Proseminar Technomathematik
WS 2007/2008

Berechnung des PageRank mithilfe von linearen Gleichungssystemen

(Google-Matrizen II)

Philipp Cordes

12.02.2008

Inhaltsverzeichnis

Zusammenfassung

Die Ermittlung des PageRank wird traditionell als Eigenwertproblem gesehen, das mithilfe der Potenzmethode gelöst wird. Ein Eigenwertproblem ist jedoch ein lineares Gleichungssystem, das sich auch auf andere Weise lösen lässt. Hier werden einige Eigenschaften dieses Gleichungssystems gezeigt und ein einfacher Algorithums vorgestellt, mit dem das System schnell gelöst werden kann.

Außerdem wird anhand des Gleichungssystems die Stabilität der Lösung durch Ableitung nach α ermittelt.

Kapitel 1

Konventionen und Einführendes

Diese Ausarbeitung baut auf die Arbeit von Christian Abramowski [1] auf, dort wird insbesondere auf den Ursprung der hier verwendeten Matrizen P, $\bar{\mathrm{P}}$ und $\bar{\bar{\mathrm{P}}}$ eingegangen. Daher ist zumindest ein Überblick darüber unbedingt erforderlich, um zu verstehen, was hier gemacht wird.

Ausgehend von der in [1] hergeleiteten Gleichung des Potenzverfahrens, werden hier folgende Bezeichnungen verwendet, weil sie Konvention in der Hauptquelle [3] sind:

$$\mathrm{P} := \mathrm{Q}^T, \qquad \bar{\mathrm{P}} := \bar{\mathrm{Q}}^T, \qquad \bar{\bar{\mathrm{P}}} := \bar{\bar{\mathrm{Q}}}^T$$

Außerdem wird die Einheitsmatrix mit $\mathbf{I}$, die Transponierte einer Matrix A mit $A^{\boldsymbol{T}}$ und der Vektor $(1, 1, \ldots, 1)^T$ mit $\mathbf{e}$ bezeichnet.

Damit ergeben sich folgende Gleichungen, die im ganzen Dokument benutzt werden:

$$\begin{aligned}
\bar{\mathrm{P}} &= \mathrm{P} + a\mathrm{v}^T, & \mathrm{v}^T\mathrm{e} &= 1, & \bar{\mathrm{P}}\mathrm{e} &= \mathrm{e}, \\
\bar{\bar{\mathrm{P}}} &= \alpha\bar{\mathrm{P}} + (1-\alpha)\mathrm{e}\mathrm{v}^T, & \pi^T\bar{\bar{\mathrm{P}}} &= \pi^T, & \bar{\bar{\mathrm{P}}}\mathrm{e} &= \mathrm{e}, \\
& & \pi^T\mathrm{e} &= 1 & &
\end{aligned}$$

1.1 Definitionen und Hilfssätze

Außerdem führe ich noch folgende Begriffe ein, die in den Vorlesungen Lineare Algebra und Analysis in dieser Form nicht behandelt wurden:

1. Eine reelle Matrix A heißt **nichtnegativ**, wenn jeder ihrer Einträge nichtnegativ ist:

   $$A \geq 0 :\Longleftrightarrow \forall i, j : a_{i,j} \geq 0$$

   (Analog: positiv, negativ, nichtpositiv.)

2. Der **Spektralradius** ρ einer Matrix A ist der größte Betrag der Eigenwerte der Matrix A:

   $$\rho(A) := \max_{1 \leq i \leq n} |\lambda_i(A)|$$

3. Die ∞**-Norm** einer Matrix A ist die größte absolute Zeilensumme der Einträge von A:

   $$||A||_\infty := \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}|$$

   U.a. gilt damit: $||A||_\infty \geq \rho(A)$.

