

## Workshop - PageRank

I denne Workshop betragter vi en person som surfer rundt på internettet. For at simplificere det en smule kigger vi kun på 5 mulige internetsider givet i mængden  $W = \{w_1, w_2, w_3, w_4, w_5\}$ . Vi antager, at personen via links på siderne går fra en af hjemmesiderne i mængden  $W$  til en anden. Dette illustrerer vi ved de stokastiske variable  $W^{(t)}$ . Udfaldet af  $W^{(t)}$  er nummeret på den internetside personen er på på tidspunktet  $t$ . Lad nu sandsynlighederne for første udfald være

$$P(W^{(0)} = 1) = \frac{1}{2}, \quad P(W^{(0)} = i) = \frac{1}{8}, \text{ hvor } i \in \{2, 3, 4, 5\}. \quad (1)$$

Og sandsynlighederne for de resterende være

$$\begin{aligned} P(W^{(t+1)} = 1 | W^{(t)} = 1) &= 0, & P(W^{(t+1)} = i | W^{(t)} = 1) &= \frac{1}{4} \\ P(W^{(t+1)} = 1 | W^{(t)} = i) &= \frac{1}{2}, & P(W^{(t+1)} = i | W^{(t)} = j) &= \frac{1}{8} \end{aligned}$$

for  $i, j \in \{2, 3, 4, 5\}$ . Dette er illustreret i Figur 1.

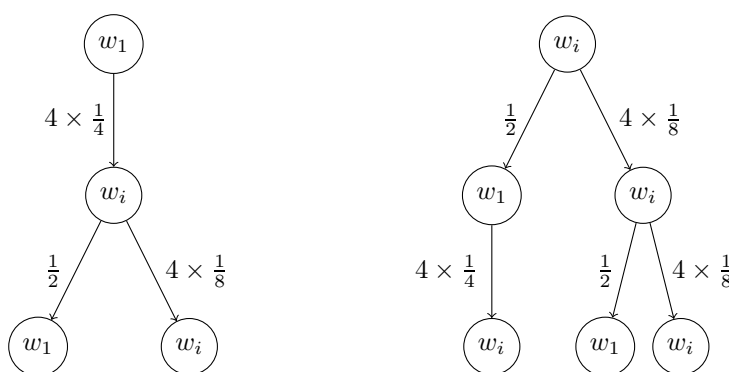


Figure 1: Sandsynligheder. Til venstre ses, hvordan det vil se ud, hvis udfaldet i første eksperiment er  $w_1$ , hvilket sker med sandsynlighed  $\frac{1}{2}$ , og til højre ses hvordan det vil se ud, hvis udfaldet af første eksperiment er en af de andre  $w_i$ ,  $i = 2, \dots, 5$ , hvilket sker med sandsynlighed  $4 \times \frac{1}{8}$  (dvs, sandsynlighed  $\frac{1}{8}$  for hver  $i = 2, \dots, 5$ ).

Lad nu  $X$  være en stokastisk variabel som tæller hvor mange gange internetside 1 besøges i tidsintervallet  $t = 0, 1, 2$ . Dvs

$$X = \text{antallet af gange } W^{(t)} = 1 \text{ er opfyldt, } t = 0, 1, 2.$$

### Delopgave 1

1. Vi lader udfaldsrummet være

$$S = \{w_i w_j w_k \mid i, j, k \in \{1, 2, 3, 4, 5\}\},$$

hvor hvert udfald svarer til de 3 internetsider besøgte i løbet af tidsintervallet  $t = 0, 1, 2$ . Husk, at  $X$  er en funktion fra  $S$  til  $\mathbb{R}$  og man kan derfor snakke om Urbilledet af  $y \in \mathbb{R}$ , hvilket er givet ved  $\{s \in S \mid X(s) = y\}$ . Forklar, hvorfor dette Urbillede svarer til en hændelse.

**Svar** Udfaldsrummet  $S$  er alle de mulige udfald - deres sum er 1. Et udfald er givet ved  $s$ . Urbilledet af  $y$  svarer til en hændelse da  $S$  indeholder alle hændelser hvoraf  $s$  er en enkelt hændelse.

