UNIVERSITÄT DES SAARLANDES
Fachrichtung 6.2 - Informatik
Lehrstuhl Prof. Weikum

Informationssysteme SS 2002

Übung 10 Musterlösung

Aufgabe 1:

Zeigen Sie, dass für normalisierte Vektor (der Länge 1) das Cosinus-Ähnlichkeitsmaß und die Euklidische Distanz im Vektorraummodell dasselbe Anfrageresultats-Ranking produzieren.

Zu zeigen: Für normalisierte Vektoren ergeben das Cosinus-Ähnlichkeitsmaß und die Euklidische Distanz dasselbe Ranking.
Die Euklidische Distanz zweier Vektoren v1 und v2 ist

$$|\vec{v}_1 - \vec{v}_2| = \sqrt{(\vec{v}_1 - \vec{v}_2)^2}$$

Das Cosinus-Ähnlichkeitsmaß zweier normierter Vektoren v1 und v2 ist

$$\begin{aligned} \cos\alpha &= \frac{\vec{v}_1 \cdot \vec{v}_2}{\underbrace{|\vec{v}_1|}_{1} \cdot \underbrace{|\vec{v}_1|}_{1}} \\ &= \vec{v}_1 \cdot \vec{v}_2 \end{aligned}$$

Da der Cosinus zweier identischer Vektoren 1 ist und die Euklidische Distanz aber 0, ist, ist zu erwarten, daß die Reihenfolge der beiden Rankings unterschiedlich sein wird.

$$\begin{array}{rcllcl}
\sqrt{(\vec{v}_1 - \vec{q})^2} & > & \sqrt{(\vec{v}_2 - \vec{q})^2} & \Leftrightarrow \quad \vec{v}_1 \cdot \vec{q} & < & \vec{v}_2 \cdot \vec{q} \\
(\vec{v}_1 - \vec{q})^2 & > & (\vec{v}_2 - \vec{q})^2 & \Leftrightarrow \quad \vec{v}_1 \cdot \vec{q} & < & \vec{v}_2 \cdot \vec{q} \\
\sum_{i=1}^{n} \left(\vec{v}_{1,i}^2 - 2\vec{v}_{1,i} \cdot \vec{q}_i + \vec{q}_i^2\right) & > & \sum_{i=1}^{n} \left(\vec{v}_{2,i}^2 - 2\vec{v}_{2,i} \cdot \vec{q}_i + \vec{q}_i^2\right) & \Leftrightarrow \quad \vec{v}_1 \cdot \vec{q} & < & \vec{v}_2 \cdot \vec{q} \\
1 - 2\vec{v}_1 \cdot \vec{q} + 1 & > & 1 - 2\vec{v}_2 \cdot \vec{q} + 1 & \Leftrightarrow \quad \vec{v}_1 \cdot \vec{q} & < & \vec{v}_2 \cdot \vec{q} \\
-2\vec{v}_1 \cdot \vec{q} & > & -2\vec{v}_2 \cdot \vec{q} & \Leftrightarrow \quad \vec{v}_1 \cdot \vec{q} & < & \vec{v}_2 \cdot \vec{q} \\
\vec{v}_1 \cdot \vec{q} & < & \vec{v}_2 \cdot \vec{q} & \Leftrightarrow \quad \vec{v}_1 \cdot \vec{q} & < & \vec{v}_2 \cdot \vec{q}
\end{array}$$

Aufgabe 2:

Betrachten Sie den folgenden Korpus, der aus 4 Dokumenten besteht.

d1 : *Marcus tried to assissinate Caesar .*
d2 : *Marcus was a Roman .*
d3 : *Caesar was a ruler. All Romans were either loyal to Caesar or hated him.*
d4: *Everyone is loyal to someone. People only try to assissinate rulers they are not loyal to.*

Bei der Extraktion von Features dienen die folgenden Wörter als "Stoppwörter" (werden also nicht betrachtet):

- *a, all, and, are, either, everyone, her, him, is, not, only, or, someone, they, to, was, were, who.*

Ferner sollen alle restlichen Wörter nach der folgenden Abbildung auf ihre jeweilige Stammform reduziert werden, und Groß-/Kleinschreibung soll grundsätzlich unwesentlich sein:

- *assissinate assassin*
- *assissinated assassin*
- *assassination assassin*
- *loyality loyal*
- *rulers rule*
- *tried try*
- *hated hate*
- *Roman Rome*
- *Romans Rome*
- *ruler rule*

a) Bestimmen Sie die *idf*-Werte aller Terme (also der nach Stoppwortelimination und Stammformreduktion verbleibenden Wörter).

b) Bestimmen Sie für jedes der vier Dokumente den gewichteten Dokumentvektor aufgrund der *tf*idf*-Formel mit normalisierten *tf*-Werten und (mit Zweierlogarithmus) gedämpften *idf*-Werten:

$$\frac{tf_{ij}}{\max_k tf_{kj}} \log_2 \frac{N}{df_i} \qquad \text{für Term i in Dokument j}$$

c) Betrachten Sie die folgenden Anfragen:

q1: *Who assissinated Caesar?*
q2: *Loyality and assassination.*

Berechnen Sie die Resultatsranglisten für die beiden Anfragen gemäß Cosinus-Ähnlichkeit.

Lösung

Nach Entfernen der Stoppwörter und der Reduktion auf die entsprechenden Wortstämme lassen sich durch die Formel:

$$idf_i = \frac{N}{df_i}$$

die *idf*-Werte (Teilaufgabe a) aller Terme berechnen. Weiter werden mit den folgenden Formeln die Werte *tf* (normalisiert) und *idf* (gedämpft) sowie die gewichteten Dokumentenvektoren[1] berechnet (Teilaufgabe b).

$$\begin{aligned} tf_{ij,\,\text{normalisiert}} &= \frac{tf_{ij}}{\max_k tf_{kj}} \\ idf_{i,\,\text{gedämpft}} &= log_2 \frac{N}{df_i} \\ w_{ij} &= tf_{ij,\,\text{normalisiert}} \cdot idf_{i,\,\text{gedämpft}} \end{aligned}$$

Die entsprechenden Werte sind in den beiden folgenden Tabellen dargestellt:

Term	i	df_i	idf_i		absolute tf_{ij}				normalisierte tf_{ij}			
			absolut	gedämpft	tf_{i1}	tf_{i2}	tf_{i3}	tf_{i4}	tf_{i1}	tf_{i2}	tf_{i3}	tf_{i4}
Marcus	1	2	2	1	1	1	0	0	1	1	0	0
try	2	2	2	1	1	0	0	1	1	0	0	$\frac{1}{2}$
assassin	3	2	2	1	1	0	0	1	1	0	0	$\frac{1}{2}$
Caesar	4	2	2	1	1	0	2	0	1	0	1	0
Rome	5	2	2	1	0	1	1	0	0	1	$\frac{1}{2}$	0
rule	6	2	2	1	0	0	1	1	0	0	$\frac{1}{2}$	$\frac{1}{2}$
loyal	7	2	2	1	0	0	1	2	0	0	$\frac{1}{2}$	1
hate	8	1	4	2	0	0	1	0	0	0	$\frac{1}{2}$	0
people	9	1	4	2	0	0	0	1	0	0	0	$\frac{1}{2}$

Term	i	w_{i1}	w_{i2}	w_{i3}	w_{i4}	q_{i1}	q_{i2}
Marcus	1	1	1	0	0	0	0
try	2	1	0	0	$\frac{1}{2}$	0	0
assassin	3	1	0	0	$\frac{1}{2}$	1	1
Caesar	4	1	0	1	0	1	0
Rome	5	0	1	$\frac{1}{2}$	0	0	0
rule	6	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0
loyal	7	0	0	$\frac{1}{2}$	1	0	1
hate	8	0	0	1	0	0	0
people	9	0	0	0	1	0	0

[1] Eine Normierung der Dokumentenvektoren ist nicht notwendig, da in Teilaufgabe c) die Cosinus- Ähnlichkeit verwendet wird, bei welcher die Länge der Vektoren unerheblich ist.

Wird die Cosinus-Ähnlichkeit als Ähnlichkeitsfunktion verwendet, so ist

$$\text{sim}\left(\vec{d},\vec{q}\right) = \frac{\vec{d}\cdot\vec{q}}{\sqrt{\vec{d}^{\,2}\cdot\vec{q}^{\,2}}}.$$

Die sich daraus ergebenden Ähnlichkeiten sind in der folgenden Tabelle aufgeführt.