Lemma 1 (Neumann-Reihe)**.** *Für eine $n \times n$-Matrix A mit $||A||_\infty < 1$ (oder $\rho(A) < 1$) gilt, falls $(\mathrm{I} - A)^{-1}$ existiert:*

$$(\mathrm{I} - A)^{-1} = \sum_{k=0}^{\infty} A^k$$

Beweis. Wegen $||A||_\infty < 1$ (oder $\rho(A) < 1$) konvergiert die Summe. Multiplikation der Gleichung mit $(\mathrm{I} - A)$ ergibt:

$$\begin{aligned}
\mathrm{I} &= (\mathrm{I} - A)^{-1}(\mathrm{I} - A) = (\sum_{k=0}^{\infty} A^k)(\mathrm{I} - A) \\
&= \sum_{k=0}^{\infty} A^k(\mathrm{I} - A) = \sum_{k=0}^{\infty}(A^k - A^{k+1}) \\
&= A^0 \underbrace{-A^1 + A^1}_{=0} \underbrace{-A^2 + A^2}_{=0} \underbrace{-A^3 + \dots}_{=0} \\
&= \mathrm{I}
\end{aligned}$$

$\square$

Lemma 2. *Für eine $n \times n$-Matrix B und $r \in \mathbb{R}$ mit $\rho(B) < r$ gilt:*

$$(r\mathrm{I} - B)^{-1} = \frac{1}{r} \sum_{k=0}^{\infty} (\frac{1}{r} B)^k$$

Beweis. Lemma 1 mit $A := \frac{1}{r}B$ (sodass $\rho(A) < 1$) ergibt:

$$(\mathrm{I} - \frac{1}{r}B)^{-1} = \sum_{k=0}^{\infty} (\frac{1}{r}B)^k$$

$r\mathrm{I} - B = r(\mathrm{I} - \frac{1}{r}B)$, also $(r\mathrm{I} - B)^{-1} = \frac{1}{r}(\mathrm{I} - \frac{1}{r}B)^{-1}$ $\square$

Definition 3 (aus [4])**.** *Eine Matrix $A \in \mathrm{GL}_n(\mathbb{R})$ heißt* ***M-Matrix****,*
wenn $\forall i \neq j : a_{ij} \leq 0$ und $\mathrm{A}^{-1} \geq 0$.

Satz 4. *A M-Matrix $\iff \exists B \geq 0, r > \rho(B) : A = r\mathrm{I} - B$*

Beweis. „$\Rightarrow$“: Sei also A M-Matrix, $r := \max\limits_{1 \leq i \leq n} |a_{ii}|$.
$\Rightarrow B := r\mathrm{I} - A \geq 0$ und $A^{-1} = (r\mathrm{I} - B)^{-1}$ (nach Lemma 2)
„$\Leftarrow$“: Sei also $A = r\mathrm{I} - B$ mit $B \geq 0$ und $r > \rho(B)$.
$\Rightarrow \exists A^{-1}, A^{-1} \geq 0, a_{ij} \leq 0 \forall i \neq j$
$\Rightarrow A$ M-Matrix. $\square$

Kapitel 2

Das lineare Gleichungssystem

2.1 Herleitung des LGS

Zuerst wird die Gleichung des Potenzverfahrens in das dazugehörige lineare Gleichungssystem umgeformt:

$$\begin{aligned}
\pi^T &= \pi^T(\alpha\bar{\mathrm{P}} + (1-\alpha)\mathrm{ev}^T) \\
\pi^T &= \pi^T\alpha\bar{\mathrm{P}} + \pi^T(1-\alpha)\mathrm{ev}^T \\
\pi^T\mathrm{I} - \pi^T\alpha\bar{\mathrm{P}} &= (1-\alpha)\underbrace{\pi^T\mathrm{e}}_{=1}\mathrm{v}^T \\
\pi^T(\mathrm{I} - \alpha\bar{\mathrm{P}}) &= (1-\alpha)\mathrm{v}^T
\end{aligned}$$

Im Folgenden wird die Koeffizientenmatrix $\mathrm{I} - \alpha\bar{\mathrm{P}}$ untersucht.