2. For hvilke  $y \in \mathbb{R}$  har vi et ikke-tomt Urbillede? Og hvad er  $p(X = y)$  for disse  $y$ ?

**Svar** Vi har et ikke-tomt Urbillede af  $y$  hvis  $X$  er opfyldt mellem 0 og 3 gange. Det er altså når  $y \in \{0, 1, 2, 3\}$  da det ikke vil være muligt at få  $y = 4$  da vi maks kan besøge side 1 i alt 3 gange og mindst 0 gange.

$$p(X = 0) = p(w_i w_i w_i) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8} = 0.125$$

$$p(X = 1) = p(w_1 w_i w_i) + p(w_i w_1 w_i) + p(w_i w_i w_1) = \frac{1}{2} \cdot 1 \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{5}{8} = 0.625$$

$$p(X = 2) = p(w_1 w_1 w_i) + p(w_i w_1 w_1) + p(w_1 w_i w_1) = \frac{1}{2} \cdot 0 \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1 \cdot \frac{1}{2} = \frac{2}{8} = 0.25$$

$$p(X = 3) = p(w_1 w_1 w_1) = \frac{1}{2} \cdot 0 \cdot 0 = 0$$

Summen af sandsynligheden for disse fire udfald er 1.

3. Hvad er middelværdien og variansen for  $X$ ?

**Svar** Middelværdi:  $E(X) = \sum_{s \in S} p(s)X(s) = 0.125 \cdot 0 + 0.625 \cdot 1 + 0.25 \cdot 2 = \mathbf{1.125}$

$$\begin{aligned} \text{Varians: } V(X) &= \sum_{s \in S} (X(s) - E(X))^2 p(s) \\ &= (0 - 1.125)^2 \cdot 0.125 + (1 - 1.125)^2 \cdot 0.625 + (2 - 1.125)^2 \cdot 0.25 = \mathbf{0.359375} \end{aligned}$$

4. Hvis  $W^{(t)}$  i stedet havde fulgt en Bernoullifordeling for alle  $t$  med sandsynlighed  $\frac{3}{8}$  for at få  $w_1$  (hvilket vi betegner som vores "succes"), hvilken fordeling havde  $X$  så fulgt og hvad havde middelværdien og variansen været? Sammenlign med det foregående eksempel.

**Svar**

Bernoulli trial =  $b(k, n, p)$  where  $k$ =successes,  $n$ =independent trials,  $p$ =probability of success

$$\begin{aligned} b(k, n, p) &= C(n, k) p^k q^{n-k} \\ b\left(0; 3; \frac{3}{8}\right) &= \frac{3!}{0! \cdot (3-0)!} \cdot \frac{3^0}{8} \cdot \left(1 - \frac{3}{8}\right)^{3-0} = 0.244140625 \\ b\left(1; 3; \frac{3}{8}\right) &= \frac{3!}{1! \cdot (3-1)!} \cdot \frac{3^1}{8} \cdot \left(1 - \frac{3}{8}\right)^{3-1} = 0.439453125 \\ b\left(2; 3; \frac{3}{8}\right) &= \frac{3!}{2! \cdot (3-2)!} \cdot \frac{3^2}{8} \cdot \left(1 - \frac{3}{8}\right)^{3-2} = 0.263671875 \\ b\left(3; 3; \frac{3}{8}\right) &= \frac{3!}{3! \cdot (3-3)!} \cdot \frac{3^3}{8} \cdot \left(1 - \frac{3}{8}\right)^{3-3} = 0.052734375 \end{aligned}$$

$$\text{Middelværdi: } E(X) = \sum_{s \in S} p(s)X(s) = 0.244140625 \cdot 0 + 0.439453125 \cdot 1 + 0.263671875 \cdot 2 + 0.052734375 \cdot 3 = \mathbf{1.125}$$

$$\begin{aligned} \text{Varians: } V(X) &= \sum_{s \in S} (X(s) - E(X))^2 p(s) \\ &= (0 - 1.125)^2 \cdot 0.244 + (1 - 1.125)^2 \cdot 0.439 + (2 - 1.125)^2 \cdot 0.263 + (3 - 1.125)^2 \cdot 0.052 = \mathbf{0.703125} \end{aligned}$$

I forhold til den tidligere opgave så er middelværdien den samme, dog så er variansen forskellig.

Bemærk, at  $W^{(t)}$  kan beskrives som en Markov-kæde med stokastisk matrix  $P$  givet ved

$$P = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \end{bmatrix}$$

I Delopgave 1 betragtede vi kun  $t = 0, 1, 2$ . Fra nu af begrænser vi os ikke til dette, men kan kigge på et vilkårligt  $t \in \mathbb{N}$ .

