AI3602 Data Mining: Homework 4

Xiangyuan Xue (521030910387)

1. According to the graph, the adjacency matrix can be written as

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

The PageRank equation is specified as

$$\boldsymbol{r} = \beta \boldsymbol{M} \boldsymbol{r} + \frac{1-\beta}{n+1} \mathbf{1}$$

where $\boldsymbol{r} \in \mathbb{R}^{n+1}$ is the PageRank vector, $\mathbf{1} \in \mathbb{R}^{n+1}$ is the all-one vector. Notice that the vertices in the clique should have the same PageRank, so we might as well assume

$$\boldsymbol{r} = (x, x, x, \ldots, x, y)^T$$

Substituting this into the PageRank equation, we have

$$\begin{cases} x = \dfrac{\beta(n-1)}{n} x + \dfrac{\beta}{n+1} y + \dfrac{1-\beta}{n+1} \\ y = \beta x + \dfrac{\beta}{n+1} y + \dfrac{1-\beta}{n+1} \end{cases}$$

which yields

$$\begin{cases} x = \dfrac{n}{n^2+n+\beta} \\ y = \dfrac{n+\beta}{n^2+n+\beta} \end{cases}$$

Therefore, we can conclude that the PageRank of the vertices in the clique is $\frac{n}{n^2+n+\beta}$, and the PageRank of the additional vertex is $\frac{n+\beta}{n^2+n+\beta}$.

2. According to the graph, the adjacency matrix can be written as

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

Since the teleport set $S = \{A\}$ and the teleport probability $\beta = 0.8$, we have

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

Assume that $\boldsymbol{r} = (a, b, c, d)^T$. According to the PageRank equation $\boldsymbol{r} = \boldsymbol{A}\boldsymbol{r}$, we have

$$\begin{cases} a = \dfrac{1}{5}a + \dfrac{3}{5}b + c + \dfrac{1}{5}d \\ b = \dfrac{4}{15}a + \dfrac{2}{5}d \\ c = \dfrac{4}{15}a + \dfrac{2}{5}d \\ d = \dfrac{4}{15}a + \dfrac{2}{5}b \end{cases}$$

which yields

$$b = c = d = \frac{4}{9}a$$

Note that the PageRank vector should be normalized, namely

$$a + b + c + d = 1$$

which implies that $a = \frac{3}{7}$, $b = c = d = \frac{4}{21}$. Therefore, the PageRank vector is specified as

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