2.1.1 Eigenschaften der Koeffizientenmatrix

Die Koeffizientenmatrix $\mathbf{A} := \mathrm{I} - \alpha\bar{\mathrm{P}}$ hat folgende Eigenschaften:

1. A ist M-Matrix (nach Satz 4).

2. A regulär (nach Definition 3, M-Matrix).

3. Die Zeilensummen von A sind $1 - \alpha$.

$$\mathrm{Ae} = (\mathrm{I} - \alpha\bar{\mathrm{P}})\mathrm{e} = \mathrm{Ie} - \alpha\underbrace{\bar{\mathrm{P}}\mathrm{e}}_{=\mathrm{e}} = (1-\alpha)\mathrm{e}$$

4. $||\mathrm{A}||_\infty = 1 + \alpha$, falls es mindestens einen Nicht-Endknoten gibt ($\mathrm{a} \neq e$):

   Die Diagonaleinträge von $\bar{\mathrm{P}}$ sind 0, deswegen sind die Diagonaleinträge von A gleich 1. In jeder Zeile von A steht sonst in Summe $-\alpha$. Der Absolutwert ist deswegen $1 + \alpha$.

5. $\mathrm{A}^{-1} \geq 0$, weil A M-Matrix ist.

6. Die Zeilensummen von A^{-1} sind $\frac{1}{1-\alpha}$:

$$\begin{aligned}
\mathrm{A}^{-1}\mathrm{e} &= x && \mid (1-\alpha)\mathrm{A}\cdot \\
(1-\alpha)\mathrm{e} &= (1-\alpha)\mathrm{A}x \\
&= \mathrm{A}(1-\alpha)x
\end{aligned}$$

Nach 3. ist damit $(1-\alpha)x = \mathrm{e}$, also $x = \frac{1}{1-\alpha}\mathrm{e}$.

Deswegen ist $||\mathrm{A}^{-1}||_\infty = \frac{1}{1-\alpha}$.

7. Die Konditionszahl $\kappa_\infty(\mathrm{A}) := ||\mathrm{A}||_\infty ||\mathrm{A}^{-1}||_\infty$ von A ist $\frac{1+\alpha}{1-\alpha}$, nach den Punkten 4 und 6. Was eine Konditionszahl ist und warum sie berechnet wird, wird noch ausführlich in Abschnitt 3 ab Seite 7 behandelt.

2.2 Übertragen auf P

Wie bei der Berechnung des Eigenwertes, ist es auch hier möglich, anstelle der Matrix $\bar{\mathrm{P}}$ mit der Matrix P zu arbeiten:

$$\begin{aligned}
\pi^T(\mathrm{I} - \alpha\bar{\mathrm{P}}) &= (1-\alpha)\mathrm{v}^T \\
\pi^T(\mathrm{I} - \alpha\mathrm{P} - \alpha\mathrm{a}\mathrm{v}^T) &= (1-\alpha)\mathrm{v}^T \\
\pi^T(\mathrm{I} - \alpha\mathrm{P}) &= (1-\alpha+\alpha\underbrace{\pi^T\mathrm{a}}_{=:\gamma})\mathrm{v}^T \\
\pi^T(\mathrm{I} - \alpha\mathrm{P}) &= \mathrm{v}^T
\end{aligned}$$

Da γ quasi einen beliebigen Wert annehmen kann (die Summe der Komponenten von π^T, die zu einem Endknoten gehören) lässt sich nach [3, Seite 13] für γ ein beliebiger Wert wählen, für diesen π^T berechnen und dann bei Normierung $\pi^T\mathrm{e} = 1$ den gleichen Wert den gleichen Wert erhalten. Als Wahl bietet sich daher $\gamma = 1$ an.

2.2.1 Eigenschaften des vereinfachten Systems

Die Eigenschaften 1. bis 5. lassen sich direkt übertragen, bei den anderen Eigenschaften ändert sich Folgendes:

6. Die Zeilensumme in $(\mathrm{I}-\alpha\mathrm{P})^{-1}$ ist 1 für Endknoten (von I) und kleiner oder gleich $\frac{1}{1-\alpha}$ für die anderen Knoten.
7. Die Konditionszahl $\kappa_\infty(\mathrm{I}-\alpha\mathrm{P})$ ist deswegen kleiner oder gleich $\frac{1+\alpha}{1-\alpha}$, also potentiell besser.
8. Wenn i einen Endknoten repräsentiert, steht in der i-ten Zeile der i-te Einheitsvektor e_i.

