



openHPI

Einführung in die Mathematik der Algorithmik – PageRank Korrektheit, Teil 3

Thomas Bläsius, Karen Seidel

Hasso-Plattner-Institut

Universität Potsdam

Behauptung 1: FÜR JEDES NETZWERK EXISTIERT GENAU EIN positiver Vektor x MIT $x = G \cdot x$. Dabei bezeichnet G die zum Netzwerk gehörende Google-Matrix und x ist der gesuchte PageRank-Vektor.

Behauptung 2: FÜR JEDES NETZWERK KONVERGIERT DIE zur Google-Matrix gehörende VEKTORITERATION bei positivem Anfangsvektor. Die Zwischenergebnisse nähern sich also immer mehr dem gesuchten PageRank-Vektor an.

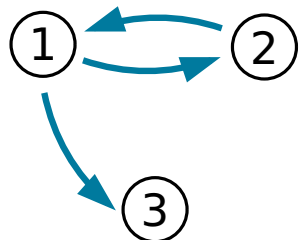
■ **Letzte offene Fragen:**

Ist die **Google-Matrix** für jedes Netzwerk **positiv**?

Hat die Google-Matrix für jedes Netzwerk **in jeder Spalte** die **Summe 1**?

Idee: Betrachte die **Berechnung der Google-Matrix** genauer.

- **Adjazenzmatrix:** $A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$
- **Link-Matrix** (modifizierte Adjazenzmatrix): $L = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & 1 \end{pmatrix}$



Modifikation 1: Ergänze auf der Diagonalen eine 1, falls in der Spalte nur 0en stehen.

Interpretation: Hat eine Seite keine ausgehenden Links, so behält sie ihren gesamten Wert.

Modifikation 2: Teile jeden Eintrag durch die Anzahl an 1en in seiner Spalte.

Interpretation: Verteile den Gesamtwert von jeder Seite gleichmäßig auf alle von dieser Seite ausgehenden Links.



FÜR JEDES NETZWERK hat die zugehörige Link-Matrix nur **Einträge größer oder gleich 0** und die **Spaltensummen** sind alle **1**.

■ Google-Matrix:

$$G = 0,85 \cdot \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & 1 \end{pmatrix} + 0,05 \cdot \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

Die **Einträge** sind **größer oder gleich 0**.

Die **Spaltensummen** betragen **0,85**.

Rechnung Spalte 1:

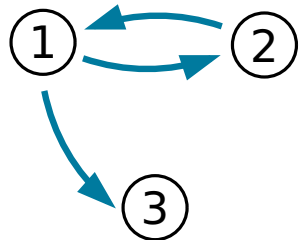
$$0,85 \cdot \frac{1}{2} + 0,85 \cdot \frac{1}{2} = 0,85$$

Alle **Einträge** sind **positiv**.

Die **Spaltensummen** betragen **0,15**.

Rechnung in jeder Spalte:

$$3 \cdot 0,05 = 0,15$$



FÜR JEDES NETZWERK ist die Google-Matrix positiv.

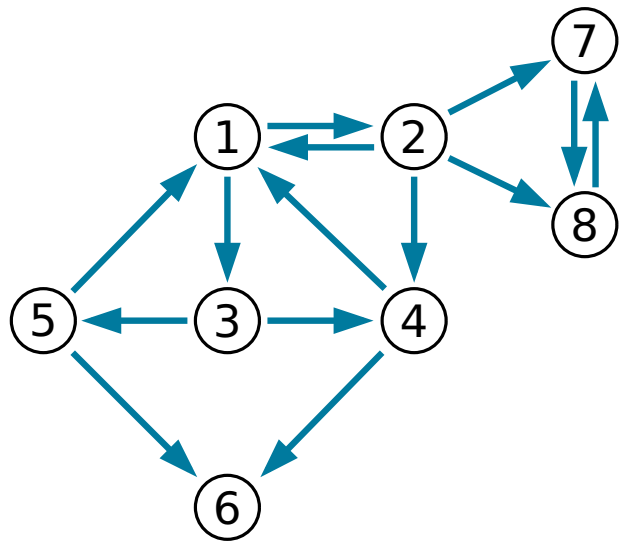


FÜR JEDES NETZWERK sind die Spaltensummen der Google-Matrix 1.



FÜR JEDES NETZWERK hat die Google-Matrix die gefragten Eigenschaften und damit GELTEN DIE BEHAUPTUNGEN 1 & 2.

