

CT5102: Programing for Data Analytics

Problem Sheet –Matrices

1. Given a matrix of exam scores for 10 students:

```
set.seed(10)
N=10
cs1 <- rnorm(N,72,10)
cs2 <- rnorm(N,65,7)
cs3 <- rnorm(N,80,9)
cs4 <- rnorm(N,55,7)
cs5 <- rnorm(N,61,5)

> m
      cs1      cs2      cs3      cs4      cs5
1  72.18746 72.71246 74.63320 42.02382 66.43276
2  70.15747 70.29047 60.33242 54.45438 57.18728
3  58.28669 63.33237 73.92621 61.77996 56.85669
4  66.00832 71.91211 60.92845 56.29448 65.17237
5  74.94545 70.18973 68.61322 45.34039 56.16174
6  75.89794 65.62543 76.63705 44.95140 60.85592
7  59.91924 58.31539 73.81200 57.53461 62.16263
8  68.36324 63.63395 72.15057 42.68639 59.49396
9  55.73327 71.47865 79.08415 52.72819 57.61193
10 69.43522 68.38085 77.71598 50.43906 64.27614
```

Generate the following ranked output using apply functions (the rank function cannot be used).

```
> ans
      cs1 cs2 cs3 cs4 cs5
Student#1  3  1  4 10  1
Student#2  4  4 10  4  8
Student#3  9  9  5  1  9
Student#4  7  2  9  3  2
Student#5  2  5  8  7 10
Student#6  1  7  3  8  5
Student#7  8 10  6  2  4
Student#8  6  8  7  9  6
Student#9 10  3  1  5  7
Student#10  5  6  2  6  3
```

Based on this output, calculate the median rank for each student as follows.

```
      cs1 cs2 cs3 cs4 cs5 median
Student#1  3  1  4 10  1      3
Student#2  4  4 10  4  8      4
Student#3  9  9  5  1  9      9
Student#4  7  2  9  3  2      3
Student#5  2  5  8  7 10      7
Student#6  1  7  3  8  5      5
Student#7  8 10  6  2  4      6
Student#8  6  8  7  9  6      7
Student#9 10  3  1  5  7      5
Student#10  5  6  2  6  3      5
```

2. The aim of this assignment is to gain familiarity with matrices and data frames in R, and also to implement useful functions. Here are the tasks:

- Create a 10x10 matrix to represent connections between people on social media (random seed=100)
- Label the people [A..J], with named rows and columns.
- Randomly populate the matrix with 1s and 0s. The number 1 means someone follows/is followed by another person. Ensure that all diagonals are 0 (you should use an appropriate R matrix operation for this).
- Each row contains information on the people a person follows. For example, {A} follows {C,D,G,J}
- Each column contains information on who follows a person. For example {A} is followed by {D,E,F,H}

Once created, the matrix should have the following entries:

	A	B	C	D	E	F	G	H	I	J
A	0	0	1	1	0	0	1	0	0	1
B	0	0	1	0	0	0	0	0	0	1
C	0	1	0	0	1	0	0	0	0	0
D	1	0	0	0	0	1	1	1	0	0
E	1	0	1	1	0	0	0	1	1	1
F	1	0	0	1	0	0	0	0	0	0
G	0	0	0	1	1	0	0	0	0	0
H	1	1	0	0	0	1	1	0	0	1
I	0	1	1	1	1	0	1	1	0	0
J	0	0	1	0	1	1	0	1	0	0

The following R functions should be written.

- ***follows*** <- ***function (node){}***, which returns the number of nodes an input node follows. For example:

```
> follows("A")
[1] 4
```
- ***followers*** <- ***function (node){}***, which returns the number of nodes that follow an input node. For example:

```
> followers("B")
[1] 3
```
- Both functions should be robust and should ensure that an invalid node is not processed. For example:

```
> followers("b")
Error in followers("b") : Error, node does not exist in Matrix
```

Finally, without using loops, the following data frame should be created using the two functions follows and followers (first 3 rows shown).

```
> df[1:3,]
  Person Follows FollowedBy
1     A       4          4
2     B       2          3
3     C       2          5
```