Auswirkungen auf die Lösung des Gleichungssystem

Durch Eigenschaft 8. lässt sich das Gleichungssystem

$$\pi^T(\mathrm{I}-\alpha\mathrm{P}) = \mathrm{v}^T$$

durch Umsortieren der Koeffizientenmatrix nach Endknoten und Nicht-Endknoten vereinfachen:

$$\mathrm{P} = \begin{array}{cc} & \begin{array}{cc} \mathsf{NE} & \mathsf{E} \end{array} \\ \begin{array}{c} \mathsf{NE} \\ \mathsf{E} \end{array} & \begin{pmatrix} \mathrm{P}_{11} & \mathrm{P}_{12} \\ 0 & 0 \end{pmatrix} \end{array}$$

$$\begin{aligned}
\Rightarrow \mathrm{I}-\alpha\mathrm{P} &= \begin{pmatrix} \mathrm{I}-\alpha\mathrm{P}_{11} & -\alpha\mathrm{P}_{12} \\ 0 & \mathrm{I} \end{pmatrix} \\
\Rightarrow (\mathrm{I}-\alpha\mathrm{P})^{-1} &= \begin{pmatrix} (\mathrm{I}-\alpha\mathrm{P}_{11})^{-1} & \alpha(\mathrm{I}-\alpha\mathrm{P}_{11})^{-1}\mathrm{P}_{12} \\ 0 & \mathrm{I} \end{pmatrix} \\
\Rightarrow \pi^T &= (\mathrm{I}-\alpha\mathrm{P})^{-1}\mathrm{v}^T \\
&= \left(\mathrm{v}_1^T(\mathrm{I}-\alpha\mathrm{P}_{11}) \;\middle|\; \alpha\mathrm{v}_1^T(\mathrm{I}-\alpha\mathrm{P}_{11})^{-1}\mathrm{P}_{12} + v_2^T \right),
\end{aligned}$$

wenn π^T und v^T in Nichtendknoten- (π_1^T bzw. v_1^T) und Endknotenanteil (π_2^T bzw. v_2^T) zerlegt werden.
$I - \alpha P_{11}$ erbt viele Eigenschaften von $I - \alpha P$, insbesondere sind beide Matrizen regulär.

Der Algorithmus, der sich daraus ableitet, ist also:

1. π_1^T in $\pi_1^T(I - \alpha P_{11}) = v_1^T$ berechnen.
2. $\pi_2^T = \alpha \pi_1^T P_{12} + v_2^T$ berechnen.
3. $\pi^T = \frac{1}{||\pi_1^T \; \pi_2^T||_1}[\pi_1^T \; \pi_2^T]$ normieren.

Damit lässt sich das lineare Gleichungssystem der „California“-Beispielmatrix [2] von der Größe 9664×9664 auf 5132×5132 für den ersten Schritt reduzieren, bei Verbesserung des Verfahrens (laut Meyer [3, Seite 15]) sogar auf eine Größe von 2622×2622 (siehe Abbildung 2.1).

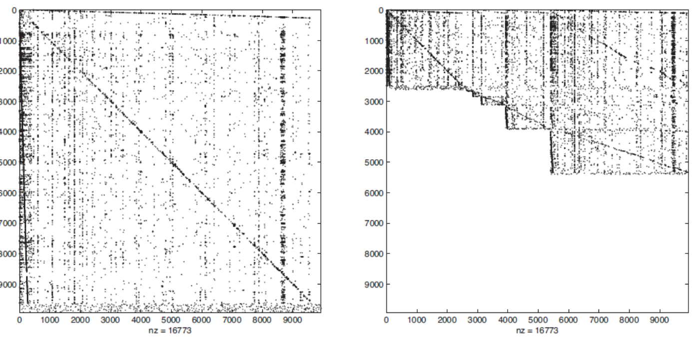

Abbildung 2.1: Hyperlinks aus der California-Matrix: links unsortiert, rechts sortiert.

Kapitel 3

Stabilität der Lösung

Im vorherigen Kapitel wurde im Rahmen der Eigenschaften der Koeffizientenmatrizen schon auf die Konditionszahl eingegangen, die als $\kappa(A) = ||A||_\infty ||A^{-1}||_\infty$ für eine reguläre Matrix A definiert ist. Die Konditionszahl $\kappa(A)$ ist der Faktor, um den eine Veränderung an A höchstens verstärkt wird. Die Kondition bezieht sich damit auf Änderungen an den Eingangsdaten (also der Hyperlinkstruktur).

