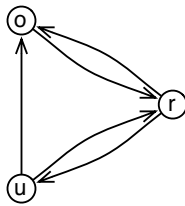


**Web-Informationssysteme, WS 2009/10**  
**Übungsblatt 4**

Besprechung am Di 17.11.2009

**Aufgabe 4-1 PageRank: Von Webs zu Matrizen**

Gegeben sei wieder der Graph aus Aufgabe 3.3 zusammen mit seiner Adjazenzmatrix:



$$\begin{array}{c}
 o \quad u \quad r \\
 \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}
 \end{array}$$

- a) In Aufgabe 3.3(d) haben wir betrachtet, wie sich Surfer über die Web-Seiten verteilen, wenn wir annehmen, dass die Surfer in jedem Schritt zufällig Links verfolgen und sich zu Beginn 12 Surfer auf o, u, und r befinden.

Sei  $o^n, u^n, r^n$  die Anzahl der Surfer auf o, u, und r nach  $n$  Schritten. Dann gilt

$$o^{n+1} = \frac{1}{2}u^n + \frac{1}{2}r^n \qquad u^{n+1} = \frac{1}{2}r^n \qquad r^{n+1} = o^n + \frac{1}{2}u^n$$

Beschreiben Sie den Übergang der Surfer von Schritt  $n$  zu Schritt  $n+1$  in Form einer Matrixgleichung.

*Hinweis:* Wir können eine Verteilung von Surfern auf Seiten durch einen Vektor  $\mathbf{v}^n$  darstellen:  $\mathbf{v}^n = (o^n, u^n, r^n)^T$ . Dann ist beispielsweise  $\mathbf{v}^0 = (12, 12, 12)^T$  (die initiale Verteilung).

- b) Angenommen es gibt eine stabile Konfiguration, in der die Anzahl der Surfer pro Seite beim nächsten Schritt gleich bleibt. In welchem Zusammenhang steht die Verteilung von Surfern auf Seiten (also ein Vektor  $\mathbf{w}$ ) zu der Matrix aus Teilaufgabe (a)?

**Aufgabe 4-2 Lineare Algebra: Eigenvektoren und Eigenwerte**

Zur Erinnerung: Ein Vektor  $\mathbf{v}$  heißt *Eigenvektor* einer linearen Transformation  $T$  (und der zugehörigen Matrix), falls sich unter der Transformation allenfalls seine Länge, aber nicht seine Richtung verändert. Es gilt also  $A\mathbf{v} = \lambda\mathbf{v}$  für einen Skalar  $\lambda$ , der dann *Eigenwert* zu  $\mathbf{v}$  genannt wird.

- a) Bestimmen Sie die Eigenwerte und Eigenvektoren für die folgenden Matrizen:

$$A_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad A_2 = \begin{pmatrix} 3 & 0 \\ 2 & -3 \end{pmatrix} \qquad A_3 = \begin{pmatrix} 15 & -10 \\ 21 & -14 \end{pmatrix} \qquad A_4 = \begin{pmatrix} -3 & 2 \\ -6 & 4 \end{pmatrix}$$

- b) Sei  $M$  eine quadratische Matrix und  $\lambda$  ein Eigenwert von  $M$ . Zeigen Sie, dass  $\lambda^k$  ein Eigenwert von  $M^k$  und  $t\lambda$  ein Eigenwert von  $tM$  für jedes  $t$ .

- c) Berechnen Sie die Eigenvektoren und Eigenwerte zu der folgenden Matrix

$$A(p) = \frac{1}{2} \begin{pmatrix} 1-p & 1+p \\ 1+p & 1-p \end{pmatrix}$$

als Funktion von  $p$ . Kommentieren Sie die Lösung im Hinblick auf den PageRank algorithmus und das Perron-Frobenius Theorem.

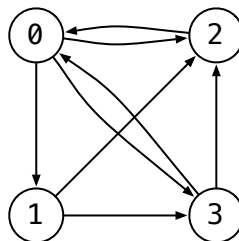
#### Aufgabe 4-3 PageRank: Einleitung

Ein Web  $W$  sei charakterisiert als ein vollständiger bipartiter Graph mit 3 “Hubs” (Knoten mit nur ausgehenden Kanten) und 2 “Autoritäten” (Knoten mit nur eingehenden Kanten). Desweiteren soll jeder Knoten in  $W$  eine Schleife zu sich selbst besitzen.

- Geben Sie den Graphen, seine Adjazenzmatrix und die für den PageRank verwendete Übergangsmatrix ohne *random leap* (normalisiert, Spalten-stochastisch) an!
- Berechnen Sie den PageRank der Seiten in  $W$  ohne *random leap*? Stimmt dies mit ihrer Erwartung überein?.
- Fügen Sie einen Link von einer der Autoritäten zu einem der Hubs hinzu und berechnen Sie den neuen PageRank (wieder kein *random leap*!).

#### Aufgabe 4-4 PageRank mit *random leap*

Berechnen Sie für den folgenden Web-Graphen die Google-Matrix mit einem *random leap*-Faktor von  $\alpha = 0.15$ :



In diesem Web hat Seite 2 einen geringeren PageRank als Seite 0. Das gefällt dem Besitzer von Seite 2 gar nicht. Er legt eine neue Seite 4 an, die nur auf Seite 2 verweist. Zu Seite 2 fügt er einen Link auf Seite 4 *hinzu*. Steigt dadurch der Rang von Seite 2 über den von Seite 0?

#### Aufgabe 4-5 PageRank: Spam

Gegeben ein Web mit  $N$  untereinander verlinkten Seiten. Desweiteren nehmen wir an, dass ein *Random surfer* mit einer Wahrscheinlichkeit von  $1 - q$  zu einer beliebigen Seite springt.

Ein Spammer möchte zu diesem Web seine Webseite `rollex.com` hinzufügen mit einem möglichst hohen PageRank. Der Spammer kann höchstens  $k$  weitere Seiten (außer `rollex.com`) hinzufügen (so dass es schließlich  $N + k + 1$  Seiten gibt). Wie viele Seiten sollte er hinzufügen und wie sollten sie verlinkt werden, um den PageRank von `rollex.com` zu maximieren. Betrachten Sie zwei Fälle:

- Keine der bereits existierenden  $N$  Seiten kann auf die  $k + 1$  neuen Seiten linken, aber die neuen Seiten können untereinander beliebig verlinkt sein.

- b) Der Spammer kann neue Links auch auf bestehenden Seiten erzeugen (aber keine bestehenden Löschen!).
- c) Diskutieren Sie den Effekt dieser Annahme im Kontext von Google und dem WWW.

*Hinweis:* Hinzufügen von Links in bestehende Seiten kann beispielsweise erreicht werden durch das Posten von Kommentaren, Forenbeiträgen etc. Daher werden Links in solchen Fällen oft mit `rel="nofollow"` versehen.