20667036
Tomas Sousa Pereira

Assignment 2

1.

a.

i)

Each entry in the matrix H means the amount of references that journal j received from each journal i divided by the amount of references given by the journal j, and it is given by the following formula:

$$h_{ij} = \frac{c_{ij}}{c_j}$$

This formula in a ratio on how much a journal is citing my journal, divided by the amount of citations my journal gives.

ii)

Using the formula above and the formula below we computed the matrix H and the impact factor table.

$$\pi_j = \sum_i \pi_i \frac{c_{ij}}{c_j} = \sum_i \pi_i h_{ij}$$

$$\text{H} = \begin{bmatrix} 0 & 8/11 & 0 & 0 \\ 5/8 & 1/11 & 0 & 5/15 \\ 0 & 8/11 & 0 & 0 \\ 0 & 10/11 & 5/8 & 0 \end{bmatrix}$$

Iteration	0	1	2
IF(A)	1/4	0.25*5/8 = 0.15625	0.6136*5/8 = 0.3835
IF(B)	11/4	(0.25*8+0.25*1+ 0.25*8+0.25*10)/11 = 0.6136	(0.15625*8+0.6136*1+ 0.15625*8+0.0.083*10)/11 = 0.3585
IF(C)	1/4	0.25*5/8 = 0.15625	0.083*5/8 = 0.051875
IF(D)	1/4	0.25*5/15 = 0.083	0.6136*5/15 = 0.2045

b.

i)

Each entry of matrix W is weighted number of arbitrary paths from some journal i to journal j, and it is given by the following formula:

$$w_{ij} = \sum_{k=1} (\alpha l_{ij})^k$$

It is the amount of recomendations, that a person has weighted by the required steps to get to a person. e.g. Lets imagine that Bob recomends Alice and Alice recomends Foo. In the end Bob also kind of recomends Foo because of Alice but has less importance than Alice.

ii)

$$\text{W} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.04 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.04 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.04 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.04 & 0.08 & 0.04 & 0 & 0.04 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.04 & 0.04 & 0.04 & 0 & 0.04 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.04 & 0.04 & 0.04 & 0 & 0.04 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.04 & 0.04 & 0.04 & 0 & 0.04 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.04 & 0.04 & 0.04 & 0 & 0.04 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.04 & 0 & 0.04 & 0 & 0.04 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.04 & 0 & 0.04 & 0 & 0.04 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

c.

The Econometrics link matrix is more similar to Pinski and Narin's (PN) model than the Katz's (K) model. Primarily the concept of Pinski and Narin's model it its similiar with the Katz's model, what once was the weight of a reference from another journal divided by the total weights I give now becomes the quantity that some industry produce for my use divided by everything I produce.
Secondarily, in a more mathematics way, the Econometrics model only use node which are directly connected to him, exactly as it is in Pinski and Narin's model while in Katz's model the nodes non directly connected still produce some weight. Another diferences is the entrys in the matrix. In both PN model an Econometrics model each connection has some weight based on the connection meaninngm the reference weight vs the quantity produce. While on the Katz's Model the weigth value is 0 no connection, or $1 * \alpha^k$ (where k is the number of steps to achieve that node). In the end, the concept and the way of Econometrics model matrix is built is similar to the Pinski and Narin's Model.

2.

a.

```
> svd(A)
$d
[1] 4.098872 2.361571 1.273669

$u
            [,1]        [,2]        [,3]
 [1,] -0.4201216 -0.07479925 -0.04597244
 [2,] -0.2994868  0.20009226  0.40782766
 [3,] -0.1206348 -0.27489151 -0.45380010
 [4,] -0.1575610  0.30464762 -0.20064670
 [5,] -0.1206348 -0.27489151 -0.45380010
 [6,] -0.2625606 -0.37944687  0.15467426
 [7,] -0.4201216 -0.07479925 -0.04597244
 [8,] -0.4201216 -0.07479925 -0.04597244
 [9,] -0.2625606 -0.37944687  0.15467426
[10,] -0.3151220  0.60929523 -0.40129339
[11,] -0.2994868  0.20009226  0.40782766

$v
           [,1]       [,2]       [,3]
[1,] -0.4944666 -0.6491758 -0.5779910
[2,] -0.6458224  0.7194469 -0.2555574
[3,] -0.5817355 -0.2469149  0.7749947
```

