



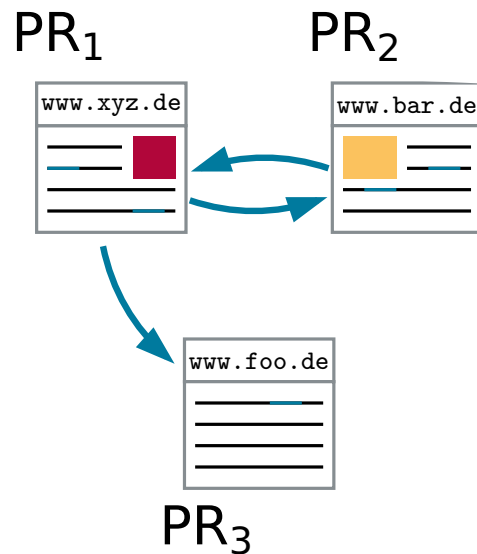
openHPI

Einführung in die Mathematik der Algorithmik – PageRank Korrektheit, Teil 1

Timo Kötzing, Karen Seidel

Hasso-Plattner-Institut

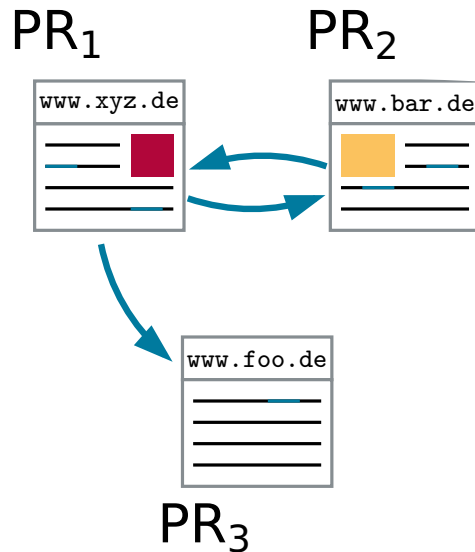
Universität Potsdam



	PR ₁	PR ₂	PR ₃
0	0,3333	0,3333	0,3333
1	0,3333	0,1917	0,4750
2	0,2129	0,1917	0,5954
3	0,2129	0,1405	0,6466
4	0,1694	0,1405	0,6901
5	0,1694	0,1220	0,7086
⋮	⋮	⋮	⋮
18	0,1448	0,1116	0,7436

- **Algorithmus:** Anfangswert $\frac{1}{3}$ für jede der drei Webseiten wählen und solange die skalierte Aktualisierungsvorschrift anwenden, bis die Werte die gewünschte Genauigkeit erreicht haben.
- **Offene Frage:**

Warum führt der Algorithmus zum ZIEL?

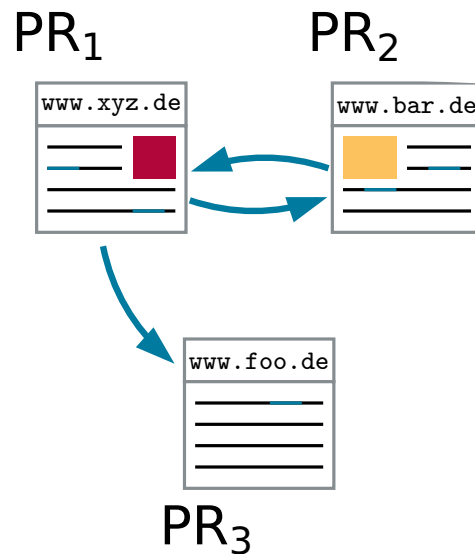


		PR ₁	PR ₂	PR ₃
	0	0,3333	0,3333	0,3333
x	1	0,3333	0,1917	0,4750
y	2	0,2129	0,1917	0,5954
	3	0,2129	0,1405	0,6466
	4	0,1694	0,1220	0,7086
	5	0,1694	0,1220	0,7086
	⋮	⋮	⋮	⋮
	18	0,1448	0,1116	0,7436

Was ist das Ziel?