$\text{sim}\left(\vec{d}_i,\vec{q}_j\right)$	q_1	q_2
d_1	$\frac{1}{\sqrt{2}} \approx 70\%$	$\frac{1}{2\sqrt{2}} \approx 35\%$
d_2	0	0
d_3	$\frac{2}{\sqrt{22}} \approx 43\%$	$\frac{1}{\sqrt{22}} \approx 21\%$
d_4	$\frac{1}{\sqrt{22}} \approx 21\%$	$\frac{3}{\sqrt{22}} \approx 64\%$

Damit ergeben sich folgende Ranglisten für q1 und q2:

q1: d_1,d_3,d_4
q2: d_4,d_1,d_3

Dokument d2 ist für keine der beiden Anfragen relevant und wird deswegen nicht in die Ranglisten mit aufgenommen.

Aufgabe 3

Betrachten Sie die vier Dokumente d_1 bis d_4 der letzten Aufgabe. Nehmen Sie an, die Indexterme seien wie folgt geordnet:

1. *assassin*
2. *Caesar*
3. *hate*
4. *loyal*
5. *Marcus*
6. *people*
7. *Rome*
8. *rule*
9. *try*

Die 9x4-Term-Dokument-Ähnlichkeitsmatrix *A*, die sich aufgrund (einer Variante) der *tf*idf*-Formel ergibt, sieht folgendermaßen aus:

$$A = \begin{pmatrix} 0.5 & 0 & 0 & 0.318 \\ 0.5 & 0 & 0.539 & 0 \\ 0 & 0 & 0.637 & 0 \\ 0 & 0 & 0.318 & 0.539 \\ 0.5 & 0.707 & 0 & 0 \\ 0 & 0 & 0 & 0.637 \\ 0 & 0.707 & 0.318 & 0 \\ 0 & 0 & 0.318 & 0.318 \\ 0.5 & 0 & 0 & 0.318 \end{pmatrix}$$

$(0.707=\frac{1}{\sqrt{2}})$

Die Singulärwertzerlegung von A in U, $\boldsymbol{\Delta}$ und V liefert das folgende Resultat (mit Abrundungen):

$$\boldsymbol{\Delta} = diag\ (1.318,\ 1.003,\ 0.861,\ 0.716)$$

$$U = \begin{pmatrix} -0.331 & -0.170 & 0.385 & 0.207 \\ -0.433 & -0.021 & -0.226 & 0.586 \\ -0.248 & -0.080 & -0.611 & 0.174 \\ -0.306 & -0.407 & -0.144 & -0.304 \\ -0.459 & 0.549 & 0.319 & -0.101 \\ -0.214 & -0.434 & 0.190 & -0.462 \\ -0.360 & 0.462 & -0.277 & -0.452 \\ -0.231 & -0.257 & -0.210 & -0.144 \\ -0.331 & -0.170 & 0.385 & 0.207 \end{pmatrix} \qquad V = \begin{pmatrix} -0.590 & 0.0934 & 0.500 & 0.627 \\ -0.439 & 0.713 & 0.034 & -0.546 \\ -0.513 & -0.126 & -0.826 & 0.196 \\ -0.444 & -0.684 & 0.257 & -0.520 \end{pmatrix}$$

Wenn Sie nur die zwei größten Singulärwerte betrachten, also $k=2$ setzen, erhalten Sie die folgende Matrix A_k:

$$A_k = \begin{pmatrix} 0.241 & 0.070 & 0.245 & 0.310 \\ 0.335 & 0.236 & 0.296 & 0.268 \\ 0.185 & 0.086 & 0.178 & 0.200 \\ 0.199 & -0.115 & 0.258 & 0.458 \\ 0.408 & 0.658 & 0.241 & -0.108 \\ 0.126 & -0.186 & 0.200 & 0.423 \\ 0.322 & 0.539 & 0.184 & -0.107 \\ 0.155 & -0.050 & 0.189 & 0.311 \\ 0.241 & 0.070 & 0.245 & 0.310 \end{pmatrix}$$

a) Vergleichen Sie Dokument d4 mit allen anderen Dokumenten mit der LSI-Methode. Welches ist das ähnlichste?

Die Dokument-Ähnlichkeit von d_4 zu den anderen Dokumenten kann berechnet werden, in dem man $A^T_{k*4} \times A_k$ berechnet. Dabei kommt man zu folgenden Ergebnissen:

$$
\begin{aligned}
sim\,(d_4, d_1) &= 0{,}3903 \\
sim\,(d_4, d_2) &= -0{,}1518 \\
sim\,(d_4, d_3) &= 0{,}4827
\end{aligned}
$$

Somit ist Dokument d_4 am ähnlichsten zu d_3.

b) Vergleichen Sie den Term "*Marcus*" mit allen anderen Termen mit der LSI-Methode. Welches ist der ähnlichste Term, welches der zweitähnlichste? Begründen Sie diese Ähnlichkeitswerte intuitiv aufgrund der Termverteilung in den vier Dokumenten.

