Data Mining Homework 02

Qiu Yihang

April 2023

1 Problem 01

Solution. Let the adjacency matrix of the graph be $\boldsymbol{A} \in \mathbb{R}^{(n+1)\times(n+1)}$.

Let the PageRank of the graph be $\boldsymbol{r}$, with $\boldsymbol{r}_u$ being the PageRank of node u.

We know

$$\boldsymbol{A} = \begin{pmatrix} 0 & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} & 0 \\ \frac{1}{n} & 0 & \frac{1}{n} & \cdots & \frac{1}{n} & 0 \\ \frac{1}{n} & \frac{1}{n} & 0 & \cdots & \frac{1}{n} & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots & 0 & 0 \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} & 0 \end{pmatrix}$$

Since there is a dead end, we introduce the teleport and adjust the adjacency matrix.

$$\boldsymbol{A} = \begin{pmatrix} 0 & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} & \frac{1}{n+1} \\ \frac{1}{n} & 0 & \frac{1}{n} & \cdots & \frac{1}{n} & \frac{1}{n+1} \\ \frac{1}{n} & \frac{1}{n} & 0 & \cdots & \frac{1}{n} & \frac{1}{n+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots & 0 & \frac{1}{n+1} \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} & \frac{1}{n+1} \end{pmatrix}$$

$$\boldsymbol{r} = \beta \boldsymbol{A} \cdot \boldsymbol{r} + \left[\frac{1-\beta}{n+1}\right]_{(n+1)\times 1}$$

Thus,

$$\boldsymbol{r} = \left(\frac{n}{n^2+n+\beta}, \ \frac{n}{n^2+n+\beta}, \ \cdots, \ \frac{n}{n^2+n+\beta}, \ \frac{n+\beta}{n^2+n+\beta} \right). \qquad \blacksquare$$

2 Problem 02

The original adjacency matrix is

$$\boldsymbol{W} = \begin{pmatrix} 0 & \frac{1}{2} & 1 & 0 \\ \frac{1}{3} & 0 & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{pmatrix}$$

2.1 Teleport Set is $\{\boldsymbol{A}\}$

Solution. The adjusted adjacency matrix when $\beta = 0.8$ is

$$\boldsymbol{A} = \beta \boldsymbol{W} + (1-\beta) \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & \frac{2}{5} & \frac{4}{5} & 0 \\ \frac{4}{15} & 0 & 0 & \frac{2}{5} \\ \frac{4}{15} & 0 & 0 & \frac{2}{5} \\ \frac{4}{15} & \frac{2}{5} & 0 & 0 \end{pmatrix} + \begin{pmatrix} \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{5} & \frac{3}{5} & 1 & \frac{1}{5} \\ \frac{4}{15} & 0 & 0 & \frac{2}{5} \\ \frac{4}{15} & 0 & 0 & \frac{2}{5} \\ \frac{4}{15} & \frac{2}{5} & 0 & 0 \end{pmatrix}$$

Since $\boldsymbol{r} = \boldsymbol{A}\boldsymbol{r}$ and $\sum_{u \in \{A,B,C,D\}} r_u = 1$, we know

$$\boldsymbol{r} = \left(\frac{3}{7}, \frac{4}{21}, \frac{4}{21}, \frac{4}{21} \right)$$

■

2.2 Teleport Set is $\{\boldsymbol{A}, \boldsymbol{C}\}$

Solution. The adjusted adjacency matrix when $\beta = 0.8$ is

$$\boldsymbol{A} = \beta \boldsymbol{W} + (1-\beta) \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{10} & \frac{1}{2} & \frac{9}{10} & \frac{1}{10} \\ \frac{4}{15} & 0 & 0 & \frac{2}{5} \\ \frac{11}{30} & \frac{1}{10} & \frac{1}{10} & \frac{1}{2} \\ \frac{4}{15} & \frac{2}{5} & 0 & 0 \end{pmatrix}$$

Since $\boldsymbol{r} = \boldsymbol{A}\boldsymbol{r}$ and $\sum_{u \in \{A,B,C,D\}} r_u = 1$, we know

$$\boldsymbol{r} = \left(\frac{27}{70}, \frac{6}{35}, \frac{19}{70}, \frac{6}{35} \right)$$

■

3 Problem 03

Solution. The adjacency matrix is

$$\boldsymbol{A} = \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}_{n\times n}$$