## Delopgave 2

1. Hvad er sandsynligheden for at være i tilstand  $w_1$  til tiden  $t = 5$ , hvis vi bruger startfordelingen fra (1) og den stokastiske matrix  $P$ ?

Svar

$$S = \begin{bmatrix} \frac{1}{2} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \end{bmatrix}^T$$

$$P^{(5)} = P^5 P^{(0)} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \end{bmatrix}^5 \begin{bmatrix} \frac{1}{2} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \end{bmatrix}^T = \begin{bmatrix} \mathbf{0.328125} \\ 0.16796875 \\ 0.16796875 \\ 0.16796875 \\ 0.16796875 \end{bmatrix}$$

Det betyder altså at sandsynligheden for at være i tilstand  $w_1$  (på internetside 1) ved tiden  $t = 5$  er 0.328125.

2. Har Markov-kæden en stationær fordeling? Hvis ja, bestem sådan en.

**Svar** Ja, Markov-kæden har en stationær fordeling da den kan beskrives som en stokastisk matrix. En stokastisk matrix er defineret ved en  $N \times N$  matrix hvor alle værdier er ikke-negative. For at finde og bestemme den stationære fordeling skal vi løse ligningen  $(P - I)x = 0$ . Så lad os gøre dette ved først at bestemme  $P - I$ :

$$P - I = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} -1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & -\frac{7}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{8} & -\frac{7}{8} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & -\frac{7}{8} & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & -\frac{7}{8} \end{bmatrix}$$

Lad os nu beregne  $(P - I)x = 0$

$$(P - I)x = 0 \iff \begin{bmatrix} -1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & -\frac{7}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{8} & -\frac{7}{8} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & -\frac{7}{8} & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & -\frac{7}{8} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & -2 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} x_1 = 2x_5 \\ x_2 = 1x_5 \\ x_3 = 1x_5 \\ x_4 = 1x_5 \\ x_5 \text{ is free} \end{bmatrix}$$

$$x_5 = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$q = \begin{bmatrix} 2/6 \\ 1/6 \\ 1/6 \\ 1/6 \\ 1/6 \end{bmatrix}$$

Markovkæden har altså en stationær fordeling som vi har bestemt til at være  $q = [2/6 \ 1/6 \ 1/6 \ 1/6 \ 1/6]^T$ . Den stationære fordeling betyder at der er større sandsynlighed for generelt at befinde sig på side 1 i forhold de andre 4 sider.

Bemærk at vi ovenfor antog, at vi kunne komme til alle andre internetsider via links uanset fra hvilken internetside vi kom fra. Det er måske fair nok i eksemplet ovenfor hvor vi kun betragter 5 sider, men ville nok ikke holde hvis vi betragtede flere sider. Bemærk, at vi stadig kan modellere det med en Markov-kæde, men at det blot vil betyde, at indgang  $i, j$  i matricen for Markov-kæden er 0, hvis der ikke er et link fra  $w_j$  til  $w_i$ .

Vi kigger nu på et lidt større eksempel med 7 hjemmesider, hvor der er links mellem siderne som illustreret på Figur 2. Vi vil med dette som eksempel kigge nærmere på Google's PageRank algoritme.

I Google's PageRank algoritme betragter vi netop en person som surfer rundt på nettet. Vi bruger denne tænkte situation til at "ranke" internetsiderne. Den første tanke går på, at når surferen står på en side vil han vælge blandt de mulige links på siden med lige stor sandsynlighed. Derfor bliver den stokastiske matrix

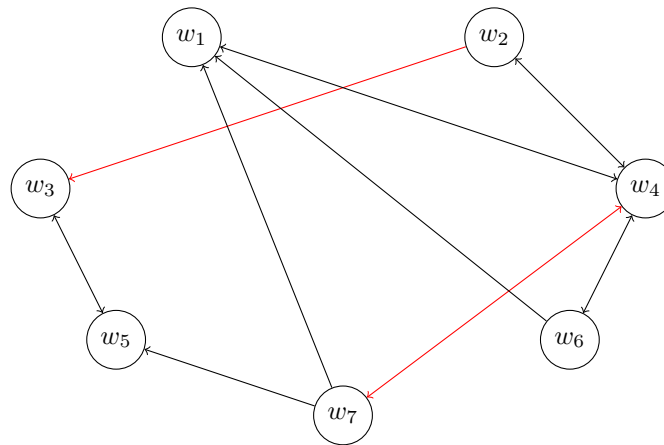


Figure 2: En pil fra en internetside til en anden betyder, at der er et link på den første til den anden side.

som illustrerer denne Markov kæde følgende (hvor både de røde og sorte pile tælles med)

$$P_1 = \begin{bmatrix} 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 1 & 0 & 0 \\ 1 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{3} \\ 0 & 0 & 1 & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & 0 \end{bmatrix} \quad (2)$$

Idéen i PageRank er mere eller mindre at finde sandsynligheden for at surferen er på en bestemt side efter “tilstrækkelig mange” klik. Det vi søger er altså en stationær fordeling.