Die Term-Ähnlichkeit von Marcus (t5) zu den anderen Termen kann berechnet werden, in dem man $A_{k*5} \times A^T_k$ berechnet. Dabei kommt man zu folgenden Ergebnissen:

$$
\begin{aligned}
sim_{Term}\,(\text{Marcus, assassin}) &= 0{,}1699 \quad (20) \\
sim_{Term}\,(\text{Marcus,Caesar}) &= 0{,}3344 \quad (21) \\
sim_{Term}\,(\text{Marcus, hate}) &= 0{,}1534 \quad (22) \\
sim_{Term}\,(\text{Marcus, loyal}) &= 0{,}0182 \quad (23) \\
sim_{Term}\,(\text{Marcus, people}) &= -0{,}0685 \quad (24) \\
sim_{Term}\,(\text{Marcus,Rome}) &= 0{,}5419 \quad (25) \\
sim_{Term}\,(\text{Marcus, rule}) &= 0{,}0423 \quad (26) \\
sim_{Term}\,(\text{Marcus, try}) &= 0{,}1700
\end{aligned}
$$

Somit ist der Term Rome (t_7) am ähnlichsten zu Marcus (t_5) und Caesar ist am zweitähnlichsten.

c) Betrachten Sie erneut die Anfrage aus *Aufgabe 2*:

- **q**: *Loyality and assassination*

Berechnen Sie das Anfrageresultatsranking mit der LSI-Methode. Vergleichen Sie das Ergebnis mit dem Ranking aus Aufgabe 2 (also mit *tf*idf*-Gewichten im vollständigen Termvektorraum und dem Cosinus-Ähnlichkeitsmaß).

Daraus ergibt sich der Queryvektor $\vec{q} = (\,1\;0\;0\;1\;0\;0\;0\;0\;0\,)$ und $\vec{q}\,' = \vec{q} \times U_2 =$ $(-0{,}637\;\; -0{,}577)$. Anschließend werden die Ähnlichkeiten berechnet, indem $\vec{q}\,'$ mit jedem A^T_{k*i} multipliziert wird:

$$
\begin{aligned}
sim(q, d_1) &= \vec{q}' \times V_2^T{}_{*1} \\
&= \mathbf{0{,}3219} \\
sim(q, d_2) &= \vec{q}' \times V_2^T{}_{*2} \\
&= \mathbf{-0{,}1317} \\
sim(q, d_3) &= \vec{q}' \times V_2^T{}_{*3} \\
&= \mathbf{0{,}3994} \\
sim(q, d_4) &= \vec{q}' \times V_2^T{}_{*4} \\
&= \mathbf{0{,}6775}
\end{aligned}
$$

Damit ergibt sich das Ranking d_4, d_3 und d_1. d_2 wird mit seinem negativen Wert wohl nicht in das Ranking mit einfließen.

d) Gegeben sei ein neues Dokument *d5*:

- *Almost all Romans were loyal to Caesar, but Marcus, a Roman, tried to assassinate Caesar.*

Behandeln Sie "*almost*" und "*but*" als zusätzliche Stoppwörter und benutzen Sie dieselben Wortstämme wie in *Aufgabe 2*. Stellen Sie einen neuen Vektor für d_5 auf der Basis der normalisierten *tf*idf*-Termgewichtung auf (wie in *Aufgabe 2*), wobei Sie die bisherigen *idf*-Werte verwenden können (also auf der Basis der bisherigen Dokumente *d1* bis *d4*). Berücksichtigen Sie das neue Dokument *d5* in der Themen-Dokument-Ähnlichkeitsmatrix V_k.

