



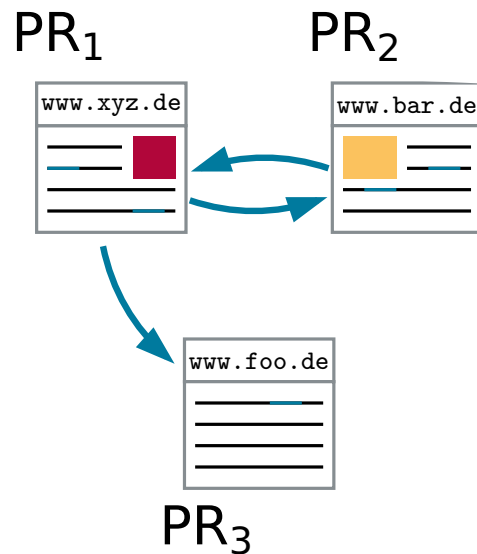
openHPI

Einführung in die Mathematik der Algorithmik – PageRank Korrektheit, Teil 2

Pascal Lenzner, Karen Seidel

Hasso-Plattner-Institut

Universität Potsdam



PageRank-Matrixgleichung:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0,05 & 0,9 & 0,05 \\ 0,475 & 0,05 & 0,05 \\ 0,475 & 0,05 & 0,9 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

18 0,1448 0,1116 0,7436

- **Ziel:** Finde einen **Vektor**, der durch Multiplikation mit der zum Netzwerk gehörigen Google-Matrix **nicht verändert wird**.
- **Offene Fragen:**
 - Warum **EXISTIERT** für jedes Netzwerk **GENAU EIN** PageRank-Vektor?
 - Wieso berechnet der PageRank-Algorithmus diesen **KORREKT**?

Mathematischer Fakt 1:

(Folgerung aus dem Perron-Frobenius-Theorem)

Für jede Matrix M mit ausschließlich positiven Einträgen, deren spaltenweise Summen alle Eins ergeben, **existiert genau ein** Vektor x **mit** ausschließlich positiven Einträgen und $x = M \cdot x$.

	PR ₁	PR ₂	PR ₃
0	0,2222	0,2222	0,2222
5	0,1094	0,1220	0,7086
18	0,1448	0,1116	0,7436

PageRank-Matrixgleichung:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0,05 & 0,9 & 0,05 \\ 0,475 & 0,05 & 0,05 \\ 0,475 & 0,05 & 0,9 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

■ Eigenschaften der Google-Matrix:

- Die **Summe** der Einträge **in jeder Spalte** ist **Eins**.
- **Alle Einträge** sind **positiv**, d.h. größer als Null.



Es gibt genau einen PageRank-Vektor x mit ausschließlich positiven Einträgen, der die PageRank-Matrixgleichung erfüllt.

Eine **Vektoriteration** ist ein Algorithmus, für den es eine Matrix M gibt, sodass beginnend mit einem Startvektor in jedem Schritt das nächste Zwischenergebnis y aus dem vorhergehenden x durch Multiplikation mit M entsteht; kurz: $y = M \cdot x$.

PR₁ PR₂ PR₃

0,2222 0,2222 0,2222

PageRank-Matrixgleichung:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0,05 & 0,9 & 0,05 \\ 0,475 & 0,05 & 0,05 \\ 0,475 & 0,05 & 0,9 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

0,1094 0,1220 0,7086

Der PageRank-Algorithmus ist eine Vektoriteration.

■ PageRank-Algorithmus als Vektoriteration:

- Jeder Eintrag des Startvektors hat den Wert $\frac{1}{3}$, also größer als Null.
- Ausgehend vom Startvektor multiplizieren wir wiederholt das jeweilige Zwischenergebnis mit der Google-Matrix.

Mathematischer Fakt 2:

(Konvergenz der Vektoriteration kombiniert mit Fakt 1)

Die **Vektoriteration** mit positivem Startvektor **liefert** für jede positive Matrix M , deren spaltenweise Summen alle Eins ergeben, den positiven Vektor x mit $x = M \cdot x$.

PR₁

PR₂

PR₃

PageRank-Matrixgleichung:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0,05 & 0,9 & 0,05 \\ 0,475 & 0,05 & 0,05 \\ 0,475 & 0,05 & 0,9 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

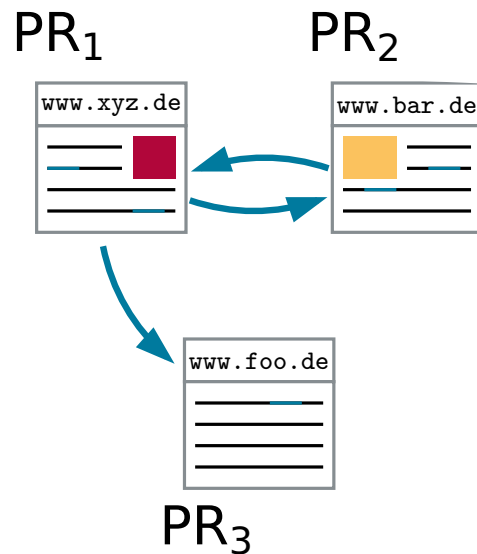
Der PageRank-Algorithmus ist eine Vektoriteration.

