" Seiten jeweils zuerst aufgelistet werden sollen.

Eine besonders häufig verwendete Suchmaschine ist z. B. Google, wobei der Name Google „etwas Riesengroßes" bezeichnen sollte – nach der unglaublichen Fülle des WWW. Er kommt von einer Abwandlung des Begriffes „Googol", ein Wort, das in der ersten Hälfte des 20. Jahrhunderts in Amerika für die riesengroße Zahl 10^{100} etabliert wurde. Ein Googol ist zwar deutlich größer als die Anzahl der Atome im sichtbaren Universum (ca. 10^{80}), andererseits ist ein Googol aber „nur" etwa $70 \cdot 69 \cdot 68 \cdot \ldots \cdot 2 \cdot 1 = 70!$ und entspricht somit der Anzahl der Möglichkeiten, 70 verschiedene Gegenstände in einer Reihe anzuordnen. Da merkt man wieder einmal, wie mächtig so manche Schreibweisen in der Mathematik sind und wie schnell Fakultäten wachsen: 70! ist ca. 10^{20}-mal so groß wie die Anzahl der Atome im sichtbaren Universum! Wer würde so aus dem Bauch heraus nicht glauben, dass es doch viel, viel mehr Atome im Universum gibt als Möglichkeiten, 70 Gegenstände in einer Reihe anzuordnen?

Google ist unter den Suchmaschinen unbestritten die Nummer 1 (gegenüber z. B. Yahoo, MSN, etc.). Warum kam es so weit, warum konnte sich Google quasi „durchsetzen"? Der Grund dafür liegt im „PageRank-Algorithmus", der Google damals schneller machte als die Konkurrenz. Und gerade beim Aufteilen des potentiellen Marktes ist es wichtig, besser zu sein als die Konkurrenz. Auch wenn die Qualität der anderen Suchmaschinen seitdem gestiegen ist, ist selbst bei gleich guter Qualität für die vielen Anwender von Google (ich selbst gehöre auch dazu) kein Grund zum Umsteigen gegeben. Dies hat einfach auch mit Gewohnheit zu tun: warum soll man umsteigen, wenn anderswo die Qualität nicht besser ist?

Die Suchmaschinen beginnen damit, mit einem so genannten *spider* oder *webcrawler* (spezielles Computerprogramm) das WWW zu „durchforsten“: Verlinkungsstruktur und Inhalte der WWW-Seiten und -Dokumente werden „aufgezeichnet“: Auf welchen Seiten ist zum gesuchten Begriff etwas zu erfahren, wo kommt er vor? Ziel dieser umfangreichen Suchtätigkeit ist eine möglichst gute „Momentaufnahme“ der Inhalte und der Vernetzungsstruktur.

Jede Anfrage bei Google wird in Form einer Liste von „Treffern“ beantwortet. Um für wichtige Informationen nicht seitenweise blättern zu müssen, sollen zum jeweiligen Thema wichtige Seiten auch zuerst aufgelistet werden, es stellt sich also bei jeder Suchmaschine die Frage: in welcher Reihenfolge werden die Treffer dem Anfragesteller präsentiert? Wie kann bewerkstelligt werden, dass die Liste der Treffer mit den „relevanten“ und „wichtigen“ Webseiten beginnt?

Bei Wikipedia heißt es dazu: „Der **PageRank**-Algorithmus ist ein Verfahren, eine Menge verlinkter Dokumente, wie beispielsweise das WWW, anhand ihrer Struktur zu bewerten bzw. zu gewichten. Dabei wird jedem Element ein Gewicht, der PageRank, aufgrund seiner Verlinkungsstruktur zugeordnet. Der Algorithmus wurde von Larry PAGE (daher der Name PageRank) und Sergej BRIN an der Stanford University entwickelt und von dieser zum Patent angemeldet. Er diente dem von BRIN und PAGE gegründeten Unternehmen Google als Grundlage für die Bewertung von Seiten.“

Lawrence (Larry) PAGE (geb. 1973, im Bild links) ist ein US-amerikanischer Informatiker und Mitbegründer der Suchmaschine Google. An der Stanford University absolvierte er das Masterstudium in Informatik. Gemeinsam mit seinem Studienkollegen **Sergej Michailowitsch BRIN** (geb. 1973 in Moskau) schuf er den Prototypen einer Suchmaschine für das WWW. Keines der großen Portale (heutige Konkurrenten wie z. B. Yahoo) interessierte sich dafür, daher gründeten sie gemeinsam 1998 die Firma Google Inc. mit einer 100 000 $-Unterstützung durch die Firma „Sun Microsystems“. Begonnene Promotionen lassen sie seit der Gründung von Google angeblich ruhen. Der Aktienwert dieses Unternehmens ging in unermessliche Höhen, so dass sie sicher zu den reichsten Menschen der Welt zu zählen sind, durch den Börsengang der Firma wurden sie zu Multimilliardären.

Insgesamt muss man sagen, dass der Algorithmus von Google in der Praxis natürlich komplizierter abläuft als hier dargestellt (es fließen da noch viele andere Nebenbedingungen ein), aber die prinzipielle hinter dem PageRank-Algorithmus von Google steckende mathematische Idee ist eigentlich eine sehr elementare.

Es ist einerseits verblüffend, mit welch elementarer zugrunde liegender Idee so viel Geld zu machen ist, wie es den Gründern von Google gelungen ist. Andererseits liegt darin wieder einmal mehr eine wohltuende Bestätigung, dass grundlegende mathematische Ideen sehr wichtig werden – hier für die Millionen User von Google und wirtschaftlich gesehen für die Begründer dieser Firma und für die zugehörigen Aktionäre.
Dies soll die Leistung der beiden oben genannten Herren in keiner Weise schmälern. Die konkrete Umsetzung dieser Idee in ein Programm, das diese Reihung auch bei 1000 000 oder noch mehr Seiten in akzeptabel kurzer Zeit erledigt, ist eine höchst schwierige Aufgabe und eine tolle Leistung!

