Seminar 11

Definition 1 (Markov Transition Matrix)
Given the graph $G = (V, E)$ and teleport probability α, let $N = |V|$ and A be the $N \times N$ link matrix with elements

$$\forall u, v \in V \; : \; A_{uv} = \begin{cases} 1 & (u, v) \in E \\ 0 & \textit{otherwise} \end{cases}$$

The transition probability matrix P is then calculated in the following way:

1. *If a row of A has all 0s, then substitute all of them with 1s.*
2. *Divide each 1 by the number of 1s in that row.*
3. *Multiply each entry by $1 - \alpha$.*
4. *Add $\frac{\alpha}{N}$ to each entry.*

Algorithm 1 (PageRank)

```
function PageRank(P)
    i ← 0
    x⃗_i = (1, 0, ..., 0)
    x⃗_{i+1} = (0, 0, ..., 0)
    repeat
        x⃗_{i+1} = x⃗_i · P
        i = i + 1
    until x_i = x_{i-1}
end function
```

Definition 2 (Hubs and authorities)
Given the link matrix A, let $h(v)$ denote the hub score and $a(v)$ the authority score. First, set the $h(v)$ a $a(v)$ vectors to 1^N for all vertices $v \in V$. The scores are calculated as

$$\begin{aligned} h(v) &= A \cdot a(v) \\ a(v) &= A^T \cdot h(v) \end{aligned}$$

which is equivalent to

$$\begin{aligned} h(v) &= A \cdot A^T \cdot h(v) \\ a(v) &= A^T \cdot A \cdot a(v) \end{aligned}$$

Exercise 1

Assume the web graph $G = (V = \{a, b, c\}, \; E = \{(a, b), (a, c), (b, c), (c, b)\})$. Count PageRank, hub and authority scores for each of the web pages. Rank the pages by the individual scores and observe the connections. You can assume that in each step of the random walk we teleport to a random page with probability 0.1 and uniform distribution. Normalize the hub and authority scores so that the maximum is 1.

By Definition 1, rewrite the graph as a link matrix

$$\begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Trying to apply the step 1 the algorithm, does not contain a row with all 0s, so we proceed to step 2. First row contains two 1s so we divide both of them by 2. Second and third lines only contain one 1, so dividing them by 1 makes no change

$$\begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{matrix} :2 \\ :1 \\ :1 \end{matrix} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Now apply step 3. Since $\alpha = 0.1$, multiply all entries by 0.9

$$\begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \cdot \ 0.9 = \begin{pmatrix} 0 & \frac{9}{20} & \frac{9}{20} \\ 0 & 0 & \frac{9}{10} \\ 0 & \frac{9}{10} & 0 \end{pmatrix}$$

and by step 4 add $\frac{\alpha}{N} = \frac{0.1}{3} = \frac{1}{30}$

$$\begin{pmatrix} 0 & \frac{9}{20} & \frac{9}{20} \\ 0 & 0 & \frac{9}{10} \\ 0 & \frac{9}{10} & 0 \end{pmatrix} + \ \frac{1}{30} \ = \begin{pmatrix} \frac{1}{30} & \frac{29}{60} & \frac{29}{60} \\ \frac{1}{30} & \frac{1}{30} & \frac{14}{15} \\ \frac{1}{30} & \frac{14}{15} & \frac{1}{30} \end{pmatrix} = \ P.$$

Using the Algorithm 1 for PageRank, we select $\overrightarrow{x}_0 = (1, 0, 0)$ and the transition probability matrix P from the previous calculation.

$$\overrightarrow{x}_1 = \overrightarrow{x}_0 \cdot P \tag{1}$$

$$\overrightarrow{x}_1 = (1, 0, 0) \cdot \begin{pmatrix} \frac{1}{30} & \frac{29}{60} & \frac{29}{60} \\ \frac{1}{30} & \frac{1}{30} & \frac{14}{15} \\ \frac{1}{30} & \frac{14}{15} & \frac{1}{30} \end{pmatrix} \tag{2}$$

$$\overrightarrow{x}_1 = \left(\frac{1}{30}, \frac{29}{60}, \frac{29}{60}\right) \tag{3}$$

$$\overrightarrow{x}_2 = \overrightarrow{x}_1 \cdot P \tag{4}$$

$$\overrightarrow{x}_2 = \left(\frac{1}{30}, \frac{29}{60}, \frac{29}{60}\right) \cdot \begin{pmatrix} \frac{1}{30} & \frac{29}{60} & \frac{29}{60} \\ \frac{1}{30} & \frac{1}{30} & \frac{14}{15} \\ \frac{1}{30} & \frac{14}{15} & \frac{1}{30} \end{pmatrix} \tag{5}$$

$$\overrightarrow{x}_2 = \left(\frac{1}{30}, \frac{29}{60}, \frac{29}{60}\right) \tag{6}$$

Since $x_i = x_{i-1}$, we claim the entries of x_2 as PageRanks of the individual web pages. By Definition 2, we set the hub score $h(v)$ and the authority score $a(v)$ to 1s

$$a(v) = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad h(v) = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Substituting into the second equation in the definition, we get the hub score

$$
\begin{aligned}
h(v) &= A \cdot A^T \cdot h(v) \\
h(v) &= \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\
h(v) &= \begin{pmatrix} 2 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\
h(v) &= \begin{pmatrix} 4 \\ 2 \\ 2 \end{pmatrix}
\end{aligned}
$$

and the authority score

$$
\begin{aligned}
a(v) &= A^T \cdot A \cdot a(v) \\
a(v) &= \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\
a(v) &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\
a(v) &= \begin{pmatrix} 0 \\ 3 \\ 3 \end{pmatrix}.
\end{aligned}
$$

Finally, we normalize them to obtain

$$
h(v) = \begin{pmatrix} 4 \\ 2 \\ 2 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 \\ 0.5 \\ 0.5 \end{pmatrix} \qquad \text{and} \qquad a(v) = \begin{pmatrix} 0 \\ 3 \\ 3 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}.
$$