Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

Lehrstuhl für
INFORMATIK I
Effiziente Algorithmen und
wissensbasierte Systeme

Institut für Informatik

Algorithmische Graphentheorie

Sommersemester 2019

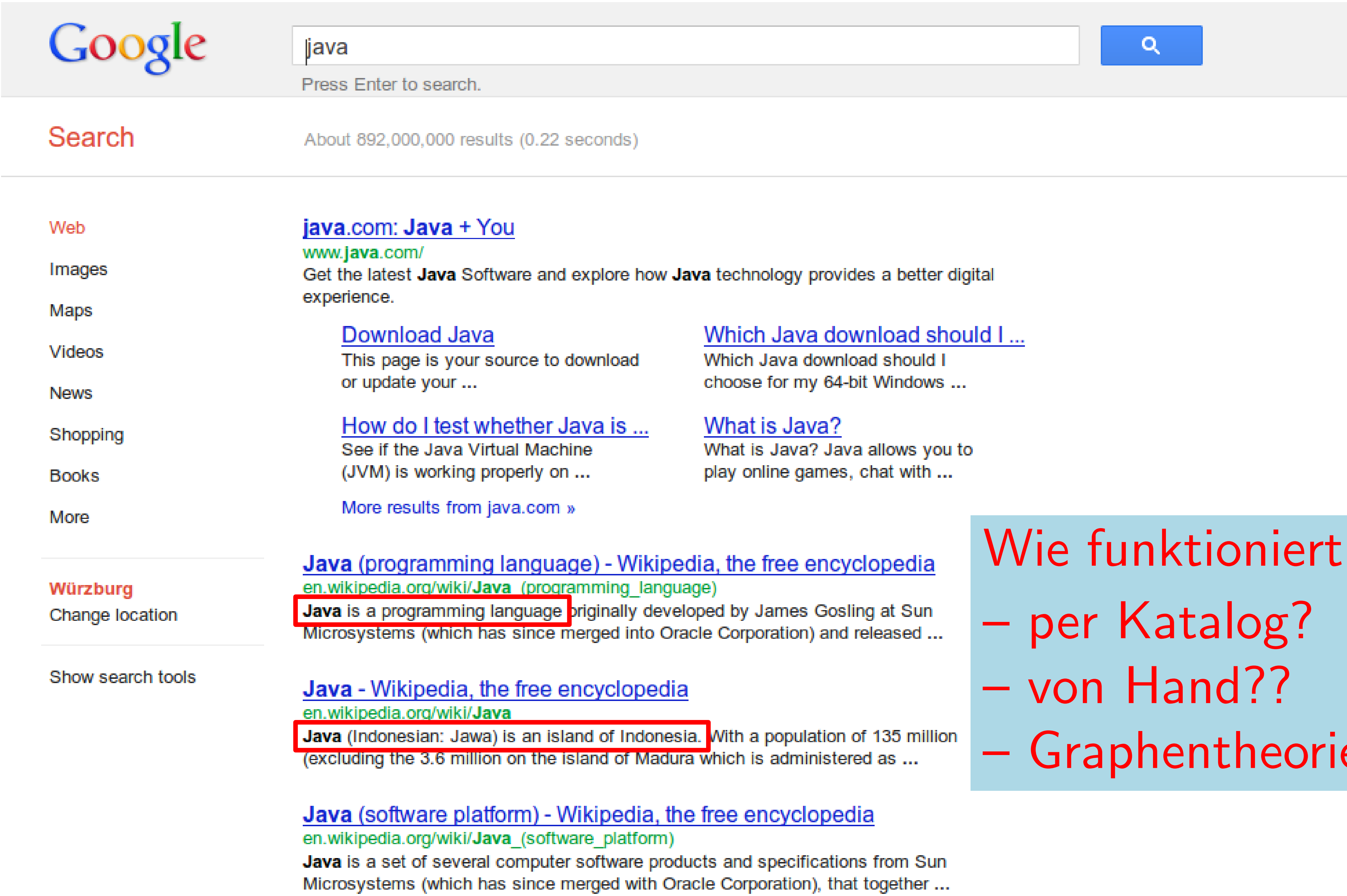
12. Vorlesung

PageRank und Power-Methode

Text siehe
www.math.cornell.edu/˜mec/Winter2009/RalucaRemus/Lecture3/lecture3.html