Term	i	idf_i	$\log idf_i$	tf_{i5}	tf_{i5} (normiert)	w_{i5}
assassin	1	2	1	1	$\frac{1}{2}$	$\frac{1}{2}$
Caesar	2	2	1	2	1	1
hate	3	4	2	0	0	0
loyal	4	2	1	1	$\frac{1}{2}$	$\frac{1}{2}$
Marcus	5	2	1	1	$\frac{1}{2}$	$\frac{1}{2}$
people	6	4	2	0	0	0
Rome	7	2	1	2	1	1
rule	8	2	1	0	0	0
try	9	2	1	1	$\frac{1}{2}$	$\frac{1}{2}$

Anschließend wird $\vec{d_5}'$ berechnet durch $U_2^T \times \vec{d_5}$:

$$\vec{d_5}' = U_2^T \times \vec{d_5} = \begin{pmatrix} -1{,}507 \\ 0{,}342 \end{pmatrix}$$

Dieser Vektor wird nun einfach als 5. Spalte an V_2^T angehängt:

$$V_2^T = \begin{pmatrix} -0{,}590 & -0{,}439 & -0{,}513 & -0{,}444 & -1{,}507 \\ 0{,}093 & 0{,}713 & -0{,}126 & -0{,}684 & 0{,}342 \end{pmatrix}$$

Aufgabe 4:

Betrachten Sie das LSI-Beispiel mit den Backrezept-Dokumenten aus der Vorlesung. Berechnen Sie die approximative SVD auf der Basis der *k=2* größten Singulärwerte. Werten Sie die Anfragen "*baking*" und "*baking bread*" auf dieser Grundlage aus, also anhand der approximierten Term-Dokument-Matrix A_2.

Zur Vereinfachung können Sie das Skalarprodukt als Ähnlichkeitsmaß zwischen Vektoren verwenden (also nicht unbedingt das Cosinus-Maß). Vergleichen Sie das Ergebnis mit dem in der Vorlesung besprochenen Resultatsranking, das sich aufgrund der 3 größten Singulärwerte ergab.

$$\begin{aligned} A_2 &= U_2 \times \Delta_2 \times V_2^T \\ &= \begin{pmatrix} 0{,}2670 & -0{,}2567 \\ 0{,}7479 & -0{,}3981 \\ 0{,}2670 & -0{,}2567 \\ 0{,}1182 & -0{,}0127 \\ 0{,}5198 & 0{,}8423 \\ 0{,}1182 & -0{,}0127 \end{pmatrix} \times \begin{pmatrix} 1{,}6950 & 0 \\ 0 & 1{,}1158 \end{pmatrix} \times \begin{pmatrix} 0{,}4366 & 0{,}3067 & 0{,}4412 & 0{,}4909 & 0{,}5288 \\ -0{,}4717 & 0{,}7549 & -0{,}3568 & -0{,}0346 & 0{,}2815 \end{pmatrix} \end{aligned}$$

$$= \begin{pmatrix} 0{,}3327 & -0{,}0774 & 0{,}3019 & 0{,}2321 & 0{,}1587 \\ 0{,}7630 & 0{,}0535 & 0{,}7178 & 0{,}6377 & 0{,}5453 \\ 0{,}3327 & -0{,}0774 & 0{,}3019 & 0{,}2321 & 0{,}1587 \\ 0{,}0941 & 0{,}0507 & 0{,}0934 & 0{,}0988 & 0{,}1019 \\ -0{,}0586 & 0{,}9797 & 0{,}0534 & 0{,}4000 & 0{,}7305 \\ 0{,}0808 & 0{,}0722 & 0{,}0833 & 0{,}0978 & 0{,}1099 \end{pmatrix}$$

Die beiden Queryvektoren der Querys q_1 = *baking* und q_2 = *baking bread* sowie ihre Projektion in Themen $\vec{q}\,' = \vec{q} \times U_2$ sehen wie folgt aus:

$$\begin{aligned} \vec{q}_1 &= \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \\ \vec{q}_2 &= \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \\ \vec{q}_1{}' &= \begin{pmatrix} 0{,}2670 & -0{,}2567 \end{pmatrix} \\ \vec{q}_2{}' &= \begin{pmatrix} 0{,}5340 & -0{,}5134 \end{pmatrix} \end{aligned}$$