■ Zweite Anwendung der Aktualisierungsvorschrift:

- $y_1 = x_2 \cdot 0,85 + 0,05 \approx 0,1917 \cdot 0,85 + 0,05 \approx 0,2129$
- $y_2 = \frac{x_1}{2} \cdot 0,85 + 0,05 \approx \frac{0,3333}{2} \cdot 0,85 + 0,05 \approx 0,1917$
- $y_3 = \left(\frac{x_1}{2} + x_3\right) \cdot 0,85 + 0,05 \approx \left(\frac{0,3333}{2} + 0,4750\right) \cdot 0,85 + 0,05 \approx 0,5954$

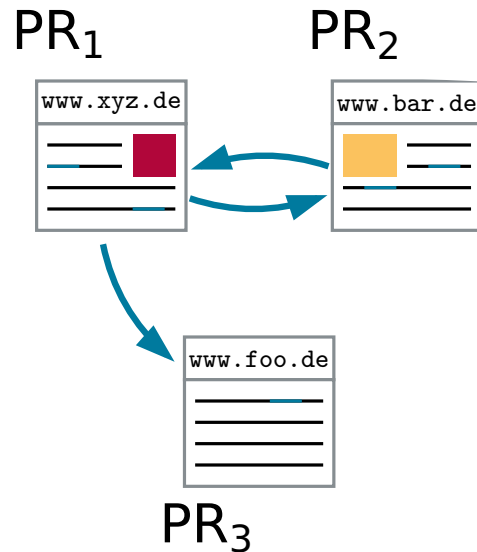


	PR₁	PR₂	PR₃
0	0,3333	0,3333	0,3333
1	0,3333	0,1917	0,4750
2	0,2129	0,1917	0,5954
3	0,2129	0,1405	0,6466
4	0,1694	0,1405	0,6901
5	0,1694	0,1220	0,7086
⋮	⋮	⋮	⋮
18	0,1448	0,1116	0,7436

■ Aktualisierungsvorschrift:

$$\begin{aligned}
 y_1 &= 0,85 \cdot \frac{x_2}{2} + 0,05 \\
 y_2 &= 0,85 \cdot \frac{x_1}{2} + 0,05 \\
 y_3 &= 0,85 \cdot \left(\frac{x_1}{2} + x_3 \right) + 0,05
 \end{aligned}$$

Die **PageRank-Werte** des Netzwerks bleiben bei Anwendung der Aktualisierungsvorschrift alle gleichzeitig unverändert.



PR₁ PR₂ PR₃

PageRank-Gleichungen:

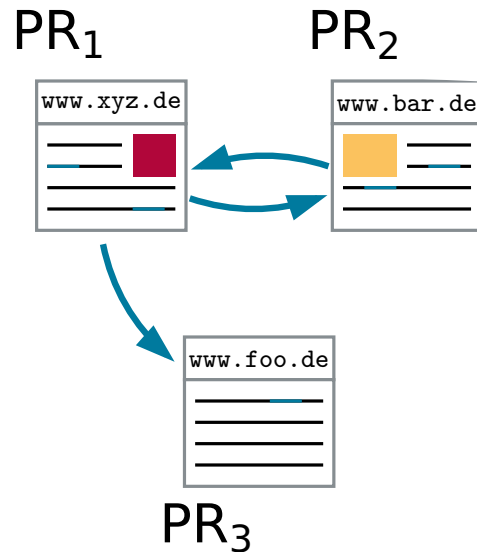
$$\begin{aligned}
 x_1 &= 0,85 \cdot x_2 + 0,05 \\
 x_2 &= 0,85 \cdot \frac{x_1}{2} + 0,05 \\
 x_3 &= 0,85 \cdot \left(\frac{x_1}{2} + x_3 \right) + 0,05
 \end{aligned}$$

18 0,1448 0,1116 0,7436

■ Aktualisierungsvorschrift:

$$\begin{aligned}
 y_1 &= 0,85 \cdot x_2 + 0,05 \\
 y_2 &= 0,85 \cdot \frac{x_1}{2} + 0,05 \\
 y_3 &= 0,85 \cdot \left(\frac{x_1}{2} + x_3 \right) + 0,05
 \end{aligned}$$

Die **PageRank-Werte** des Netzwerks bleiben bei Anwendung der Aktualisierungsvorschrift alle gleichzeitig unverändert.