Dr. Joachim Spoerhase und Prof. Dr. Alexander Wolff
Lehrstuhl für Informatik I

Internet-Suche



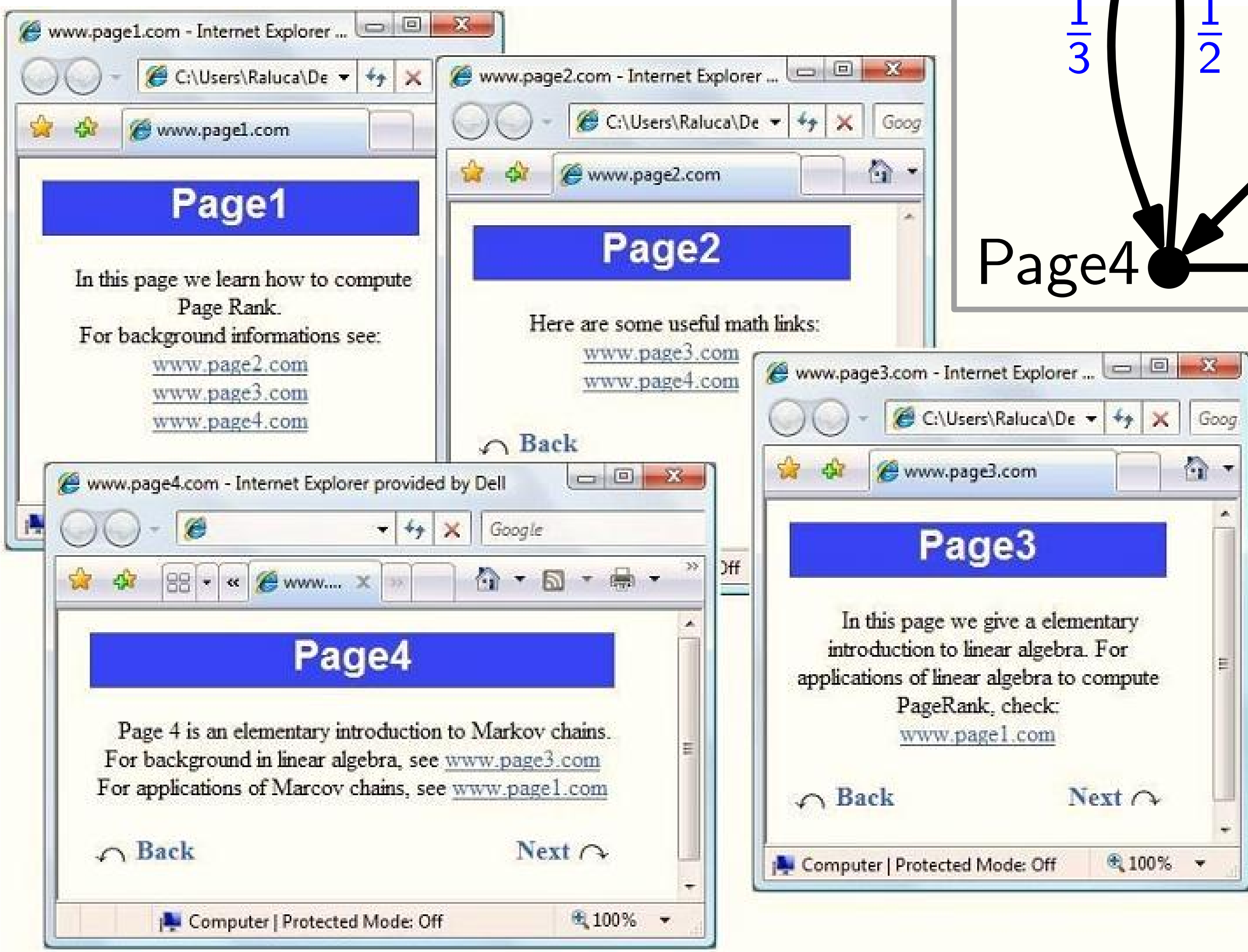
Wie funktioniert das?