Daraus ergeben sich folgende Ähnlichkeitswerte:

$$
\begin{aligned}
sim(q_1,d_1) &= \vec{q}_1' \times V_2^T{}_{*1} \approx 0{,}2377\\
sim(q_1,d_2) &= \vec{q}_1' \times V_2^T{}_{*2} \approx -0{,}1119\\
sim(q_1,d_3) &= \vec{q}_1' \times V_2^T{}_{*3} \approx 0{,}2094\\
sim(q_1,d_4) &= \vec{q}_1' \times V_2^T{}_{*4} \approx 0{,}1400\\
sim(q_1,d_5) &= \vec{q}_1' \times V_2^T{}_{*5} \approx 0{,}0689
\end{aligned}
$$

$$
\begin{aligned}
sim(q_2,d_1) &= \vec{q}_2' \times V_2^T{}_{*1} \approx 0{,}4753\\
sim(q_2,d_2) &= \vec{q}_2' \times V_2^T{}_{*2} \approx -0{,}2238\\
sim(q_2,d_3) &= \vec{q}_2' \times V_2^T{}_{*3} \approx 0{,}4188\\
sim(q_2,d_4) &= \vec{q}_2' \times V_2^T{}_{*4} \approx 0{,}2799\\
sim(q_2,d_5) &= \vec{q}_2' \times V_2^T{}_{*5} \approx 0{,}1379
\end{aligned}
$$

, sowie die folgenden beiden (identischen) Rankings:

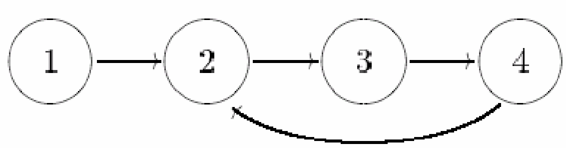
$$
\begin{aligned}
Ranking_1 &= d_1, d_3, d_4, d_5, d_2\\
Ranking_2 &= d_1, d_3, d_4, d_5, d_2
\end{aligned}
$$

Aufgabe 5:

Betrachten Sie den Graphen *G=(V,E)* mit der Knotenmenge *V={1,2,3,4}* und der Kantenmenge *E={(1,2),(2,3),(3,4),(4,2)}*. Bestimmen Sie die Transitionsmatrix *P* für den Random Walk eines Web-Surfers auf *G* mit dem Epsilon-Wert *0.1*. Berechnen Sie die Autoritätswerte von *V* nach der Methode von Page und Brin:

1. iterativ mit der Initialisierung *r(1)=r(2)=r(3)=r(4)=0.25* und 4 Iterationsschritten sowie
2. durch die entsprechende Eigenvektorberechnung bzw. das Lösen eines linearen Gleichungssystems.

Gesucht: Die Transitionsmatrix P für den Random Walk eines Web-Surfers auf dem Graphen G(V,E) mit der Knotenmenge V = {1, 2, 3, 4} (also n = |V|- 1 = 3, da beim Random Walk davon ausgegangen wird, daß man nicht "zufällig" wieder auf die Seite geht, auf der man gerade war und damit ein Dokument wegfällt), der Kantenmenge E = {(1,2),(2,3),(3,4),(4,2)} und ε = 0.1
Weiter sind die Page-Rank-Werte der Knoten gesucht.



Berechnung der Transitionsmatrix:

$$
\begin{aligned}
notequal\,(i,j) &= \begin{cases} 0 & \text{falls } i = j \\ 1 & \text{sonst} \end{cases} \\
A_{i,j} &= \begin{cases} \frac{1}{outdegree(i)} & \text{falls } (i,j) \in E \\ 0 & \text{sonst} \end{cases} \\
&= \begin{cases} 1 & \text{falls } (i,j) \in E \\ 0 & \text{sonst} \end{cases} \\
A &= \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix} \\
P_{i,j} &= \frac{\varepsilon}{n} \cdot notequal\,(i,j) + (1-\varepsilon) \cdot A_{i,j} \\
&= \frac{0{,}1}{3} \cdot notequal\,(i,j) + (1-0{,}1) \cdot A_{i,j} \\
&= 0{,}0\overline{3} \cdot notequal\,(i,j) + 0{,}9 \cdot A_{i,j} \\
r\,(i) &= \frac{\varepsilon}{n} + (1-\varepsilon) \cdot \sum_{(j,i)\in G} \frac{r\,(j)}{outdegree\,(j)} \\
&= 0{,}0\overline{3} + 0{,}9 \cdot \sum_{(j,i)\in G} \frac{r\,(j)}{outdegree\,(j)}
\end{aligned}
$$