### Delopgave 3

1. Find stationær fordeling for Markov-kæden beskrevet ud fra den stokastiske matrix  $P_1$ .

**Svar**

$$P_1 - I = \begin{bmatrix} 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 1 & 0 & 0 \\ 1 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{3} \\ 0 & 0 & 1 & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & 0 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{3} \\ 0 & -1 & 0 & \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & -1 & 0 & 1 & 0 & 0 \\ 1 & \frac{1}{2} & 0 & -1 & 0 & \frac{1}{2} & \frac{1}{3} \\ 0 & 0 & 1 & 0 & -1 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{4} & 0 & -1 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & -1 \end{bmatrix}$$

$$(P_1 - I)x = 0 \iff \begin{bmatrix} -1 & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{3} & 0 \\ 0 & -1 & 0 & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & -1 & 0 & 1 & 0 & 0 & 0 \\ 1 & \frac{1}{2} & 0 & -1 & 0 & \frac{1}{2} & \frac{1}{3} & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & -1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Dette betyder at den stationære fordeling bliver  $q$  hvor  $x_5$  er free:

$$w = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, q = \begin{bmatrix} 0 \\ 0 \\ 0.5 \\ 0 \\ 0.5 \\ 0 \\ 0 \end{bmatrix}$$

2. Find stationær fordeling for Markov-kæden beskrevet ud fra Figur 2 hvis vi fjerner de røde pile. Derudover siger vi, at sandsynligheden for at gå fra en tilstand til den næste bliver fordelt ligeligt mellem de kanter der går ud fra tilstanden. (F.eks. går der to sorte pile væk fra  $w_7$ , hvorfor indgangene her bliver  $\frac{1}{2}$ ).

**Svar** Markov-kæden  $P$  bliver dermed

$$P = \begin{bmatrix} 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$(P - I)x = 0 \iff \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -\frac{3}{2} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Efter at have lavet rækkeoperationer/row reduction så kommer vi frem til den overstående matrix. Deraf er løsningen:

$$x = \begin{bmatrix} \frac{3}{2}x_6 \\ x_6 \\ x_5 \\ 3x_6 \\ x_5 \\ x_6 \\ 0 \end{bmatrix}$$

Ved at vælge  $x_5 = 1$  og  $x_6 = 1$  får vi følgende

$$x = \begin{bmatrix} \frac{3}{2} \\ 1 \\ 1 \\ 3 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

$$q = \begin{bmatrix} \frac{3}{2}/8.5 \\ 1/8.5 \\ 1/8.5 \\ 3/8.5 \\ 1/8.5 \\ 1/8.5 \\ 0 \end{bmatrix}$$

Vi har nu fundet den stationære fordeling hvor  $x_5$  og  $x_6$  er *free variables*.

3. Er der en entydig stationær fordeling i de to tilfælde? Kan I forklare hvorfor der er/ikke er?

**Svar** For at der er en entydig stationær fordeling kræves det at den stokastiske matrix er **regular**, hvilket den er når alle elementer er positive og ikke nul i mindst en af  $P^{(k)}$  potens (Theorem 18). Der findes ikke en entydig stationær fordeling i opgave 3.1 og 3.2 da ingen  $P^k$  kun indeholder strengt positive indgange.

Problemet med denne første tanke er, at vi ikke altid vil have en entydig stationær fordeling. For at sikre os, at vi altid opnår dette, tilføjer PageRank en sandsynlighed for at surferen starter et helt nyt sted (altså blot indtaster en ny internetadresse i stedet for at følge de links der er på siden).