Funktioniert der PageRank-Algorithmus auch für andere Werte des Dämpfungsfaktors d ?



$$A = \begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

$$L = \begin{pmatrix} 0 & \frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 1 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & \frac{1}{4} & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Modifikation 1: Ergänzung der Diagonalen

Modifikation 2: Spaltensummen 1

■ Allgemeine Berechnung der Google-Matrix:

Abhängigkeit von Webseitenzahl n und Dämpfungsfaktor d

$$G = \underbrace{0,85}_{d} \cdot \underbrace{\begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & 1 \end{pmatrix}}_L + \underbrace{0,05}_{\frac{1-d}{n}} \cdot \underbrace{\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}}_{n \times n\text{-Einsmatrix}}$$

■ Google-Matrix:

$$G = d \cdot L + \frac{1-d}{n} \cdot n \times n\text{-Einsmatrix}$$

$$= 0,6 \cdot \begin{pmatrix} 0 & \frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 1 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & \frac{1}{4} & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} + 0,05 \cdot \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Die Einträge sind größer oder gleich 0.

Die Spaltensummen betragen $d = 0,6$.

Alle Einträge sind positiv.

Die Spaltensummen betragen $1 - d = 0,4$.

→ FÜR jedes Netzwerk und JEDEN DÄMPFUNGSAKTOR $d < 1$ hat die Google-Matrix die gefragten Eigenschaften und damit KONVERGIERT DIE VEKTORITERATION gegen den eindeutigen PageRank-Vektor.

■ **Google-Matrix:** $G =$

$$\begin{pmatrix} 0,05 & 0,2 & 0,05 & 0,35 & 0,35 & 0,05 & 0,05 & 0,05 \\ 0,35 & 0,05 & 0,05 & 0,05 & 0,05 & 0,05 & 0,05 & 0,05 \\ 0,35 & 0,05 & 0,05 & 0,05 & 0,05 & 0,05 & 0,05 & 0,05 \\ 0,05 & 0,2 & 0,35 & 0,05 & 0,05 & 0,05 & 0,05 & 0,05 \\ 0,05 & 0,05 & 0,35 & 0,05 & 0,05 & 0,05 & 0,05 & 0,05 \\ 0,05 & 0,05 & 0,05 & 0,35 & 0,35 & 0,65 & 0,05 & 0,05 \\ 0,05 & 0,2 & 0,05 & 0,05 & 0,05 & 0,05 & 0,05 & 0,65 \\ 0,05 & 0,2 & 0,05 & 0,05 & 0,05 & 0,05 & 0,65 & 0,05 \end{pmatrix}$$

Der PageRank-Algorithmus ist die **Vektoriteration** $y = G \cdot x$ mit Startvektor x_{Start} .

Diese **berechnet** beliebig genau den eindeutigen positiven Vektor x mit $x = G \cdot x$.

Nach Brin und Page ist dieses x der **PageRank-Vektor**.

■ **Startvektor:** $x_{Start} =$

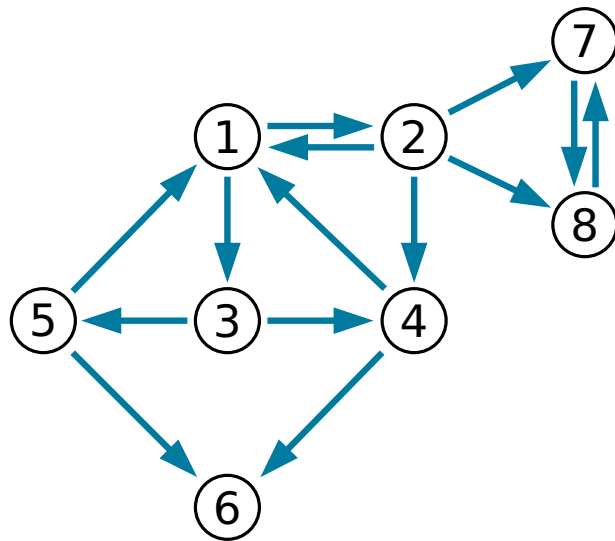
$$\begin{pmatrix} 1 \\ 0,01 \\ 0,01 \\ 0,01 \\ 0,01 \\ 0,01 \\ 0,01 \\ 0,01 \\ 0,01 \end{pmatrix}$$