Die Änderungen an der Matrix A, die im Folgenden betrachtet werden, sind Änderungen von α. Durch Wahl von α in $]0,1[$ lässt sich die „Teleportations"wahrscheinlichkeit, also die Wahrscheinlichkeit, dass der „random surfer" zu einer zufälligen Seite wechselt, modellieren. Je nach Wahl von α ergibt sich so ein PageRank, der sich stärker an der tatsächlichen Hyperlinkstruktur orientiert oder einer, der den Personalisierungsvektor v^T stärker berücksichtigt. Dass α damit maßgeblichen Einfluss auf die Konvergenz der Potenzmethode hat, wird in [1] behandelt; hier geht es um die Stabilität des PageRank für verschiedene Werte von α, also eine Änderung am Algorithmus.
Dazu wird der PageRank-Vektor π^T auf Differenzierbarkeit nach α untersucht, abgeleitet und die Ableitung schließlich abgeschätzt.

3.1 Differenzierbarkeit

Satz 5 (Satz von Perron-Frobenius, aus [4]). *Für $A \in \mathbb{R}^{n\times n} \geq 0$ irreduzibel gilt:*

(1) $\lambda = \rho(A)$ ist Eigenwert von A und $\lambda > 0$.

(2) Die algebraische Vielfachheit von λ ist 1.

(3) $\exists x > 0$ mit $Ax = \lambda x$.

(4) $\exists! x > 0$ mit $Ax = \lambda x$, $||x||_1 = 1$, alle anderen Eigenvektoren haben mindestens eine negative Komponente.

Satz 6. $\pi^T(\alpha) = \frac{1}{\sum_{i=1}^{n} \mathrm{D}_i(\alpha)}(\mathrm{D}_1(\alpha), \ldots, \mathrm{D}_n(\alpha))$ *mit $\mathrm{D}_i(\alpha)$ i-ter Hauptminorendeterminante ($= \det(\mathrm{I} - \bar{\bar{\mathrm{P}}})_{ii}$), also die Diagonalelemente von $\mathrm{adj}(\mathrm{I} - \bar{\bar{\mathrm{P}}}(\alpha))$.*
Damit besteht π^T aus Summen und Produkten aus $\bar{\bar{\mathrm{P}}}(\alpha)$ und ist damit auf $]0,1[$ nach α differenzierbar.

Aus Übersichtlichkeitsgründen werden folgende Abkürzungen eingeführt:

$$\mathrm{P} := \bar{\bar{\mathrm{P}}}(\alpha), \quad \pi^T := \pi^T(\alpha), \quad \mathrm{D}_i := \mathrm{D}_i(\alpha), \quad \mathrm{A} := \mathrm{I} - \mathrm{P}$$

Beweis. Perron-Frobenius sagt: π^T ist einziger Eigenvektor von $\bar{\bar{\mathrm{P}}}$ zum Eigenwert 1. Deswegen ist $\mathrm{rg}\,\mathrm{A} = n - 1$; außerdem sind π^T und e Eigenvektoren von A zum Eigenwert 0.

Da $\mathrm{A} \cdot \operatorname{adj} \mathrm{A} = \operatorname{adj} \mathrm{A} \cdot \mathrm{A} = \det \mathrm{A} \cdot \mathrm{I} = 0_{n \times n}$, sind die Zeilen von $\operatorname{adj} \mathrm{A}$ Vielfache von π^T und die Spalten Vielfache von e, also

$$\operatorname{adj} \mathrm{A} = k \cdot \mathrm{e}\pi^T \text{für ein } k \in \mathbb{R}, k \neq 0.$$

Aber $[\operatorname{adj} \mathrm{A}]_{ii} = \mathrm{D}_i$, damit sind die $\pi_i^T = \mathrm{D}_i/k$ Vielfache von D_i und k ergibt sich aus der Normierung $\pi^T e = 1$.