PR_1 PR_2 PR_3

PageRank-Gleichungen:

$$\begin{aligned} x_1 &= 0,85 \cdot x_2 + 0,05 \\ x_2 &= 0,85 \cdot \frac{x_1}{2} + 0,05 \\ x_3 &= 0,85 \cdot \left(\frac{x_1}{2} + x_3 \right) + 0,05 \end{aligned}$$

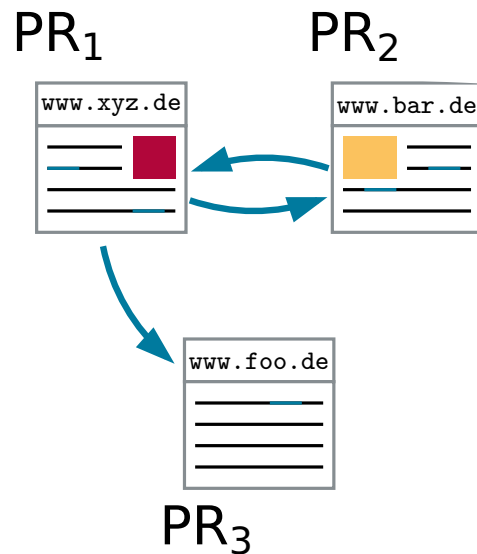
18 0,1448 0,1116 0,7436

■ PageRank-LGS umgeformt:

$$\begin{aligned} x_1 - 0,85x_2 &= 0,05 \\ -0,425x_1 + x_2 &= 0,05 \\ -0,425x_1 + 0,15x_3 &= 0,05 \end{aligned}$$

Eindeutige Lösung: $x_1 \approx 0,1448, x_2 \approx 0,1116, x_3 \approx 0,7436$





PR_1 PR_2 PR_3

PageRank-Gleichungen:

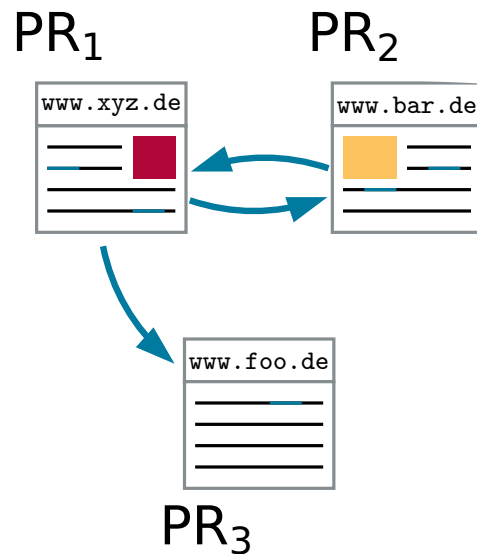
$$\begin{aligned} x_1 &= 0,85 \cdot x_2 + 0,05 \\ x_2 &= 0,85 \cdot \frac{x_1}{2} + 0,05 \\ x_3 &= 0,85 \cdot \left(\frac{x_1}{2} + x_3 \right) + 0,05 \end{aligned}$$

18 0,1448 0,1116 0,7436

■ Aktualisierungsvorschrift:

$$\begin{aligned} y_1 &= 0,85 \cdot (0 + x_2 + 0) + 0,05 \\ y_2 &= 0,85 \cdot (\frac{x_1}{2} + 0 + 0) + 0,05 \\ y_3 &= 0,85 \cdot (\frac{x_1}{2} + 0 + x_3) + 0,05 \end{aligned}$$

Wir suchen eine Darstellung als Matrix-Vektor-Multiplikation.