■ **1. Schritt:** $y = G \cdot x_{Start}$

$$= \begin{pmatrix} 0,05 & 0,2 & 0,05 & 0,35 & 0,35 & 0,05 & 0,05 & 0,05 \\ 0,35 & 0,05 & 0,05 & 0,05 & 0,05 & 0,05 & 0,05 & 0,05 \\ 0,35 & 0,05 & 0,05 & 0,05 & 0,05 & 0,05 & 0,05 & 0,05 \\ 0,05 & 0,2 & 0,35 & 0,05 & 0,05 & 0,05 & 0,05 & 0,05 \\ 0,05 & 0,05 & 0,35 & 0,05 & 0,05 & 0,05 & 0,05 & 0,05 \\ 0,05 & 0,05 & 0,05 & 0,35 & 0,35 & 0,65 & 0,05 & 0,05 \\ 0,05 & 0,2 & 0,05 & 0,05 & 0,05 & 0,05 & 0,05 & 0,65 \\ 0,05 & 0,2 & 0,05 & 0,05 & 0,05 & 0,05 & 0,65 & 0,05 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0,05 \\ 0,05 \\ 0,05 \\ 0,05 \\ 0,05 \\ 0,05 \\ 0,05 \end{pmatrix} = \frac{1}{8} \cdot \begin{pmatrix} 1,15 \\ 0,7 \\ 0,7 \\ 0,85 \\ 0,7 \\ 1,6 \\ 1,15 \\ 1,15 \end{pmatrix} = \begin{pmatrix} 0,14375 \\ 0,0875 \\ 0,0875 \\ 0,10625 \\ 0,0875 \\ 0,2 \\ 0,14375 \\ 0,14375 \end{pmatrix}$$

■ **2. Schritt:** $y = G \cdot (G \cdot x_{Start})$

$$= \begin{pmatrix} 0,05 & 0,2 & 0,05 & 0,35 & 0,35 & 0,05 & 0,05 & 0,05 \\ 0,35 & 0,05 & 0,05 & 0,05 & 0,05 & 0,05 & 0,05 & 0,05 \\ 0,35 & 0,05 & 0,05 & 0,05 & 0,05 & 0,05 & 0,05 & 0,05 \\ 0,05 & 0,2 & 0,35 & 0,05 & 0,05 & 0,05 & 0,05 & 0,05 \\ 0,05 & 0,05 & 0,35 & 0,05 & 0,05 & 0,05 & 0,05 & 0,05 \\ 0,05 & 0,05 & 0,05 & 0,35 & 0,35 & 0,65 & 0,05 & 0,05 \\ 0,05 & 0,2 & 0,05 & 0,05 & 0,05 & 0,05 & 0,05 & 0,65 \\ 0,05 & 0,2 & 0,05 & 0,05 & 0,05 & 0,05 & 0,65 & 0,05 \end{pmatrix} \cdot \begin{pmatrix} 0,14375 \\ 0,0875 \\ 0,0875 \\ 0,10625 \\ 0,0875 \\ 0,2 \\ 0,14375 \\ 0,14375 \end{pmatrix} = \begin{pmatrix} 0,12125 \\ 0,093125 \\ 0,093125 \\ 0,089375 \\ 0,07625 \\ 0,228125 \\ 0,149375 \\ 0,149375 \end{pmatrix}$$



	PR ₁	PR ₂	PR ₃	PR ₄	PR ₅	PR ₆	PR ₇	PR ₈
0	0,1250	0,1250	0,1250	0,1250	0,1250	0,1250	0,1250	0,1250
1	0,1438	0,0875	0,0875	0,1063	0,0875	0,2000	0,1438	0,1438
2	0,1213	0,0931	0,0931	0,0894	0,0763	0,2281	0,1494	0,1494
3	0,1137	0,0864	0,0864	0,0919	0,0779	0,2366	0,1536	0,1536
4	0,1139	0,0841	0,0841	0,0889	0,0759	0,2429	0,1551	0,1551
5	0,1121	0,0842	0,0842	0,0878	0,0752	0,2452	0,1557	0,1557
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
13	0,1113	0,0834	0,0834	0,0875	0,0750	0,2469	0,1563	0,1563

- Der **PageRank-Wert** einer Webseite liegt zwischen 0 und 1. Er wird interpretiert **als** deren **relative Wichtigkeit** in der Hyperlink-Struktur.
- Eine **Vektoriteration** (PageRank-Algorithmus) **berechnet** gleichzeitig beliebig genau **alle PageRank-Werte** der Webseiten des Netzwerks
(bei einem Dämpfungsfaktor $d < 1$).
- Die **Konvergenz ergibt sich aus** Eigenschaften der Google-Matrix zum modellierenden Graphen. Wegen ihnen kann mit **von Mathematikern bewiesenen Theoremen** (Fakten 1 & 2) argumentiert werden.