Bei Google wurden schon 2005 etwas über 8 Milliarden URLs durchsucht und etwas über 1 Milliarde Bilder. Google bedient in jeder Sekunde[2] sehr viele Anfragen in über 100 „Domains“ und Sprachen, und alle wollen ihr Ergebnis sofort, d. h. auf Knopfdruck ohne zu warten. Es wurde und wird eine Antwortzeit von höchstens einer halben Sekunde als Richtwert angestrebt. Diese schnelle Lieferung von Ergebnissen hat auch frühzeitig zur Popularität von Google beigetragen, die Konkurrenz hat sich mit der Anfragebeantwortung oft mehr Zeit gelassen und ist dadurch eindeutig ins Hintertreffen geraten. Inzwischen beschäftigt Google eine ganze Schar hervorragender Programmierer, aber die ersten Schritte dürften die beiden Gründungsherren schon weitgehend alleine bewerkstelligt haben – Hut ab!

Im Magazin *Stern* ist online zu lesen: „Während ihr Unternehmen inzwischen der mit Abstand populärste Betreiber von Internet-Suchmaschinen ist und Milliarden an der Börse einnehmen will, versuchen die beiden, die lockere Atmosphäre einer Studenten-Firma in die Welt des Big Business zu tragen. In dem Börsenprospekt werden sie und andere Manager mit Vornamen genannt, ‚nicht bösartig zu sein’ wird zu einem der obersten Grundsätze postuliert und vergangenes Jahr verdienten die baldigen Online-Milliardäre gerade mal 350 000 Dollar. Google ist keine gewöhnliche Gesellschaft und will auch keine werden.“ Dann heißt es weiter: „Der als Konzernchef verpflichtete Eric Schmidt hat eine ganz andere Geschichte. Mit 48 Jahren ist er viel älter und war bereits vier Jahre lang Konzernchef beim Softwareunternehmen Novell. Zuvor war er unter anderem Technologie-Chef beim Computerkonzern Sun Microsystems. Schmidts Erfahrungen in der hart umkämpften Branche gelten als wichtige Ergänzung zu den Visionen der jungen Firmengründer. Wohl auch deshalb verdient er deutlich mehr als sie: 550 000 Dollar im vergangenen Jahr.“

Angeblich wollen Page, Brin und Schmidt noch bis mindestens 2024 im Konzern bleiben, na dann „ad multos annos“!

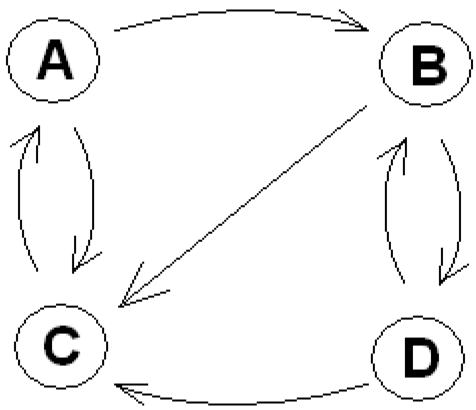
Es hat sich bei Internetrecherchen sogar schon das zugehörige Verbum „googeln“ eingebürgert, auch im Englischen spricht man von „to google“. Wenn jemand zu einem bestimmten Begriff jemand anderen fragt, so hört man oft: „Hast du diesen Begriff schon *gegoogelt*, um mehr darüber zu erfahren?“

Bei Wikipedia ist nachzulesen: Der Rechtschreib-Duden hat das Verb 2004 in die 23. Auflage aufgenommen. Dabei steht das Wort nicht nur für die Suche mit Google, sondern teilweise schon allgemein als Synonym für Websuche mit beliebigen anderen Suchmaschinen. Seit August 2006 will Google rechtlich klarstellen, dass die Verwendung der Verben „to google“ oder deutsch „googeln“ nur noch in Verbindung mit der Suchmaschine Google korrekt ist. Die aktuelle 24. Auflage des Duden erläutert das Verb googeln nun explizit als „mit Google im Internet suchen“. Google will offenbar verhindern, dass mit seinem Namen das Gleiche passiert wie z. B. mit Nutella, Tesa, Fön, Tixo, Aspirin, dem Grammophon oder dem escalator (engl. für „Rolltreppe“) – alle waren ursprünglich geschützte Markennamen und sind heute gewissermaßen zu Gattungsnamen geworden.

---

[2] Allein in Amerika gab es schon 2005 pro Tag durchschnittlich 60.000.000 Suchanfragen von *Erwachsenen*, d. h. pro Sekunde ca. 700 (vgl. WILLS 2006, S. 6). CHARTIER (2006, S. 17) schreibt, dass Google pro Sekunde mehr als 3000 Anfragen bekommt; ich vermute, auch das ist für nur Amerika gemeint, nicht weltweit. In einem Spiegel-Artikel 2/2010 (S. 62) ist zu lesen, dass Google weltweit jeden Tag ungefähr 3 Milliarden Anfragen, d. h. ca. 35.000 pro Sekunde hat.

3 Das WWW als gerichteter Graph: Beschreibung durch Übergangsmatrizen



Zunächst werden bei Google so genannte *spider* ausgeschickt, die das www, d. h. die darin befindlichen html-, pdf- etc. -dokumente samt ihrer Vernetzung (hier: Verlinkung) als gerichteten Graphen darstellen. Wenn man einen Suchbegriff eingibt, so wird jener Teilgraph von Dokumenten genommen, in denen der Suchberiff vorkommt. Auf die Details, wie ein solcher erstellt wird, kommt es hier nicht an. Wir wählen zunächst ein ganz einfaches Beispiel. A, B, C, D seien 4 verschiedene Internetseiten, die auf eine einfache Weise miteinander verlinkt sind, so gibt es z. B. von Seite A einen Link zu den Seiten B und C, von der Seite B gibt es Links zu den Seiten C und D usw. (vgl. den nebenstehenden Graphen).