Derfor benyttes i stedet den stokastiske matrix

$$P = \alpha P_1 + (1 - \alpha)P_2 \quad (3)$$

hvor  $P_1$  er sandsynligheden via links (som vist i et eksempel i (2)) og  $P_2$  er en matrix hvor alle indgange er 1 over antallet af tilstande/internetsider (i vores eksempel  $\frac{1}{7}$ ).  $\alpha$  er en variabel som kan justeres alt efter hvor sandsynligt, vi mener det er, at surferen følger links eller skifter til en ny side.

#### Delopgave 4

1. Hvorfor vil en Markov-kæde med stokastisk matrix  $P$  som i (3) altid have en entydig stationær fordeling hvis  $0 < \alpha < 1$ ?

**Svar**

$$P = \alpha \begin{bmatrix} 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 1 & 0 & 0 \\ 1 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{3} \\ 0 & 0 & 1 & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & 0 \end{bmatrix} + (1 - \alpha) \begin{bmatrix} \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} \end{bmatrix}$$

Når  $0 < \alpha < 1$  så vil alle indgange/entries være strengt/strictly positive og dermed er  $P$  en regulær stokastisk matrix.  $P^1$  indeholder nemlig kun positive tal hvilket ved definition i Theorem 18 derfor har en entydig stationær fordeling.

2. Forklar hvad et lille/stort  $\alpha$  betyder for vores vurdering af surferens adfærd.

**Svar** Et lille  $\alpha$  vil betyde at der er stor sandsynlighed for at brugeren blot indtaster en ny internetadresse i stedet for at følge links på siden.

Et stort  $\alpha$  må så betyde at der er stor sandsynlighed for at brugeren følger links på siden.

3. Beregn den entydige stationære fordeling for de to eksempler i Delopgave 3 hvis vi sætter  $\alpha = 0.85$ , som typisk er blevet brugt i litteraturen. Ud fra resultaterne, hvordan vil I så "ranke" internetsiderne i de to eksempler?

**Svar** Ved at reducere  $P = \alpha P_1 + (1 - \alpha)P_2$  til reduced row-echelon form (ved brug af MATLAB) får vi følgende værdier.

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & -1.7083 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -4.8769 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & -2.9570 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -4.8003 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Deraf kan vi nu finde en simpel basis for løsningen:

$$x_7 = \begin{bmatrix} 1.7083 \\ 1 \\ 4.8769 \\ 2.9570 \\ 4.8003 \\ 1 \\ 1 \end{bmatrix}$$

Ved at dividere hver indgang med summen af alle indgange får vi nu sandsynlighedsvektoren  $\mathbf{q}$ .

$$\mathbf{q} = \begin{bmatrix} 1.7083/17.3425 \\ 1.000/17.3425 \\ 4.8796/17.3425 \\ 2.9570/17.3425 \\ 4.8003/17.3425 \\ 1/17.3425 \\ 1/17.3425 \end{bmatrix} = \begin{bmatrix} 0.0985048 \\ 0.05766135 \\ 0.28120974 \\ 0.17050718 \\ 0.27679424 \\ 0.05766135 \\ 0.05766135 \end{bmatrix}$$

Denne entydige stationære fordeling  $\mathbf{q}$  (steady-state vector) fortæller os at den mest vigtige internetside er nr. 3 efterfulgt af henholdsvis 5, 4, 1 og 2, 6, 7. Derfor vil jeg "ranke" internetsiderne i netop denne rækkefølge hvor side 3 er mest vigtig og 2,6,7 er mindst vigtige.

Hvis vi kigger på det andet eksempel så får vi

$$x_7 = \begin{bmatrix} 7.1058 \\ 4.6883 \\ 7.9685 \\ 13.0176 \\ 8.1982 \\ 4.6883 \\ 1 \end{bmatrix} \quad (4)$$

Og så dividere vi med summen af alle indgange og får sandsynlighedsvektoren  $\mathbf{q}$ .

$$\mathbf{q} = \begin{bmatrix} 7.1058/46.6667 \\ 4.6883/46.6667 \\ 7.9685/46.6667 \\ 13.0176/46.6667 \\ 8.1982/46.6667 \\ 4.6883/46.6667 \\ 1/46.6667 \end{bmatrix} = \begin{bmatrix} 0.15 \\ 0.10 \\ 0.17 \\ 0.28 \\ 0.18 \\ 0.10 \\ 0.02 \end{bmatrix} \quad (5)$$

Jeg vil ranke siderne fra mest vigtig til mindst vigtig på følgende måde:  $w_4, w_5, w_6, w_1, w_2, w_6, w_7$ .