PR_1 PR_2 PR_3

PageRank-Gleichungen:

$$\begin{aligned} x_1 &= 0,85 \cdot \frac{x_1}{2} + 0,05 \\ x_2 &= 0,85 \cdot \frac{x_1}{2} + 0,05 \\ x_3 &= 0,85 \cdot \left(\frac{x_1}{2} + x_3 \right) + 0,05 \end{aligned}$$

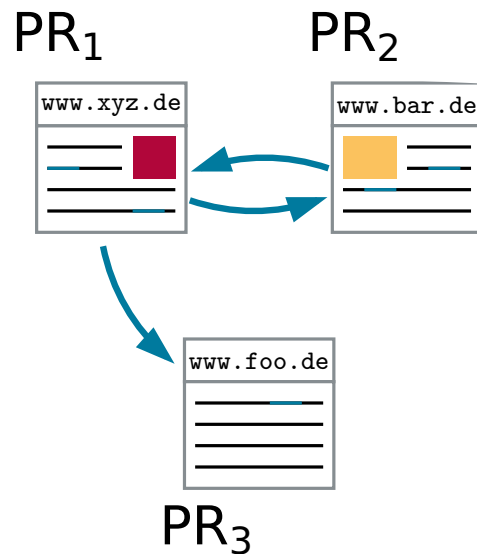
18 0,1448 0,1116 0,7436

■ Aktualisierungsvorschrift:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = 0,85 \cdot \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + 0,05 \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

In jedem Schritt des Algorithmus gilt

$$1 = x_1 + x_2 + x_3.$$



PR_1 PR_2 PR_3

PageRank-Gleichungen:

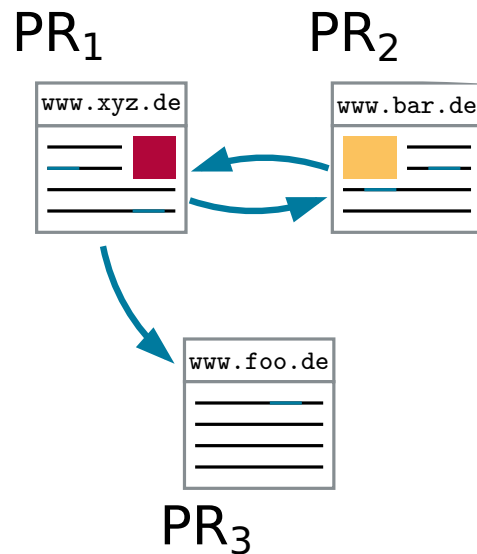
$$\begin{aligned} x_1 &= 0,85 \cdot \frac{x_1}{2} + 0,05 \\ x_2 &= 0,85 \cdot \frac{x_1}{2} + 0,05 \\ x_3 &= 0,85 \cdot \left(\frac{x_1}{2} + x_3 \right) + 0,05 \end{aligned}$$

18 0,1448 0,1116 0,7436

■ Aktualisierungsvorschrift:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = 0,85 \cdot \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + 0,05 \cdot \begin{pmatrix} x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 \end{pmatrix}$$

Wieder suchen wir eine Darstellung als Matrix-Vektor-Multiplikation.



PR₁ **PR₂** **PR₃**

0 0 2 2 2 2 0 2 2 2 2 0 2 2 2 2

PageRank-Gleichungen:

$$\begin{aligned}
 x_1 &= 0,85 \cdot x_2 + 0,05 \\
 x_2 &= 0,85 \cdot \frac{x_1}{2} + 0,05 \\
 x_3 &= 0,85 \cdot \left(\frac{x_1}{2} + x_3 \right) + 0,05
 \end{aligned}$$