$$
\begin{aligned}
P &= \begin{pmatrix} 0 & 0{,}0\overline{3}+0{,}9\cdot 1 & 0{,}0\overline{3}+0{,}9\cdot 0 & 0{,}0\overline{3}+0{,}9\cdot 0 \\ 0{,}0\overline{3}+0{,}9\cdot 0 & 0 & 0{,}0\overline{3}+0{,}9\cdot 1 & 0{,}0\overline{3}+0{,}9\cdot 0 \\ 0{,}0\overline{3}+0{,}9\cdot 0 & 0{,}0\overline{3}+0{,}9\cdot 0 & 0 & 0{,}0\overline{3}+0{,}9\cdot 1 \\ 0{,}0\overline{3}+0{,}9\cdot 0 & 0{,}0\overline{3}+0{,}9\cdot 1 & 0{,}0\overline{3}+0{,}9\cdot 0 & 0 \end{pmatrix} \\
&= \begin{pmatrix} 0 & 0{,}9\overline{3} & 0{,}0\overline{3} & 0{,}0\overline{3} \\ 0{,}0\overline{3} & 0 & 0{,}9\overline{3} & 0{,}0\overline{3} \\ 0{,}0\overline{3} & 0{,}0\overline{3} & 0 & 0{,}9\overline{3} \\ 0{,}0\overline{3} & 0{,}9\overline{3} & 0{,}0\overline{3} & 0 \end{pmatrix} \\
P^T &= \begin{pmatrix} 0 & 0{,}0\overline{3} & 0{,}0\overline{3} & 0{,}0\overline{3} \\ 0{,}9\overline{3} & 0 & 0{,}0\overline{3} & 0{,}9\overline{3} \\ 0{,}0\overline{3} & 0{,}9\overline{3} & 0 & 0{,}0\overline{3} \\ 0{,}0\overline{3} & 0{,}0\overline{3} & 0{,}9\overline{3} & 0 \end{pmatrix}
\end{aligned}
$$

wobei $0{,}9\overline{3} = \frac{28}{30} = \frac{14}{15}$ und $0{,}0\overline{3} = \frac{1}{30}$.

a) iterativ, beginnend mit r (1) = r (2) = r (3) = r (4) = 0.25

$$\begin{aligned}
\Pi^{(0)} &= \begin{pmatrix} 0{,}25 \\ 0{,}25 \\ 0{,}25 \\ 0{,}25 \end{pmatrix} \\
\Pi^{(1)} &= P^T \cdot \Pi^{(0)} = \begin{pmatrix} 0 & 0{,}0\overline{3} & 0{,}0\overline{3} & 0{,}0\overline{3} \\ 0{,}9\overline{3} & 0 & 0{,}0\overline{3} & 0{,}9\overline{3} \\ 0{,}0\overline{3} & 0{,}9\overline{3} & 0 & 0{,}0\overline{3} \\ 0{,}0\overline{3} & 0{,}0\overline{3} & 0{,}9\overline{3} & 0 \end{pmatrix} \cdot \begin{pmatrix} 0{,}25 \\ 0{,}25 \\ 0{,}25 \\ 0{,}25 \end{pmatrix} = \begin{pmatrix} 0{,}025 \\ 0{,}475 \\ 0{,}25 \\ 0{,}25 \end{pmatrix} \\
\Pi^{(2)} &= P^T \cdot \Pi^{(1)} = \begin{pmatrix} 0 & 0{,}0\overline{3} & 0{,}0\overline{3} & 0{,}0\overline{3} \\ 0{,}9\overline{3} & 0 & 0{,}0\overline{3} & 0{,}9\overline{3} \\ 0{,}0\overline{3} & 0{,}9\overline{3} & 0 & 0{,}0\overline{3} \\ 0{,}0\overline{3} & 0{,}0\overline{3} & 0{,}9\overline{3} & 0 \end{pmatrix} \cdot \begin{pmatrix} 0{,}025 \\ 0{,}475 \\ 0{,}25 \\ 0{,}25 \end{pmatrix} = \begin{pmatrix} 0{,}0325 \\ 0{,}265 \\ 0{,}4525 \\ 0{,}25 \end{pmatrix} \\
\Pi^{(3)} &= P^T \cdot \Pi^{(2)} = \begin{pmatrix} 0 & 0{,}0\overline{3} & 0{,}0\overline{3} & 0{,}0\overline{3} \\ 0{,}9\overline{3} & 0 & 0{,}0\overline{3} & 0{,}9\overline{3} \\ 0{,}0\overline{3} & 0{,}9\overline{3} & 0 & 0{,}0\overline{3} \\ 0{,}0\overline{3} & 0{,}0\overline{3} & 0{,}9\overline{3} & 0 \end{pmatrix} \cdot \begin{pmatrix} 0{,}0325 \\ 0{,}265 \\ 0{,}4525 \\ 0{,}25 \end{pmatrix} = \begin{pmatrix} 0{,}03225 \\ 0{,}27851\overline{6} \\ 0{,}256741\overline{6} \\ 0{,}432241\overline{6} \end{pmatrix} \\
\Pi^{(4)} &= P^T \cdot \Pi^{(3)} = \begin{pmatrix} 0 & 0{,}0\overline{3} & 0{,}0\overline{3} & 0{,}0\overline{3} \\ 0{,}9\overline{3} & 0 & 0{,}0\overline{3} & 0{,}9\overline{3} \\ 0{,}0\overline{3} & 0{,}9\overline{3} & 0 & 0{,}0\overline{3} \\ 0{,}0\overline{3} & 0{,}0\overline{3} & 0{,}9\overline{3} & 0 \end{pmatrix} \cdot \begin{pmatrix} 0{,}03225 \\ 0{,}27851\overline{6} \\ 0{,}256741\overline{6} \\ 0{,}432241\overline{6} \end{pmatrix} = \begin{pmatrix} 0{,}03225 \\ 0{,}4420836\overline{1} \\ 0{,}2754319\overline{4} \\ 0{,}24998\overline{4} \end{pmatrix}
\end{aligned}$$