1. Modellannahme: Man geht der Einfachheit halber davon aus, dass alle Links auf einer Seite mit jeweils derselben relativen Häufigkeit bzw. Wahrscheinlichkeit benutzt werden[3], so dass bei zwei ausgehenden Pfeilen bei beiden ½ stehen soll, bei drei ausgehenden Pfeilen bei allen drei jeweils 1/3 etc. Wegen dieser bewusst vereinfachenden Annahme sind die Wahrscheinlichkeiten bei den einzelnen Pfeilen gar nicht dazugeschrieben.

Nun kann man sich wieder wie bei den Telefongesellschaften vorstellen, dass eine Menge „User" sich auf die Seiten A, B, C, D verteilen, zu Beginn mit den relativen Anteilen (A_0, B_0, C_0, D_0). Der obige Wechsel der Telefongesellschaft entspricht hier einem Wechsel der Internetseite. Wir stellen uns dieses System wieder in diskreten Schritten vor: Die User wechseln in gewissen Zeiteinheiten die Internetseite (indem sie einem Link folgen oder auch nicht), so dass sich nach n Zeitschritten eine Verteilung von (A_n, B_n, C_n, D_n) ergibt (*relative Anteile*). Wir können die zugehörigen Rekursionen wieder leicht aus dem Übergangsgraphen ablesen und aufschreiben:

$$\begin{array}{lllll}
 & & C_n & & = A_{n+1} \\
0{,}5A_n & & & +0{,}5D_n & = B_{n+1} \\
0{,}5A_n & +0{,}5B_n & & +0{,}5D_n & = C_{n+1} \\
 & 0{,}5B_n & & & = D_{n+1}
\end{array}$$

Um zu überprüfen, ob bzw. bei welcher Verteilung $(\overline{A}, \overline{B}, \overline{C}, \overline{D})$ sich die User dieser 4 Seiten à la longue einpendeln werden, könnten wir wieder mit EXCEL arbeiten. Aber man kann lineare Gleichungssysteme bekanntlich auch gut mit Matrizen und Vektoren beschreiben, das obige z. B. durch:

---

[3] Natürlich entspricht dies nicht ganz der Realität; ein Link am Ende einer langen html-Seite wird wahrscheinlich nicht dieselbe Wahrscheinlichkeit haben benutzt zu werden wie einer, der prominent ganz oben in der Seite „thront". Aber dieses Vereinfachen und Idealisieren (so dass der „Sache" mit mathematischen Mitteln beizukommen ist) ist ein typischer Schritt bei so gut wie allen Modellierungen.

$$\underbrace{\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0{,}5 & 0 & 0 & 0{,}5 \\ 0{,}5 & 0{,}5 & 0 & 0{,}5 \\ 0 & 0{,}5 & 0 & 0 \end{pmatrix}}_{=:U} \cdot \underbrace{\begin{pmatrix} A_n \\ B_n \\ C_n \\ D_n \end{pmatrix}}_{=:\vec{v}_n} = \underbrace{\begin{pmatrix} A_{n+1} \\ B_{n+1} \\ C_{n+1} \\ D_{n+1} \end{pmatrix}}_{\vec{v}_{n+1}}$$

Alle Übergänge $\vec{v}_n \rightarrow \vec{v}_{n+1}$ werden durch dieselbe Matrix U („Übergangsmatrix“) vermittelt, dabei stehen in der Spalte i die Wahrscheinlichkeiten, dass ein User, der sich gerade auf Seite i befindet, sich im nächsten Schritt auf Seite j befinden wird d. h. einen Link zur Seite j benutzt (durch den Eintrag in Zeile j gegeben; $i, j = 1, \ldots, 4$). So muss z. B. jemand, der sich auf Seite C befindet, im nächsten Schritt zwangsweise (Wahrscheinlichkeit 1) zu Seite A kommen, was im Übergangsgraphen und in der Übergangsmatrix anhand der 1 in Spalte 3 und Zeile 1 abzulesen ist.

Für die Übergänge erhält man der Reihe nach:

$$U \cdot \vec{v}_0 = \vec{v}_1, \quad \underbrace{U \cdot \overbrace{\left(U \cdot \vec{v}_0\right)}^{\vec{v}_1}}_{U^2 \cdot \vec{v}_0} = \vec{v}_2, \quad \underbrace{U \cdot \overbrace{\left(U \cdot \underbrace{\left(U \cdot \vec{v}_0\right)}_{\vec{v}_1}\right)}^{\vec{v}_2}}_{U^3 \cdot \vec{v}_0} = \vec{v}_3, \; \ldots, \; \boxed{U^n \cdot \vec{v}_0 = \vec{v}_n}.$$

Wir haben mit Hilfe von Matrizen somit eine Möglichkeit, eine *geschlossene* Formel für die Verteilung $\vec{v}_n$ zu erhalten (*nicht* nur *iterativ* wie mit Tabellenkalkulation).

Vektoren, die relative Anteile oder Wahrscheinlichkeiten enthalten, heißen in der Mathematik auch „stochastische Vektoren“[4]. Eine quadratische Matrix heißt „stochastisch“, wenn ihre Spaltenvektoren stochastisch sind[5]. Übergangsmatrizen sind natürlich stochastische Matrizen, sie sind quadratisch, und in der ersten Spalte z. B. stehen die relativen Häufigkeiten, mit der User von A im nächsten Schritt zu den Seiten A, B, C, D wechseln. Diese Zahlen sind naturgemäß aus dem Intervall [0; 1] und ergeben in Summe 1 (analog bei den anderen Spalten).