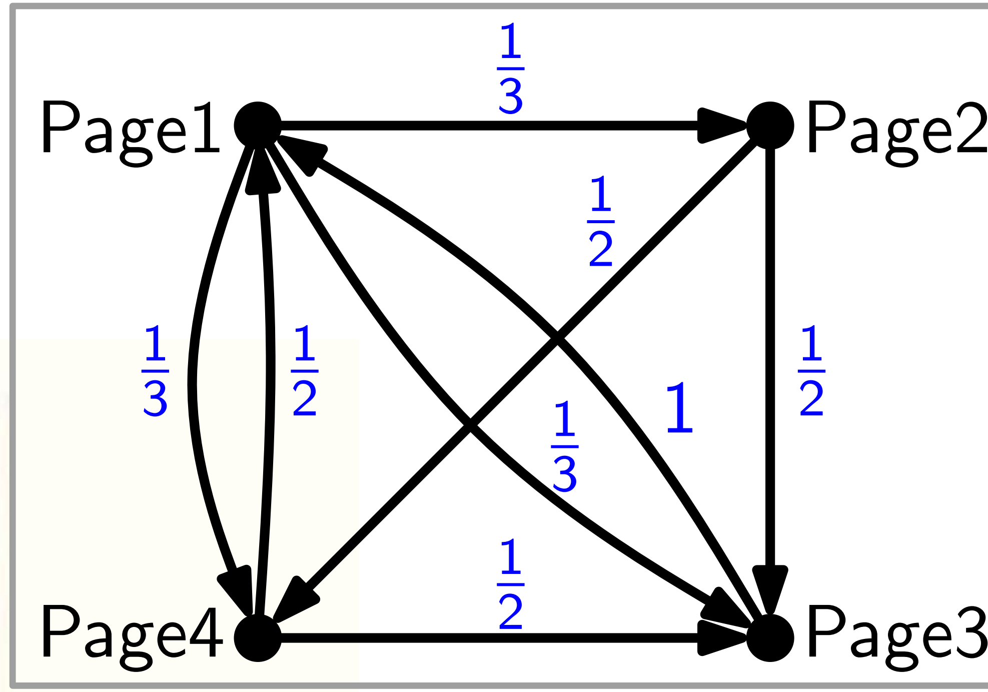
– per Katalog?

– von Hand??

– Graphentheorie???

Ein Graphen-Modell





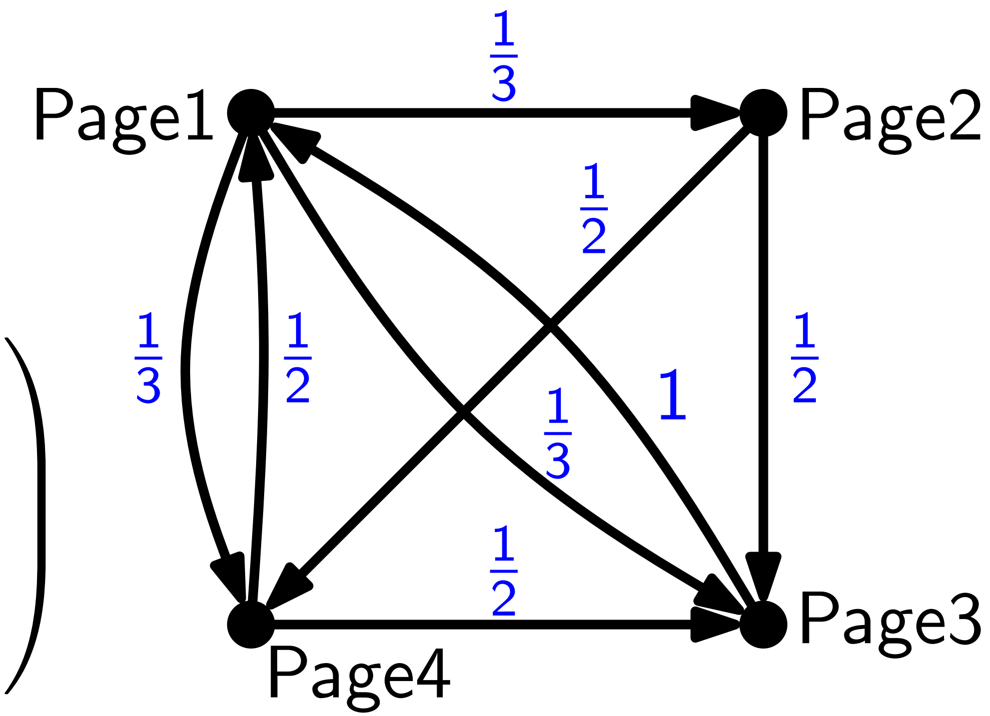
- Ein Knoten für jede Webseite.
- Eine Kante uv, falls Seite u per Link auf Seite v verweist.
- Kantengewichte $w(uv) = 1/\mathrm{outdeg}(u)$