□

3.2 Berechnen der Ableitung

Zur Erinnerung:

$$\pi^T \mathrm{e} = 1 \quad \Rightarrow \quad \frac{d\pi^T(\alpha)}{d\alpha}\mathrm{e} = 0$$

$$\begin{aligned}
\pi^T(\alpha) &= \pi^T(\alpha) \cdot (\alpha\bar{\mathrm{P}} + (1-\alpha)\mathrm{e}\mathrm{v}^T) \qquad \Big| \frac{d}{d\alpha} \\
\frac{d\pi^T(\alpha)}{d\alpha} &= \frac{d\pi^T(\alpha)}{d\alpha} \cdot (\alpha\bar{\mathrm{P}} + (1-\alpha)\mathrm{e}\mathrm{v}^T) + \pi^T(\alpha) \cdot (\bar{\mathrm{P}} - \mathrm{e}\mathrm{v}^T) \\
&= \frac{d\pi^T(\alpha)}{d\alpha}\alpha\bar{\mathrm{P}} + \underbrace{(1-\alpha)\frac{d\pi^T(\alpha)}{d\alpha}\mathrm{e}\mathrm{v}^T}_{=0} + \pi^T(\alpha) \cdot (\bar{\mathrm{P}} - \mathrm{e}\mathrm{v}^T) \\
\frac{d\pi^T(\alpha)}{d\alpha}(\mathrm{I} - \alpha\bar{\mathrm{P}}) &= \pi^T(\alpha)(\bar{\mathrm{P}} - \mathrm{e}\mathrm{v}^T) \\
\frac{d\pi^T(\alpha)}{d\alpha} &= \pi^T(\alpha)(\bar{\mathrm{P}} - \mathrm{e}\mathrm{v}^T)(\mathrm{I} - \alpha\bar{\mathrm{P}})^{-1}
\end{aligned}$$

Komponentenweise ist es dann

$$\frac{d\pi_j^T(\alpha)}{d\alpha} = \pi^T(\alpha)(\bar{\mathrm{P}} - \mathrm{e}\mathrm{v}^T)(\mathrm{I} - \alpha\bar{\mathrm{P}})^{-1}\mathrm{e}_j.$$

3.3 Abschätzen der Ableitung

Das Abschätzen der Ableitung geschieht in zwei Teilen, zuerst werden die einzelnen Komponenten abgeschätzt und dann der gesamte Vektor.

Vorweg noch eine Abschätzung, die für beide Abschnitte gebraucht wird:

$$\begin{aligned}
||\pi^T(\alpha)(\bar{\mathrm{P}} - \mathrm{e}\mathrm{v}^T)||_1 &\leq ||\pi^T(\alpha)\bar{\mathrm{P}}||_1 + ||\pi^T(\alpha)\mathrm{e}\mathrm{v}^T||_1 \\
&\leq 1 + 1 = 2
\end{aligned} \tag{3.1}$$

3.3.1 Komponentenweise Abschätzung

Lemma 7 (Hölder-Ungleichung)**.** *Für $x \in \mathrm{e}^\perp$ (also x mit $x^T\mathrm{e} = 0$) gilt:*

$$\begin{aligned}
\forall\lambda \in \mathbb{R}: \ |x^T y| &= |x^T(y - \lambda\mathrm{e})| \\
&\leq ||x||_1 ||y - \lambda\mathrm{e}||_\infty \\
&\leq ||x||_1 \frac{y_{max} - y_{min}}{2}
\end{aligned}$$

y_{min} und y_{max} sind dabei minimale und die maximale Komponente von y.

$$\min_{\lambda\in\mathbb{R}} ||y - \lambda\mathrm{e}||_\infty = \frac{y_{max} - y_{min}}{2} \qquad \textit{(Das Minimum wird bei } \lambda = \frac{y_{max} + y_{min}}{2} \textit{ erreicht.)}$$

Für $y = (\mathrm{I} - \alpha\bar{\mathrm{P}})^{-1}\mathrm{e}_j$ ergibt sich:

$$\begin{aligned}
\left|\frac{d\pi_j(\alpha)}{d\alpha}\right| &= \left|\pi^T(\alpha)(\bar{\mathrm{P}} - \mathrm{ev}^T)(\mathrm{I} - \alpha\bar{\mathrm{P}})^{-1}\mathrm{e}_j\right| \\
&\leq ||\pi^T(\alpha)(\bar{\mathrm{P}} - \mathrm{ev}^T)||_1 \left(\frac{y_{max} - y_{min}}{2}\right) \\
&\leq y_{max} - y_{min} && \text{nach Gleichung (3.1)}
\end{aligned}$$