An dieser Stelle hat man auch die Möglichkeit leichte allgemeine Begründungen zu thematisieren, z. B.:

- das Produkt einer stochastischen Matrix mit einem stochastischen Vektor ist wieder ein stochastischer Vektor,
- das Produkt zweier stochastischer Matrizen ist wieder eine stochastische Matrix.

Eine wichtige Anwendung erfahren hier auch das Potenzieren von Matrizen und das Assoziativgesetz der Matrizenmultiplikation, z. B. bei

$\vec{v}_3 = U \cdot \left(U \cdot \left(U \cdot \vec{v}_0\right)\right) = \left(U \cdot U \cdot U\right) \cdot \vec{v}_0 = U^3 \cdot \vec{v}_0$.

---

[4] Ein Vektor $\vec{v}$ heißt stochastisch, wenn seine Komponenten Zahlen aus dem Intervall [0 ; 1] sind mit Summe 1.

[5] In manchen Büchern werden bei einer stochastischen Matrix stochastische *Zeilen*vektoren statt *Spalten*vektoren gefordert. Dann muss die zugehörige Multiplikation umgekehrt erfolgen: (Zeilen-)Vektor mal Matrix statt Matrix mal (Spalten-)Vektor. Die zweite Form der Multiplikation ist in der Schule aber üblicher. Beide Begriffe (*stochastischer Vektor* bzw. *stochastische Matrix*) sind zwar für einen möglichen Unterricht in der Schule nicht unbedingt nötig, bieten sich aber hier an.

Wie kann man die Relevanz bzw. die Wichtigkeit einer Seite „messen"?

Natürlich ist eine Seite umso relevanter (wichtiger), je mehr Seiten auf diese Seite verweisen, insbesondere dann, wenn es sich bei den verweisenden Seiten selbst um wichtige handelt; denn dann nimmt man ja berechtigt an, dass auf dieser Seite gewisse tragende „Standards" bezüglich des Suchbegriffes gesetzt werden und dort also viel Wissenswertes zu finden ist. Welche ist nun die wichtigste Seite im obigen Graphen? Welche die zweitwichtigste, usw.? Wie soll man allgemein die Wichtigkeit einer Seite in einem gerichteten Graphen feststellen? Man kann sich dazu z. B. Folgendes vorstellen:
Sehr viele User nutzen dieses Netz (gerichteter Graph): Welcher Anteil davon wird sich – à la longue – im Zuge der Recherchen bei A, B, C, D aufhalten? Wenn sich herausstellen sollte, dass eine bestimmte Seite 90% der Suchenden auf sich zieht, so ist wohl klar, dass diese Seite am wichtigsten ist und in der Liste zuerst genannte werden sollte. Diese „langfristigen relativen Anteile" sind also eine Möglichkeit, die Relevanz (Wichtigkeit) einer Seite zu beschreiben, und dafür brauchen wir „Grenzverteilungen".

Angenommen die User beginnen zufällig auf den vier Seiten zu surfen und die „Startverteilung" der User beträgt für alle vier Seiten jeweils ¼: $\vec{v}_0 = \begin{pmatrix} 0,25 \\ 0,25 \\ 0,25 \\ 0,25 \end{pmatrix}$; wenn diese User dann zufällig weitersurfen, beträgt die Verteilung im nächsten Schritt: $\vec{v}_1 = U \cdot \vec{v}_0 = \begin{pmatrix} 0,25 \\ 0,25 \\ 0,375 \\ 0,125 \end{pmatrix}$ und im darauf folgenden dann $\vec{v}_2 = U \cdot \vec{v}_1 = U^2 \cdot \vec{v}_0 = \begin{pmatrix} 0,375 \\ 0,1875 \\ 0,3125 \\ 0,125 \end{pmatrix}$; die Seiten A und C scheinen hier also den Löwenanteil abzubekommen. Dies ist auch plausibel: alle Seiten haben einen Link zur Seite C und von dort aus muss man zur Seite A . . . . Indem man immer wieder von links mit der Matrix U multipliziert, erhält man die weiteren Verteilungen $\vec{v}_n = U^n \cdot \vec{v}_0$; diese streben zu einer „Grenzverteilung" $\vec{v}_n \to \vec{v}$, die sich hier als $\vec{v} = \begin{pmatrix} 3/9 \\ 2/9 \\ 3/9 \\ 1/9 \end{pmatrix}$ ergibt. Nach diesem Ergebnis müssten die Seiten A und C ex aequo auf Platz 1 gereiht werden, vor Seite B und D.

Solche Grenzverteilungen kann man im Unterricht auf mehrere Arten bestimmen:

1) Mit Tabellenkalkulation die Iteration so lange durchführen, bis sich die Werte nicht mehr ändern.
2) Mit einem CAS eine hohe Matrixpotenz von U bestimmen, so dass man mit $\vec{v}_n = U^n \cdot \vec{v}_0$ für große n wohl nahe der „Grenzverteilung" sein wird.
3) Gesucht ist dabei ein Vektor $\vec{v}$ mit Komponentensumme 1, der sich bei Multiplikation mit U nicht mehr ändert: $U \cdot \vec{v} = \vec{v}$. Man muss also ein lineares Gleichungssystem lösen, auch mit CAS.

Probleme, die dabei auftreten:

- Kann es mehrere solche Vektoren $\vec{v}$ (Grenzverteilungen, Lösungen) geben? Wenn es mehrere solche Vektoren gibt, streben dann die Verteilungsvektoren $\vec{v}_i$ manchmal (je nach Startverteilung $\vec{v}_0$) gegen den einen und manchmal gegen den anderen? Dies alles wäre natürlich sehr unangenehm, denn immerhin soll diese Grenzverteilung die Basis und Argumentationsgrundlage für das Ranking nach der Wichtigkeit sein. Eine nicht eindeutige und von der Startverteilung abhängige Grundlage wäre allerdings sehr zweifelhaft. Am besten wäre es, wenn dieser Vektor der Grenzverteilung eindeutig wäre und auch unabhängig von der Startverteilung $\vec{v}_0$.
- Alle drei oben angegebenen Möglichkeiten zur Berechnung der Grenzverteilung $\vec{v}$ funktionieren in einfacher Weise natürlich nur für relativ kleine Dimensionen, wie oben bei einer 4×4-Matrix, evtl. auch noch bei einer 20×20-Matrix, aber klarer Weise nur mehr schwerlich bei einer 1000000×1000000-Matrix. Im Fall von größeren Matrizen werden hier andere *iterative* Algorithmen zur *näherungsweisen* Lösung verwendet. Bei Google-Anwendungen können dies einige hunderttausend oder gar Millionen Seiten sein; außerdem kommen auf Google pro Sekunde sehr viele Anfragen zu, die alle prompt erledigt werden sollten.