■ PageRank-Algorithmus als Vektoriteration:

- Jeder Eintrag des **Startvektors** hat den Wert $\frac{1}{3}$, also **größer als Null**.
- Ausgehend vom Startvektor **multiplizieren** wir **wiederholt** das jeweilige Zwischenergebnis **mit der Google-Matrix**.



Der PageRank-Algorithmus berechnet den PageRank-Vektor.



PR₁ PR₂ PR₃

0,2222 0,2222 0,2222

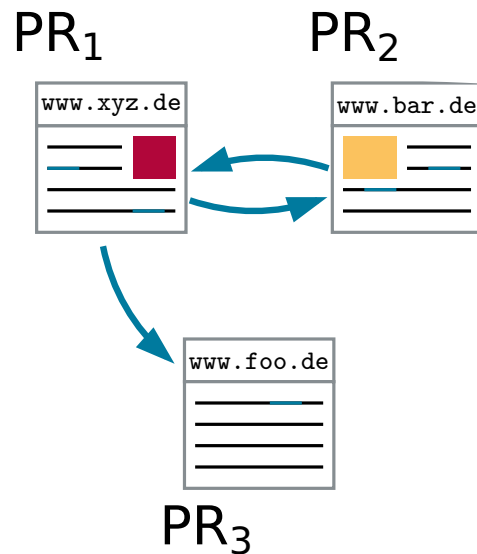
PageRank-Matrixgleichung:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0,05 & 0,9 & 0,05 \\ 0,475 & 0,05 & 0,05 \\ 0,475 & 0,05 & 0,9 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

0,1094 0,1220 0,7086

18 0,1448 0,1116 0,7436

- **Ziel:** Finde einen **Vektor**, der durch Multiplikation mit der zum Netzwerk gehörigen Google-Matrix **nicht verändert wird**.
- **Offene Fragen:** für unser Beispielnetzwerk
Warum EXISTIERT ~~für jedes Netzwerk~~ GENAU EIN PageRank-Vektor?
Wieso berechnet der PageRank-Algorithmus diesen KORREKT?



PR₁ PR₂ PR₃

0,2222 0,2222 0,2222

PageRank-Matrixgleichung:

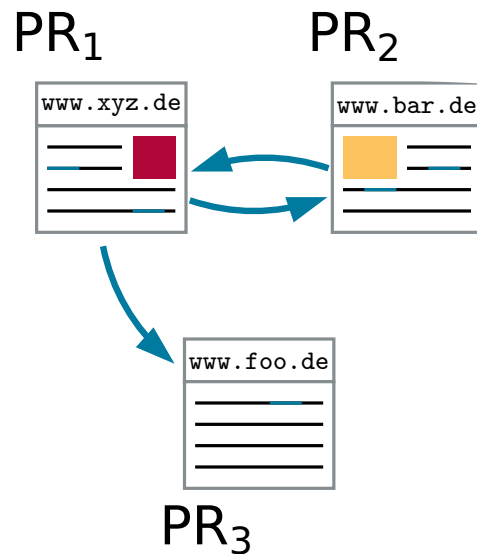
$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0,05 & 0,9 & 0,05 \\ 0,475 & 0,05 & 0,05 \\ 0,475 & 0,05 & 0,9 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

0,1094 0,1220 0,7086

18 0,1448 0,1116 0,7436

- **Ziel:** Finde einen **Vektor**, der durch Multiplikation mit der zum Netzwerk gehörigen Google-Matrix **nicht verändert wird**.
 - **Beantwortet:** für unser Beispielnetzwerk
- Warum EXISTIERT ~~für jedes Netzwerk~~ GENAU EIN PageRank-Vektor?
- Wieso berechnet der PageRank-Algorithmus diesen KORREKT?





Weiterführende Frage:

Ist die Google-Matrix für jedes Netzwerk positiv und alle spaltenweisen Summen Eins?

18 0,1448 0,1116 0,7436

- **Ziel:** Finde einen **Vektor**, der durch Multiplikation mit der zum Netzwerk gehörigen Google-Matrix **nicht verändert wird**.
 - **Beantwortet:** für unser Beispielnetzwerk
- Warum EXISTIERT ~~für jedes Netzwerk~~ GENAU EIN PageRank-Vektor?
- Wieso berechnet der PageRank-Algorithmus diesen KORREKT?