b.

```
> t(Q) %*% A
      docs
query d1 d2 d3
   q1  1  3  2
```

Figure 1: Original A

```
> t(Q) %*%  decomposeA$u
 %*% diag(decomposeA$d)
 %*% t(decomposeA$v)

query [,1] [,2] [,3]
   q1    1    3    2
```

Figure 2: A computed with decomposed matrices

c.

```
> U2
           [,1]        [,2]
 [1,] -0.4201216 -0.07479925
 [2,] -0.2994868  0.20009226
 [3,] -0.1206348 -0.27489151
 [4,] -0.1575610  0.30464762
 [5,] -0.1206348 -0.27489151
 [6,] -0.2625606 -0.37944687
 [7,] -0.4201216 -0.07479925
 [8,] -0.4201216 -0.07479925
 [9,] -0.2625606 -0.37944687
[10,] -0.3151220  0.60929523
[11,] -0.2994868  0.20009226
```

Figure 3: U_2

```
> S2
         [,1]     [,2]
[1,] 4.098872 0.000000
[2,] 0.000000 2.361571
```

Figure 4: S_2

```
> V2
           [,1]       [,2]
[1,] -0.4944666 -0.6491758
[2,] -0.6458224  0.7194469
[3,] -0.5817355 -0.2469149
> t(V2)
           [,1]       [,2]       [,3]
[1,] -0.4944666 -0.6458224 -0.5817355
[2,] -0.6491758  0.7194469 -0.2469149
```

Figure 5: V_2 & V_2^T

```
> U2 %*% S2 %*% t(V2)
            [,1]        [,2]      [,3]
 [1,]  0.9661565  0.98503618 1.0453788
 [2,]  0.3002301  1.13274606 0.5974388
 [3,]  0.6659264 -0.14770988 0.4479400
 [4,] -0.1477099  0.93469041 0.1980556
 [5,]  0.6659264 -0.14770988 0.4479400
 [6,]  1.1138664  0.05034577 0.8473231
 [7,]  0.9661565  0.98503618 1.0453788
 [8,]  0.9661565  0.98503618 1.0453788
 [9,]  1.1138664  0.05034577 0.8473231
[10,] -0.2954198  1.86938082 0.3961113
[11,]  0.3002301  1.13274606 0.5974388
```

Figure 6: A_2

d.

```
> docs.coord
   coordinates
docs          x          y
  d1 -0.4944666 -0.6491758
  d2 -0.6458224  0.7194469
  d3 -0.5817355 -0.2469149
> query.coord
     coordinates
query          x         y
   q1 -0.2140026 0.1820571
```

Figure 7: Coordinates

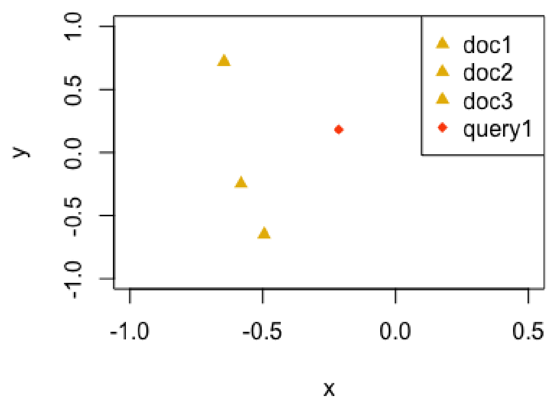

Figure 8: Plot

e.

```
> t(Q) %*% A2

query      [,1]     [,2]     [,3]
   q1 1.118677 3.052473 1.840873
```