Unabhängig davon, ob „Markoff-Ketten" als Begriff thematisiert werden, besonders wichtig in diesem Zusammenhang ist ein berühmter **Grenzwertsatz** (geht schon auf Markoff zurück), der eine einfache hinreichende Bedingung an die Übergangsmatrix *U* angibt, die garantiert, dass die Grenzverteilung $\vec{v}$ existiert, eindeutig und unabhängig von der Startverteilung $\vec{v}_0$ ist (ohne Beweis):

> *Wenn U stochastisch ist und U^n für irgendein $n \geq 1$ nur positive Elemente hat, dann streben die Matrixpotenzen U^n für $n \to \infty$ zu einer stochastischen Grenzmatrix G mit identischen Spalten*[6].

Es ist klar, dass diese Spalten dann den eindeutigen und vom Startvektor $\vec{v}_0$ unabhängigen Grenzvektor $\vec{v}$ angeben. Denn wegen $A_0 + B_0 + C_0 + D_0 = 1$ erhält man z. B. im Fall einer 4×4-Matrix für den Grenzvektor $\vec{v}$ mit dieser Grenzmatrix *G* (unabhängig von den konkreten Werten von A_0, B_0, C_0, D_0)

$$\vec{v} = \underbrace{\begin{pmatrix} u_1 & u_1 & u_1 & u_1 \\ u_2 & u_2 & u_2 & u_2 \\ u_3 & u_3 & u_3 & u_3 \\ u_4 & u_4 & u_4 & u_4 \end{pmatrix}}_{G} \cdot \underbrace{\begin{pmatrix} A_0 \\ B_0 \\ C_0 \\ D_0 \end{pmatrix}}_{\vec{v}_0} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{pmatrix}.$$

Natürlich gilt $\sum u_i = 1$, weil *G* stochastisch ist.

Bei unserem Beispiel hat zwar nicht *U* selbst, aber schon U^5 nur positive Elemente, so dass die Konvergenz und die Unabhängigkeit von der Startverteilung nach obigem Grenzwertsatz garantiert ist. Wir berechnen in unserem Fall z. B. U^{20} und erhalten mit einem CAS (4 Nachkommastellen):

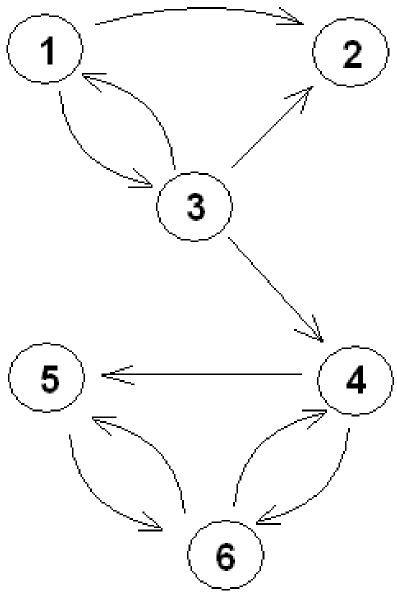
---

[6] D. h. die Zeileneinträge sind in jeder Zeile konstant. Dieser Satz braucht im Unterricht nicht bewiesen zu werden, man kann ihn einfach benutzen, um den PageRank-Algorithmus in seinen Grundzügen nachvollziehen zu können – siehe unten (bei Bedarf findet sich ein elementarer Beweis für den Spezialfall von 2×2-Matrizen z. B. in HUMENBERGER 2002a). Auch sonst braucht die zugehörige Theorie nicht breitgetreten zu werden.

$$U^{20} = \begin{pmatrix} 0{,}3333 & 0{,}3333 & 0{,}3333 & 0{,}3333 \\ 0{,}2222 & 0{,}2222 & 0{,}2222 & 0{,}2222 \\ 0{,}3333 & 0{,}3333 & 0{,}3333 & 0{,}3333 \\ 0{,}1111 & 0{,}1111 & 0{,}1111 & 0{,}1111 \end{pmatrix}.$$

Hier ist $\vec{v} = \begin{pmatrix} 3/9 \\ 2/9 \\ 3/9 \\ 1/9 \end{pmatrix}$, die schon oben angegebene Grenzverteilung, als Spalte gut abzulesen.

Ein etwas komplizierteres Beispiel



Ein immer noch sehr kleines Netzwerk aus 6 Internetseiten habe nebenstehende Verlinkungsstruktur.
Die Übergangsmatrix können wir wieder aus dem Übergangsgraphen ablesen:

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

Hier gibt es eine Situation, die oben noch nicht gegeben war: Von der Seite 2 gibt es offenbar keine weiterführenden Links, im Surfvorgang könnte man so eine Seite als „Senke“ bzw. „Sackgasse“ bezeichnen. Dies spiegelt sich in der zweiten Spalte der Matrix U wider, die nur Nullen enthält. Dies ist natürlich schlecht für unsere Zwecke (stochastische Matrix, Spaltensumme sollte 1 sein). Was wird man in so einer Situation beim Recherchieren mit Google praktisch machen?