Aus $(\mathrm{I} - \alpha\mathrm{P})^{-1} \geq 0$ folgt: $y_{min} \geq 0$.
Weiterhin gilt:

$$\begin{aligned}
y_{max} &\leq \max_{i,j} \left[(\mathrm{I} - \alpha\bar{\mathrm{P}})^{-1}\right]_{ij} \leq ||(\mathrm{I} - \alpha\bar{\mathrm{P}})^{-1}||_\infty \\
&= ||(\mathrm{I} - \alpha\bar{\mathrm{P}})^{-1}\mathrm{e}||_\infty = \frac{1}{1-\alpha}
\end{aligned}$$

Also:

$$\left|\frac{d\pi_j^T(\alpha)}{d\alpha}\right| \leq \frac{1}{1-\alpha}$$

3.3.2 Abschätzung des ganzen Vektors

$$\begin{aligned}
\left\|\frac{d\pi^T(\alpha)}{d\alpha}\right\|_\infty &= ||\pi^T(\alpha)(\bar{\mathrm{P}} - \mathrm{ev}^T)(\mathrm{I} - \alpha\bar{\mathrm{P}})^{-1}||_\infty \\
&\leq ||\pi^T(\alpha)(\bar{\mathrm{P}} - \mathrm{ev}^T)||_\infty \cdot ||(\mathrm{I} - \alpha\bar{\mathrm{P}})^{-1}||_\infty \\
&\leq 2 \cdot \frac{1}{1-\alpha} = \frac{2}{1-\alpha}
\end{aligned}$$

3.3.3 Fazit

Bei beiden Abschätzungen taucht der Faktor $\frac{1}{1-\alpha}$ auf, der für $\alpha \approx 1$ sehr groß ist. Daher ergibt sich auch hier, dass π^T für α-Werte nahe 1 nicht gut abschätzen lässt.

Kapitel 4

Ausblick

Seit 2004 wird die Berechnung des PageRank auch als lineares Gleichungssystem betrachtet, deswegen können zukünftige Verbesserungen am Algorithmus nun aus zwei Bereichen schöpfen: Markowketten und linearer Algebra. Das Ziel der Suchmaschinenentwicklung ist jedoch klar: Personalisierte Suchergebnisse in Echtzeit. Davon ist man aber noch sehr weit entfernt: Früher ließ Google den PageRank einmal pro Woche neu berechnen, inzwischen ist das nur noch einmal im Monat möglich; und das, obwohl sich fast alle größeren Seiten jede Woche ändern.

Quasi außen vor gelassen wurde der Personalisierungsvektor v^T. Er taucht zwar in der gesamten Herleitung auf, wird hier jedoch nur zum Beweis, dass es mit ihm geht, mitgeschleppt. Google benutzt ihn etwa zur Abwehr von sog. Linkfarmen, die versuchen, den PageRank von bestimmten Webseiten durch massiven Einsatz von Links künstlich in die Höhe zu treiben. Allerdings könnte man jedem Nutzer – falls es durch neue Berechnungsmethoden irgendwann so weit ist – einen eigenen Vektor zuordnen und damit Einfluss auf die Reihenfolge der Suchergebnisse nehmen. Dadurch ließen sich dem Nutzer die für ihn relevantesten Suchergebnisse liefern – oder sehr gezielte Werbung.

Wegen der Kürze wurden auch Beweise, etwa für die Hölderlin-Ungleichung oder den Satz von Perron-Frobenius, ausgespart.

Danksagungen An dieser Stelle möchte ich mich bei meinem Betreuer Georg Vossen für die Unterstützung bedanken.

Letzte Änderung am 22. Februar 2008.

Literaturverzeichnis

[1] Christian Abramowski: Google-Matrizen I, 12.02.2008

[2] Jon Kleinberg: The Structure of Information Networks. Begleitmaterial zum Kurs an der Cornell University Ithaca, Herbst 2002
(`http://www.cs.cornell.edu/Courses/cs685/2002fa/`)

[3] Amy N. Langville, Carl D. Meyer: Deeper Inside PageRank. North Carolina State University, 20.10.2004

[4] Carl D. Meyer: Matrix Analysis and Applied Linear Algebra. SIAM, Philadelphia, 2000