Some Math...

Def. *Übergangsmatrix*
$A = (w_{uv})_{uv \in V \times V}$

hier:

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



Intuition:

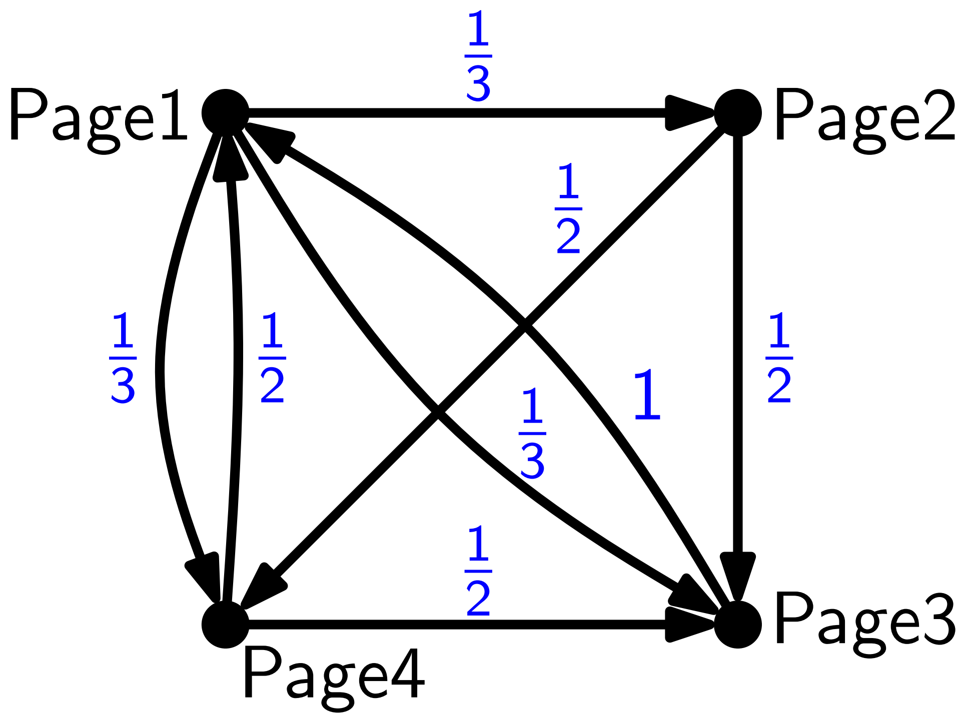
Jede Seite bekommt ein Gewicht (Rang).

Seite desto (ge-)wichtiger, je öfter sie von wichtigen Seiten verlinkt wird.

Jede Seite verteilt ihr Gewicht gleichmäßig unter ihren ausgehenden Nachbarn.