Dafür gibt es mehrere Möglichkeiten:

a) Den Suchvorgang beenden und bei der Seite ② bleiben; dies würde in der Matrix bedeuten die zweite Null in der zweiten Spalte durch eine 1 zu ersetzen, im gerichteten Graphen käme dann bei ② ein Pfeil zu sich selbst dazu. Diese Möglichkeit wollen wir hier nicht wählen. Wenn man auch die 3. Modellannahme (siehe unten) berücksichtigt, könnte man in der zweiten Spalte die angesprochene 1 setzen, ohne dass man dadurch ein Ende des Surfprozesses erzwingt.

b) Man könnte mit dem Browser eine Seite zurück gehen und von dort andere Links benützen als den zu ② (die dann hoffentlich keine Sackgassen sind). Hier müsste man unterscheiden, von welcher Seite aus man zu ② gekommen ist, was die Sache relativ kompliziert machte.

c) Wir entscheiden uns für eine dritte Variante: Man verlässt diese Seite, kehrt zu einer „Liste“ zurück (gleichgültig, ob diese Liste von Google schon nach der Wichtigkeit gereiht wurde oder nicht) und klickt zufällig eine der vielen (anderen) Seiten an.

Dies formulieren wir noch mal explizit in der

2. Modellannahme: Wenn man beim Surfvorgang in einer Sackgasse (ohne weiterführenden Link) landet, so kehrt man zu einer Liste zurück und klickt nun eine der möglichen m Internetseiten an, und zwar alle mit derselben Wahrscheinlichkeit $1/m$.

Man sieht dabei auch davon ab, dass die in Rede stehende Seite selbst normalerweise nun wohl nicht mehr angeklickt wird; aber wenn dies sehr viele Seiten sind, ergibt sich dadurch kein großer Unterschied[7]: Man ersetzt die Einträge der zweiten Spalte (Nullen) nun jeweils durch 1/6 (allgemein: $1/m$, wenn es sich um m Webseiten, d. h. um einen gerichteten Graphen mit m Knoten, d. h. um eine $m \times m$-Matrix handelt). Statt einer reinen „Nullenspalte" fügt man also jeweils den m-dimensionalen Spaltenvektor

$\begin{pmatrix} 1/m \\ \vdots \\ 1/m \end{pmatrix}$ ein und erhält: $$U_1 = \begin{pmatrix} 0 & 1/6 & 1/3 & 0 & 0 & 0 \\ 1/2 & 1/6 & 1/3 & 0 & 0 & 0 \\ 1/2 & 1/6 & 0 & 0 & 0 & 0 \\ 0 & 1/6 & 1/3 & 0 & 0 & 1/2 \\ 0 & 1/6 & 0 & 1/2 & 0 & 1/2 \\ 0 & 1/6 & 0 & 1/2 & 1 & 0 \end{pmatrix}.$$

So kann man also auch dann eine stochastische Übergangsmatrix U_1 erreichen, wenn es in der Netzstruktur Sackgassen gibt.

Frage: Was wäre, wenn es eine Seite gäbe, auf die kein Link verweist (nur Links von ihr weg)? Wäre dies ähnlich schlimm wie eine Sackgasse? [Nein, das entspräche einer Nullen*zeile* in der Übergangsmatrix, und das wäre nicht so schlimm, denn dann könnte die Matrix trotzdem stochastisch sein.]

3. Modellannahme: Dadurch auf den Plan gerufen kann man sagen: Auch wenn die Seite keine Sackgasse ist, kann es doch vorkommen, dass jemand nicht den Links auf dieser Seite folgt, sondern eben zu einer Liste zurückkehrt und eine andere Seite einfach anklickt. Mit Wahrscheinlichkeit α mögen die User irgendwelchen Links auf der jeweiligen Seite folgen, mit Wahrscheinlichkeit $1-\alpha$ zur Liste zurückkehren und neu einsteigen, d. h. eine beliebige andere Seite (mit Wahrscheinlichkeit $1/m$) anklicken. Wie kann man nun dieses Szenario mathematisch beschreiben? Wie sieht die dann zugehörige, neue Übergangsmatrix T aus?

Wenn man den Links auf der Seite folgt, ist die Übergangsmatrix durch U_1 gegeben.

Wie muss die Übergangsmatrix im Falle des Neueinstiegs lauten (Schüler(innen)aufgabe)? Beim *Neueinsteigen* muss die nächste Verteilung $(1/m, \ldots, 1/m)^t$ sein, d. h. die Übergangsmatrix muss in diesem Fall

$$U_2 = \begin{pmatrix} 1/m & \cdots & 1/m \\ \vdots & & \vdots \\ 1/m & \cdots & 1/m \end{pmatrix} \text{ lauten, denn: } \begin{pmatrix} 1/m & \cdots & 1/m \\ \vdots & & \vdots \\ 1/m & \cdots & 1/m \end{pmatrix} \cdot \underbrace{\begin{pmatrix} v_1 \\ \vdots \\ v_m \end{pmatrix}}_{\sum v_i = 1} = \begin{pmatrix} 1/m \\ \vdots \\ 1/m \end{pmatrix}.$$

Insgesamt ergibt sich also für die neue Übergangsmatrix T durch Gewichten der beiden Fälle bzw. Übergangsmatrizen mit den Faktoren α bzw. $1-\alpha$:

[7] Außerdem kann es auch wirklich passieren, dass man eine Seite ein zweites Mal anklickt, obwohl man diese schon gesehen hat und eigentlich nicht mehr darauf wollte.

$$\boxed{T = \underbrace{\alpha \cdot U_1}_{\text{Mit W' } \alpha \text{ den Links folgen}} + \underbrace{(1-\alpha)\cdot U_2}_{\text{Mit W' } (1-\alpha) \text{ neu einsteigen}}} \qquad (1)$$

Es ist sehr leicht einzusehen (Schüleraufgabe): Weil U_1 und U_2 stochastische Matrizen sind, ist auch T stochastisch.