b) Durch Lösen eines Gleichungssystemes.

Hier reicht es nicht aus, die Definition von r(q) direkt als Grundlage der Gleichungen zu verwenden, da dort zwar berechnet wird, wie hoch die Wahrscheinlichkeit für die Auswahl eines Knotens ist, wir diese aber in Abhängigkeit von dem Knoten, bei dem wir aktuell sind, brauchen. Weiter muß die Summe alle Wahrscheinlichkeitswerte gleich 1 sein. Entsprechend lautet das Gleichungssystem wie folgt:

$$\begin{aligned}
\vec{r} &= P^T \cdot \vec{r} \\
1 &= \sum_{i=1}^{|V|} r(i) \\
&\Longleftrightarrow
\end{aligned}$$

$$\begin{pmatrix} r(1) \\ r(2) \\ r(3) \\ r(4) \end{pmatrix} = \begin{pmatrix} 0 & 0{,}0\overline{3} & 0{,}0\overline{3} & 0{,}0\overline{3} \\ 0{,}9\overline{3} & 0 & 0{,}0\overline{3} & 0{,}9\overline{3} \\ 0{,}0\overline{3} & 0{,}9\overline{3} & 0 & 0{,}0\overline{3} \\ 0{,}0\overline{3} & 0{,}0\overline{3} & 0{,}9\overline{3} & 0 \end{pmatrix} \cdot \begin{pmatrix} r(1) \\ r(2) \\ r(3) \\ r(4) \end{pmatrix}$$

$$1 = r(1) + r(2) + r(3) + r(4)$$

$$\Longleftrightarrow$$

$$\begin{aligned}
r(1) &= \frac{1}{30} \cdot r(2) + \frac{1}{30} \cdot r(3) + \frac{1}{30} \cdot r(4) \\
r(2) &= \frac{14}{15} \cdot r(1) + \frac{1}{30} \cdot r(3) + \frac{14}{15} \cdot r(4) \\
r(3) &= \frac{1}{30} \cdot r(1) + \frac{14}{15} \cdot r(2) + \frac{1}{30} \cdot r(4) \\
r(4) &= \frac{1}{30} \cdot r(1) + \frac{1}{30} \cdot r(2) + \frac{14}{15} \cdot r(3) \\
1 &= r(1) + r(2) + r(3) + r(4)
\end{aligned}$$

Die 2te bzw. letze Gleichung (beginnend mit 1=…) ist notwendig, da man sonst nur eine der Variablen in Abhängigkeit der anderen bekommt.
Löst man dieses Gleichungssystem auf (z. B. mit Maple), so bekommt man

$$\begin{aligned}
r(1) &= 0{,}03225806452 \\
r(2) &= 0{,}3328056985 \\
r(3) &= 0{,}3221210922 \\
r(4) &= 0{,}3128151448
\end{aligned}$$

Somit ist

$$\vec{r} = \lim_{i \mapsto \infty} \Pi^{(i)} \approx \begin{pmatrix} 0{,}03 \\ 0{,}33 \\ 0{,}32 \\ 0{,}31 \end{pmatrix}.$$