Dynamisches System

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



Anfangs bekommt jede Seite das gleiche Gewicht: $1/n$.

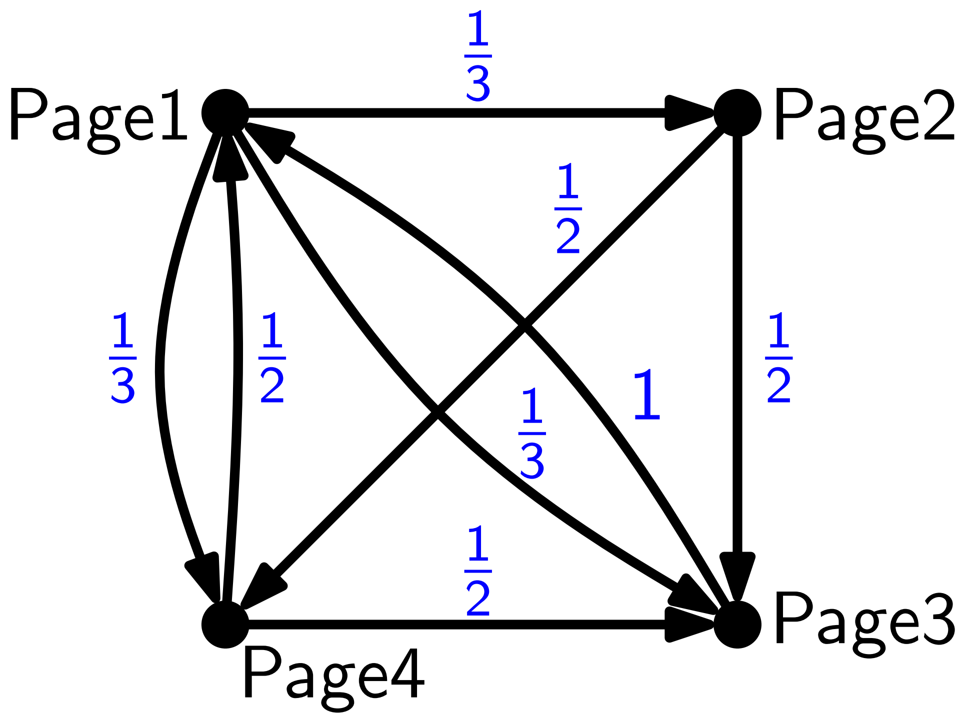
Im Beispiel: Vektor $r_0 := (1/4, 1/4, 1/4, 1/4)^\top$

Im ersten Schritt verteilt jeder Knoten sein Gewicht gleichmäßig auf seine Nachbarn:

$$r_1 := A \cdot r_0 = \begin{pmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{pmatrix} = \begin{pmatrix} 3/8 \\ 1/12 \\ 1/3 \\ 5/24 \end{pmatrix}$$

Dynamisches System

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



Iteriere dieses Verfahren!

$r_2 := Ar_1 = A^2 r_0 \approx (0{,}43; 0{,}12; 0{,}27; 0{,}16)^\top$

$r_3 := A^3 r_0 \approx (0{,}35; 0{,}14; 0{,}29; 0{,}20)^\top$

System strebt gegen Gleichgewichtszustand!