Eine entscheidende Eigenschaft von *T*:
Die Matrix T hat *nur positive Einträge*, keine Nullen mehr. Nach obigem Grenzwertsatz (Markoff) liegt mit der Übergangsmatrix T also sicher jene gewünschte und besonders einfache Situation vor, in der es eine eindeutige und von der Startverteilung unabhängige Grenzverteilung gibt. Diese Grenzverteilung kann dann die gewünschte Reihung der Seiten angeben, ihre Wichtigkeit (Relevanz) messen (→ „PageRank“).

Wie groß soll der Wert von α gewählt werden? Es ist bekannt, dass Google lange Zeit $\alpha = 0{,}85$ gewählt hat. Möglicherweise ist Google aber in der Zwischenzeit von diesem Wert abgewichen. Für obiges Beispiel ergibt sich für das lineare Gleichungssystem $T \cdot \vec{v} = \vec{v}$

($\alpha = 0{,}85$; $\vec{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_6 \end{pmatrix}$; $\underbrace{v_1 + \ldots + v_6}_{v_i \geq 0} = 1$; CAS; 4 Nachkommastellen): $\begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{pmatrix} = \begin{pmatrix} 0{,}0517 \\ 0{,}0737 \\ 0{,}0574 \\ 0{,}1999 \\ 0{,}2686 \\ 0{,}3487 \end{pmatrix}$

Dieses ergibt sich auch, wenn man eine hohe Matrixpotenz von T berechnet und eine Spalte davon nimmt. Auch mit Tabellenkalkulation könnte man zu diesem Ergebnis kommen. Nach diesem Ergebnis wären also die Seiten in absteigender Reihenfolge ihrer Wichtigkeit:

Seite 6 → Seite 5 → Seite 4 → Seite 2 → Seite 3 → Seite 1.

Bemerkung: Man muss bei der Konstruktion von T nicht unbedingt U_1 verwenden, man könnte auch an der entsprechenden Stelle in der Nullenspalte von U eine 1 setzen (dies entspräche der obigen Möglichkeit **a**). Das echte Ende des Prozesses wird ja ohnehin durch die Matrix U_2 verhindert, wenn man den „Neueinstieg“ eben jederzeit zulässt.

Explizite Lösung (Formel)
Durch Einsetzen von (1) in $T \cdot \vec{v} = \vec{v}$ und Verwenden der Matrixschreibweise kann diese Gleichung noch ein wenig umgeschrieben werden (I bezeichnet dabei die *m*-dimensionale Einheitsmatrix), so dass sich sogar eine explizite Formel für $\vec{v}$ ergibt:

$$\alpha \cdot U_1 \cdot \vec{v} + (1-\alpha) \cdot \underbrace{U_2 \cdot \vec{v}}_{=(1/m,\ldots,1/m)^t} = I \cdot \vec{v} \quad \Rightarrow \quad (\alpha \cdot U_1 - I) \cdot \vec{v} = (\alpha - 1) \cdot \begin{pmatrix} 1/m \\ \vdots \\ 1/m \end{pmatrix} \quad \Rightarrow$$

$$\boxed{\vec{v} = (\alpha - 1) \cdot (\alpha \cdot U_1 - I)^{-1} \cdot \begin{pmatrix} 1/m \\ \vdots \\ 1/m \end{pmatrix}}$$

Man kann zeigen, dass die Matrix $\alpha \cdot U_1 - I$ „nichtsingulär" ist, so dass sich für $\vec{v}$ dabei immer eine eindeutige Lösung ergibt[8].

Diese explizite Lösung ist aber für die Google-Praxis ungeeignet. Dort sind es ja sehr große lineare Gleichungssysteme (z. B. $m = 1000000$ oder mehr) und dabei kann leider nicht mit der expliziten Lösung („Formel") gerechnet werden, denn das Invertieren einer Matrix kann man nur bei relativ kleinem *m* in vernünftiger Zeit bewerkstelligen (noch dazu gibt es ja sehr viele Anfragen an Google pro Sekunde). Man kommt in der Praxis vielmehr *iterativ* zu einer (*Näherungs*-)*Lösung*, wobei die gesuchten Werte v_i in diesem Zusammenhang die Parameter für die Wichtigkeit einer Seite, die so genannten PageRank-Werte sind.

Diese Probleme sollen hier aber gar nicht im Vordergrund stehen. Es geht uns hier primär um die elementaren Überlegungen, die zur (stochastischen) Übergangsmatrix *T* mit nur positiven Einträgen führen, so dass es nach dem obigen Grenzwertsatz sicher eine von der Startverteilung unabhängige Grenzverteilung gibt. Diese Ideen sind gleichermaßen einfach und genial! Sie garantieren, dass das Verfahren „immer funktioniert".

Drei einfache Modellannahmen (siehe oben) haben hier eine enorme Wirkung. Natürlich kann dieses Modellieren nicht selbständiges Modellieren durch die Schülerinnen und Schüler sein, sie lernen dadurch aber ein ganz aktuelles Stück „angewandter Mathematik" kennen.
In vielen Fällen kann man das so genannte Relevanzparadoxon der Mathematik beobachten: Mathematik wird in der *modernen Gesellschaft* zunehmend wichtiger, denn unser Leben wird immer mehr durchsetzt von „Dingen", die sehr viel Mathematik enthalten (Handies, Internet, elektronischer Zahlungsverkehr, Computer, Versicherungen, CD-player, Netzwerke, . . . man könnte diese Liste nach Belieben fortsetzen!). Mathematik ist damit sicher eine so genannte Schlüsseltechnologie für unsere Zukunft. Aber in den meisten Fällen ist die Mathematik dahinter sehr komplex und kann von Nichtmathematiker(inne)n nicht verstanden werden. Und für den normalen Gebrauch dieser Dinge ist das ja auch nicht notwendig. Das bedeutet in gewisser Weise, dass Mathematik für das Individuum immer weniger wichtig wird. Das ist ein Grund, warum viele Leute die Wichtigkeit von Mathematik überhaupt nicht sehen. Daher sollten wir im Mathematikunterricht dem entgegenwirken, indem möglichst viele Beispiele thematisiert werden, die die Wichtigkeit von Mathematik in der modernen Gesellschaft auf elementare und „schlagende" Weise zeigen. Das hier behandelte ist m. E. jedenfalls ein solches.