Let the hub and authority vector be $\boldsymbol{h}$ and $\boldsymbol{a}$ respectively. Then we have

$$\boldsymbol{a}^{(0)} = \boldsymbol{h}^{(0)} = \left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \cdots, \frac{1}{\sqrt{n}}\right)_{n\times 1}$$

$$\boldsymbol{a}^{(i+1)} = \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{a}^{(i)} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}_{n\times n} \boldsymbol{a}^{(i)}$$

$$\boldsymbol{h}^{(i+1)} = \boldsymbol{A}\boldsymbol{A}^\top \boldsymbol{h}^{(i)} = \begin{pmatrix} 2 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}_{n\times n} \boldsymbol{h}^{(i)}$$

Thus, the authority vector and the hub vector is

$$\boldsymbol{a}^\infty = \left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \cdots, \frac{1}{\sqrt{n}}\right)_{n\times 1}$$
$$\boldsymbol{h}^\infty = (1, 0, 0, \cdots, 0)_{n\times 1}$$

■

4 Power Iteration

Proof. Suppose $\boldsymbol{M} \in \mathbb{R}^{n\times n}$. Let the eigenvalue of $\boldsymbol{M}$ be $\lambda_1, ...\lambda_n$ ($\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$).

Let the eigenvector releted to λ_i be $\boldsymbol{v}_i$. Then we have $\boldsymbol{M}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i$ for $i = 1, 2, ..., n$.

Suppose $\lambda_1 = \lambda_2 = ... = \lambda_p > \lambda_{p+1}$. We know $\lambda^* \triangleq \lambda_1 = \lambda_2 = ... = \lambda_p$ is the principal eigenvalue of $\boldsymbol{M}$ and $\boldsymbol{v}_1, \boldsymbol{v}_2, ...\boldsymbol{v}_p$ are the principal eigenvectors.

We know all eigenvectors of a matrix are a basis of $\mathbb{R}^n$.

Thus, exists $\boldsymbol{r}^{(0)}$ is a linear combination of $\boldsymbol{v}_1, ...\boldsymbol{v}_n$. Suppose $\boldsymbol{r}^{(0)} = \sum_{i=1}^n \alpha_i \boldsymbol{v}_i$.

Thereby, we have

$$\boldsymbol{M}^k \boldsymbol{r}^{(0)} = \boldsymbol{M}^k \sum_{i=1}^n \alpha_i \boldsymbol{v}_i = \boldsymbol{M}^{k-1} \sum_{i=1}^n \alpha_i \lambda_i \boldsymbol{v}_i = \cdots = \boldsymbol{M} \sum_{i=1}^n \alpha_i \lambda_i^{k-1} \boldsymbol{v}_i = \sum_{i=1}^n \alpha_i \lambda_i^k \boldsymbol{v}_i$$

When $k \to \infty$, we have $\boldsymbol{M}^k \boldsymbol{r}^{(0)} \to (\lambda^*)^k \sum_{i=1}^p \alpha_i \boldsymbol{v}_i$. After unit normalization, we have

$$\begin{aligned}
\boldsymbol{r}^{(\infty)} &= \frac{1}{\sqrt{\sum_{i=1}^p \alpha_i^2}} \sum_{i=1}^p \alpha_i \boldsymbol{v}_i \\
\boldsymbol{M}\boldsymbol{r}^{(\infty)} &= \frac{1}{\sqrt{\sum_{i=1}^p \alpha_i^2}} \sum_{i=1}^p \alpha_i \boldsymbol{M}\boldsymbol{v}_i = \frac{\lambda^*}{\sqrt{\sum_{i=1}^p \alpha_i^2}} \sum_{i=1}^p \alpha_i \lambda^* \boldsymbol{v}_i = \lambda^* \boldsymbol{r}^{(\infty)}
\end{aligned}$$

Since $\|\boldsymbol{r}^{(\infty)}\| = 1$ and $\boldsymbol{M}\boldsymbol{r}^{(\infty)} = \lambda^* \boldsymbol{r}^{(\infty)}$, we know $\boldsymbol{r}^{(\infty)}$ is the principal eigenvector of $\boldsymbol{M}$.

Therefore, the sequence $\boldsymbol{M}\boldsymbol{r}^{(0)}, \boldsymbol{M}^2\boldsymbol{r}^{(0)}, ...\boldsymbol{M}^k\boldsymbol{r}^{(0)}$ approaches the principal eigenvector. $\blacksquare$