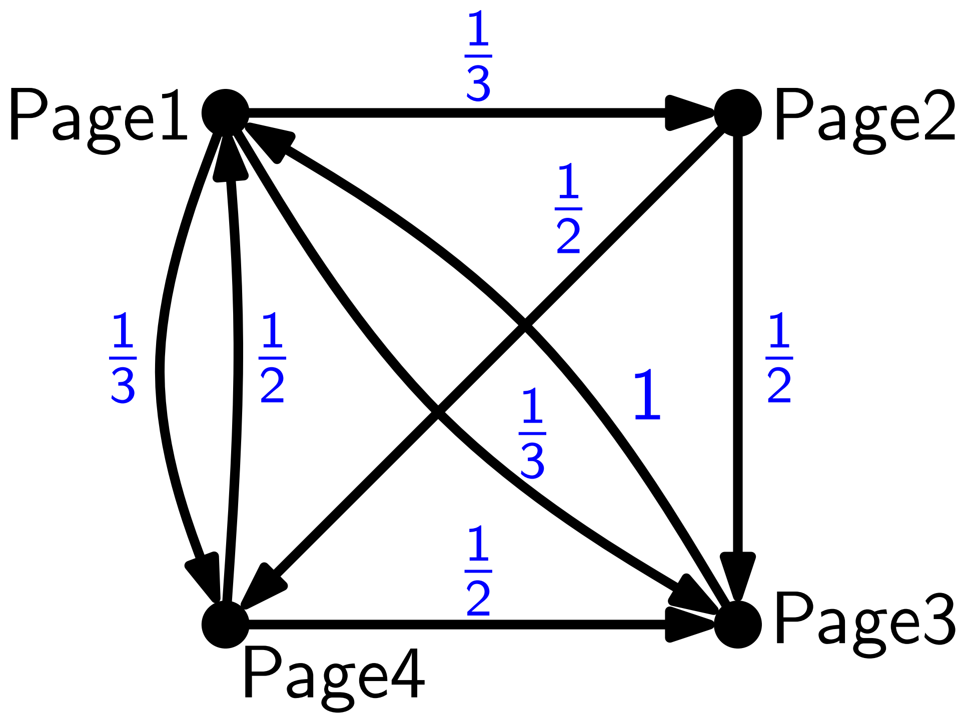
$\vdots$

$r_6 := A^6 \cdot r_0 \approx (0{,}38; 0{,}13; 0{,}29; 0{,}19)^\top$

$r_7 := A^7 \cdot r_0 \approx (0{,}38; 0{,}12; 0{,}29; 0{,}19)^\top \approx A^8 \cdot r_0 =: r_8$

Eigenvektoren

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



Im Gleichgewicht gilt $A \cdot r = r$.

D.h. r ist Eigenvektor von A für den Eigenwert 1.

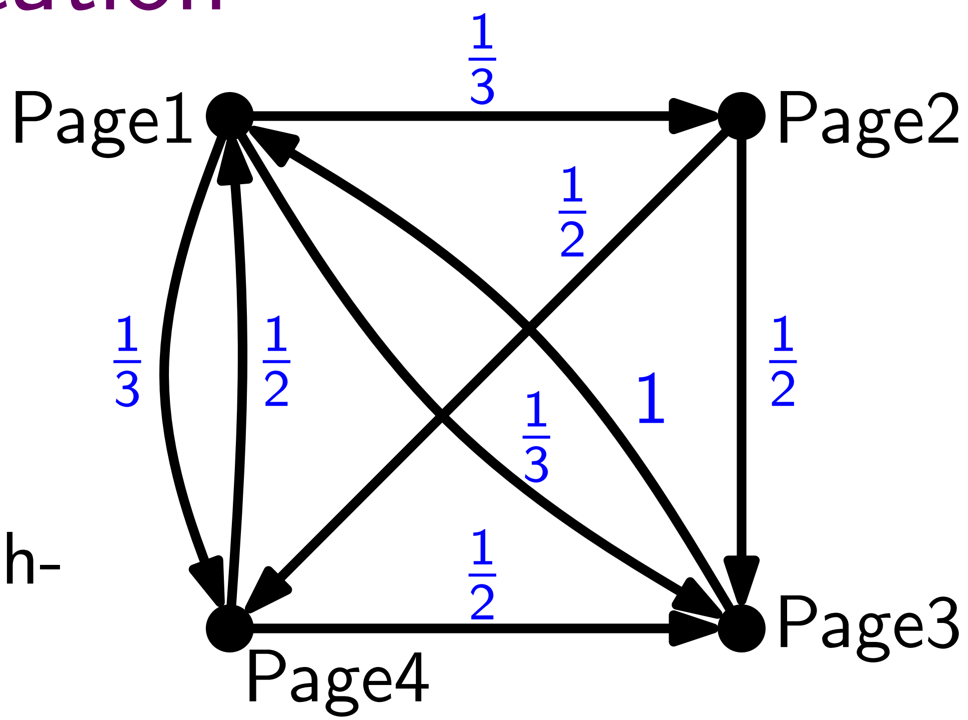
Seitenrangvektor

Lineare Algebra (LGS lösen!):
Alle Eigenvektoren zum Eigenwert 1 haben Form $c \cdot (12, 4, 9, 6)^{\top}$.