In der Literatur findet man häufig auch folgende Definition des PageRank-Algorithmus als Beispiel einer *rekursiven Definition*.

Z. B. steht bei Wikipedia: „Das Prinzip des PageRank-Algorithmus ist, dass jede Seite ein Gewicht (PageRank) besitzt, das umso größer ist, je mehr Seiten (mit möglichst hohem eigenen Gewicht) auf diese Seite verweisen. Das Gewicht PR_i einer Seite *i* berechnet sich also aus den Gewichten PR_j der auf *i* verlinkenden Seiten *j*. Verlinkt *j* auf insgesamt C_j verschiedene Seiten, so wird das Gewicht von PR_j anteilig auf diese Seiten aufgeteilt. Folgende rekursive Formel kann als Definition des PageRank-Algorithmus angesehen werden:

$$PR_i = \frac{1-d}{N} + d \sum_{j:\, j \to i} \frac{PR_j}{C_j}$$

[8] Dies folgt aber auch aus obigem Grenzwertsatz (Markoff).

Dabei ist N die Gesamtanzahl der Seiten und d ein Dämpfungsfaktor zwischen 0 und 1, mit dem ein kleiner Anteil des Gewichts $(1 - d)$ einer jeden Seite abgezogen und gleichmäßig auf alle vom Algorithmus erfassten Seiten verteilt wird. Dies ist notwendig, damit das Gewicht nicht zu Seiten ‚abfließt', die auf keine andere Seite verweisen."

In wiefern entspricht diese Definition unseren Darstellungen? Wenn wir in $T \cdot \vec{v} = \vec{v}$ die Beziehung (1) einsetzen, so erhalten wir – siehe oben: $\vec{v} = (1-\alpha)\cdot(1/m,\ldots,1/m)^t + \alpha \cdot U_1 \cdot \vec{v}$. Nun betrachten wir die i-te Komponente davon (die Komponenten v_i sind ja in der neuen Notation PR_i):

$v_i = \frac{1-\alpha}{m} + \alpha \cdot \sum_j \mathrm{P}(j \to i) \cdot v_j$, wobei die $\mathrm{P}(j \to i)$ die Einträge in der Übergangsmatrix U_1 in Zeile i sind, also die Wahrscheinlichkeiten, mit der ein User von Seite j zu Seite i kommt, wenn er den Links folgt. Wenn es keinen Link von Seite j zu Seite i gibt, so ist $\mathrm{P}(j \to i) = 0$, ansonsten ist $\mathrm{P}(j \to i)$ genau $1/C_j$, wobei C_j die Anzahl der von j ausgehenden Links ist. Damit ist die Entsprechung mit $d = \alpha$ und $N = m$ geklärt.

Die Kriterien, nach denen Google die Reihung der Seiten vornimmt, sind mittlerweile sehr vielfältig. Das erste davon war der PageRank, seither sind viele weitere Kriterien dazu gekommen, beispielsweise wird heutzutage der Seitentitel und der Text des Hyperlinks miteinbezogen. Derzeit bezieht Google etwa 200 solcher Hinweise in eine Suche mit ein, um dem User möglichst das zu liefern, was er tatsächlich sucht.

Literatur

1. BÜCHTER, A. u. H.-W. HENN ([2]2007): Elementare Stochastik. Springer, Berlin-Heidelberg.
2. CHARTIER, T. P. (2006): Googling Markov. In: The UMAP Journal **27**, 1, 17 – 30.
3. HUMENBERGER, H. (2002a): Der PALIO – das Pferderennen von Siena als Ausgangspunkt für Modelle von Auswahlprozessen und als Einstieg zum Thema Markoff-Ketten. In: Stochastik in der Schule **22**, 2, 2 – 13.
   Online: http://stochastik-in-der-schule.de/sisonline/struktur/jahrgang22-2002/heft2/2002-2_Humen.pdf (27. 04. 2011).
4. HUMENBERGER, H. (2002b): Der PALIO – das Pferderennen von Siena. In: mathematik lehren 113 (August 2002), 58 – 62.
5. LAMBACHER-SCHWEIZER (2001): Lineare Algebra mit analytischer Geometrie, Leistungskurs. Klett, Stuttgart.
6. LANGVILLE, A. N. (2006): Google's PageRank and Beyond: The Science of Search Engine Rankings. Princeton University Press, Princeton.
7. MEYER, D. (1998): Markoff-Ketten. In: Mathematik in der Schule **36**, 12, 661 – 670 und 675 – 680.
8. Spiegel: Rafaela v. BREDOW u. a.: Ende der Privatheit. Spiegel 2/2010, 58 – 69.
9. Stern: http://www.stern.de/wirtschaft/unternehmen/:Das-Google-Duo-Die-Internetstars-Page-Brin-/523515.html (27. 04. 2011).
10. WILLS, R. S. (2006): Google's PageRank: The Math Behind the Search Engine. In: The Mathematical Intelligencer **28**, 4, 6 – 11.
11. WIRTHS, H. (1997): Markow-Ketten – Brücke zwischen Analysis, linearer Algebra und Stochastik. In: Mathematik in der Schule **35**, 11, 601 – 606 und 611 – 613.

Anschrift des Verfassers: Hans HUMENBERGER
Universität Wien, Fakultät für Mathematik, Nordbergstraße 15 (UZA 4), A – 1090 Wien.
Mail: hans.humenberger@univie.ac.at