5 0,1094 0,1220 0,7080

⋮

⋮

⋮

⋮

18

0,1448

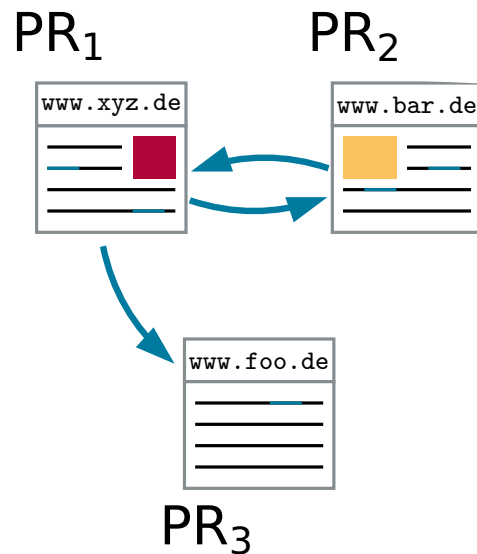
0,1116

0,7436

■ Aktualisierungsvorschrift:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = 0,85 \cdot \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + 0,05 \cdot \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

Wir klammern den x-Vektor aus und fassen alles andere zu einer einzigen Matrix zusammen.



PR₁ PR₂ PR₃

PageRank-Gleichungen:

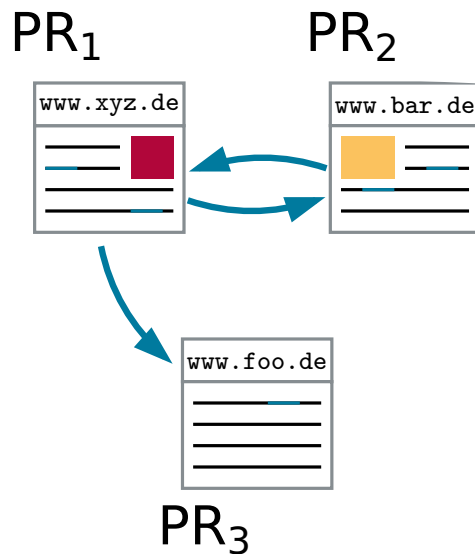
$$\begin{aligned}
 x_1 &= 0,85 \cdot \frac{x_1}{2} + 0,05 \\
 x_2 &= 0,85 \cdot \frac{x_1}{2} + 0,05 \\
 x_3 &= 0,85 \cdot \left(\frac{x_1}{2} + x_3 \right) + 0,05
 \end{aligned}$$

18 0,1448 0,1116 0,7436

■ Aktualisierungsvorschrift:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0 + 0,05 & 0,85 + 0,05 & 0 + 0,05 \\ \frac{0,85}{2} + 0,05 & 0 + 0,05 & 0 + 0,05 \\ \frac{0,85}{2} + 0,05 & 0 + 0,05 & 0,85 + 0,05 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

Wir berechnen alle Einträge der Aktualisierungsmatrix.



PR_1 PR_2 PR_3

0 0 2222 0 2222 0 2222

PageRank-Gleichungen:

$$\begin{aligned} x_1 &= 0,85 \cdot x_2 + 0,05 \\ x_2 &= 0,85 \cdot \frac{x_1}{2} + 0,05 \\ x_3 &= 0,85 \cdot \left(\frac{x_1}{2} + x_3 \right) + 0,05 \end{aligned}$$

5 0,1094 0,1220 0,7000

⋮

⋮

⋮

⋮

18

0,1448

0,1116

0,7436

■ Aktualisierungsvorschrift:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0,05 & 0,9 & 0,05 \\ 0,475 & 0,05 & 0,05 \\ 0,475 & 0,05 & 0,9 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

Die PageRank-Gleichungen sind damit äquivalent zu einer Matrixgleichung.



Ziel:
Finde 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 0,7088

18 0,1448 0,1116 0,7436

■ **Aktualisierungsmatrix:**

$$\begin{pmatrix} 0,05 & 0,9 & 0,05 \\ 0,475 & 0,05 & 0,05 \\ 0,475 & 0,05 & 0,9 \end{pmatrix}$$

Der **PageRank-Vektor** des Netzwerks wird durch die Aktualisierungsmatrix auf sich selbst abgebildet.



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

**Warum existiert für jedes Netzwerk
(genau) ein PageRank-Vektor?**

**Wieso berechnet der Algorithmus
diesen korrekt?**