Normiere: Einträge von r summieren sich zu 1.
(probabilistischer Eigenvektor).

Probabilistische Interpretation

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



Ein „zufälliger“ Surfer startet gleichverteilt auf einer zufälligen Seite.

$\Rightarrow$ Start-Wahrscheinlichkeitsverteilung $r_0 = (1/4, 1/4, 1/4, 1/4)^\top$.

In jeder Iteration klickt der Surfer zufällig und gleichverteilt auf einen ausgehenden Link der Seite.

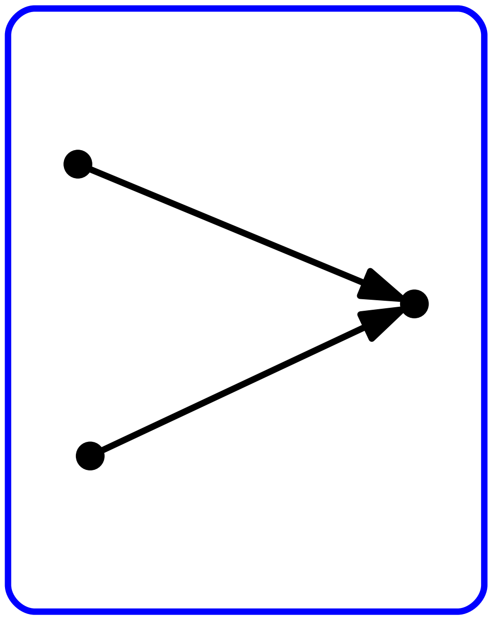
$\Rightarrow$ Wahrscheinlichkeitsverteilung in Iteration i ist $r_i = A^i \cdot r_0$.

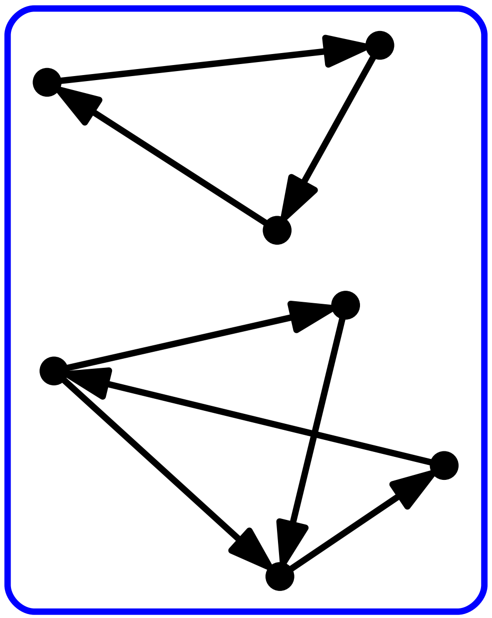
$\Rightarrow$ Stationäre Verteilung ist probabilistischer Eigenvektor für Eigenwert 1.

Probleme

- Gewicht „verschwindet“ in Knoten ohne ausgehende Kanten (nur „Eigenvektor“ 0).

- Falls Graph nicht zusammenhängend, existieren unendlich viele probabilistische Eigenvektoren zum Eigenwert 1 (Linearkombination der Eigenvektoren der ZK).





Lösung von Page und Brin


(c) Wikipedia 2008

Verbinde jede Senke mit allen anderen Knoten.

Mit einer gewissen Wahrscheinlichkeit p (Dämpfungsfaktor, beispielsweise $p = 0{,}15$) „teleportiert" sich der Surfer auf eine zufällig und gleichverteilt gewählte Seite.

