Homework-1 solutions

Karan Bania

July 2024

1 Link Analysis

Before getting to the answers, let us recall that the Personalized-PageRank (PPR) equation **for user** s is (Slide numbers 36 and 53 from slide 4 here)-

$$r_{s_t} = \beta M r_{s_{t-1}} + (1-\beta)T_s$$

where, r_{s_t} is a vector of dimension $\mathbb{R}^N$ (N is the number of nodes), $M(\in \mathbb{R}^{N\times N})$ is the transition matrix, β is the given teleport parameter, and $T_s(\in \mathbb{R}^N)$ is the teleport vector, i.e., the vector that makes the algorithm "personalized". For example, for the user A, $T_A = \frac{1}{3}[1,1,1,0,\dots]$. It is important to note that, T_s **does not** depend on time, so,

$$\begin{aligned} r_{s_{t+1}} &= \beta M r_{s_t} + (1-\beta)T_s \\ r_{s_{t+1}} &= \beta M(\beta M r_{s_{t-1}} + (1-\beta)T_s) + (1-\beta)T_s \end{aligned}$$

Solving these equations we can get to an expression for r_{s_t} for a user s, which is -

$$\begin{aligned} r_{s_t} &= \beta^t M^t r_0 + (1-\beta)(\sum_{k=0}^{t}(\beta M)^k)T_s \\ r_{s_t} &= a + b \times T_s \end{aligned}$$

Now, given that the teleport parameter is the same for all users and we have ran the algorithm up to some time t, we can see a **linear dependence** on T_s (we also assume that all users have the same base distribution r_0). Thus, if we denote the PPR vectors for uses A, B, C & D as v_A, v_B, v_C & v_D respectively, then we can perform vector manipulations & find the PPR vectors for users with different teleport sets.

$$\begin{aligned} A &= \{1,2,3\}, \\ B &= \{3,4,5\}, \\ C &= \{1,4,5\}, \\ D &= \{1\}. \end{aligned}$$

Different users will only differ in their T_ss so we need to perform operations to change that.

To solve the next three questions, we need to find v_x for some x whose T_s has been given to us, i.e.,

$$v_x = a \times v_A + b \times v_B + c \times v_C + d \times v_D$$

and then solve for a, b, c & d (This is exactly the **vector space** formed by v_A, v_B, v_C & v_D).

This can be further simplified to

$$v_{x_1} = a/3 + c/3 + d \tag{1}$$
$$v_{x_2} = a/3 \tag{2}$$
$$v_{x_3} = a/3 + b/3 \tag{3}$$
$$v_{x_4} = b/3 + c/3 \tag{4}$$
$$v_{x_5} = b/3 + c/3. \tag{5}$$

This already tells us that if $v_{x_4} \neq v_{x_5}$, then we cannot find the PPR vector.

1.1 Personalized PageRank I

Yes, we **can** do this, solving (1),

$$v_{\text{Eloise}} = v_{\{2\}} = 3 * v_A - 3 * v_B + 3 * v_C - 2 * v_D$$

1.2

We **cannot** compute $v_{\text{Felicity}} = v_{\{5\}}$ from the given information. As we cannot isolate the contribution of of 5, or formally, 5 does not lie in the vector space formed by the basis vectors v_A, v_B, v_C & v_D.

1.3

Yes, we **can** do this, solving (1),

$$v_{\text{Glynnis}} = 0.6 * v_A + 0.3 * v_B + 0.3 * v_C - 0.2 * v_D$$

1.4 Personalized PageRank II

Clearly, it is the **vector space defined by vectors in** V.

1.5 A different equation for PageRank

We have to prove,

$$\mathbf{r} = (\beta \mathbf{M} + \frac{(1-\beta)}{N} \mathbf{1}\mathbf{1}^T)\mathbf{r}$$

is equivalent to,

$$\mathbf{r} = \beta \mathbf{M} r + \frac{(1-\beta)}{N} \mathbf{1}$$

OR,

$$\mathbf{1}^T \mathbf{r} = 1$$

We also know that $\mathbf{r}$ is normalized, i.e., $\sum_{i=1}^{N} \mathbf{r_i} = 1$; clearly,

$$\mathbf{1}^T \mathbf{r} = \sum_{i=1}^{N} \mathbf{r_i} = 1.$$

2 Relational Classification I

3 Relational Classification II

4 GNN Expressiveness

5 Node Embedding and its relation to matrix factorization