Dadurch werden Senken entfernt $\rightsquigarrow$ „Zusammenhang" erreicht.

Neue Übergangsmatrix $M = (1 - p) \cdot A + p \cdot B$, wobei $B = 1/n \cdot (1)_{ij}$. Matrix M hat folgende Eigenschaften:

- M hat nur positive Einträge.
  (Dann heißt M *positiv*.)
- Alle Spaltensummen von M sind 1, da dies für A u. B gilt.
  (Dann heißt M *spalten-stochastisch*.)

Satz von Perron-Frobenius

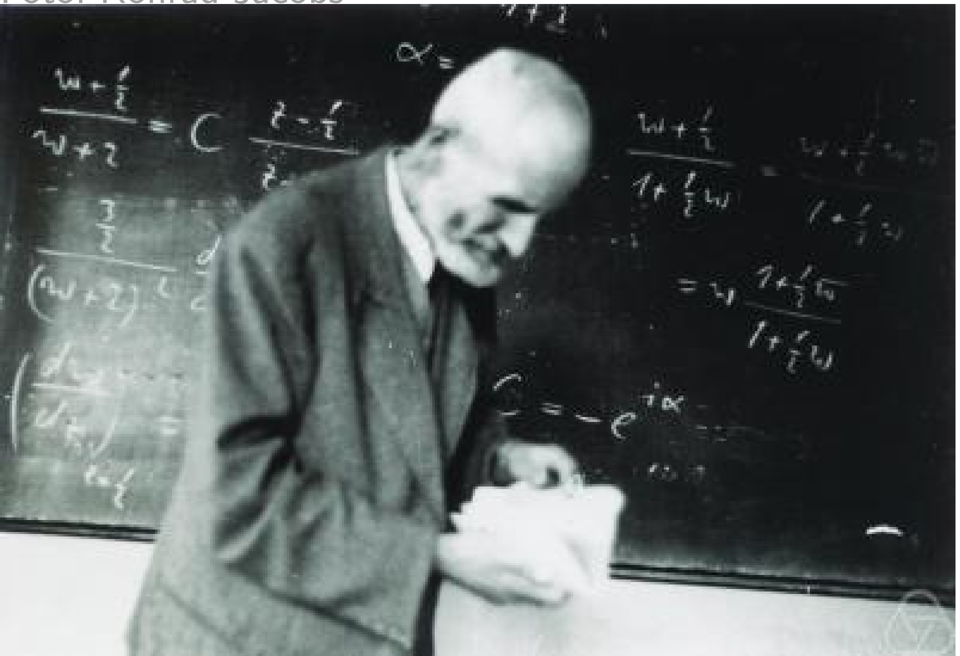
Satz. Sei M eine positive spalten-stochastische Matrix. Dann:

(i) 1 ist ein Eigenwert der Vielfachheit 1,

(ii) 1 ist der größte Eigenwert,

(iii) es gibt einen eindeutigen Eigenvektor r zum Eigenwert 1, dessen Einträge sich zu 1 summieren (probabilistischer Eigenvektor); alle Einträge von r sind positiv.

Foto: Konrad Jacobs



Oskar Perron
(Frankenthal 1880 – 1975 München)

Foto: Oberwolfach Photo Collection

Ferdinand Georg Frobenius
(Berlin 1849 – 1917 Berlin)

Power-Methode

Satz. Sei M eine positive spalten-stochastische $n \times n$ Matrix. Sei r ihr probabilistischer Eigenvektor zum Eigenwert 1. Sei r_0 ein Spaltenvektor, dessen Einträge alle $1/n$ sind. Dann konvergiert die Folge $r_0, Mr_0, M^2 r_0, \ldots, M^k r_0, \ldots$ gegen den Vektor r.

Der Seitenrangvektor für einen Web-Graphen mit
– Übergangsmatrix A und
– Dämpfungsfaktor p
kann mit Hilfe obiger Folge approximiert werden.

Dies ist häufig schneller als die exakte Berechnung.

Man muss dabei ausnutzen, dass in A fast alle Einträge 0 sind.