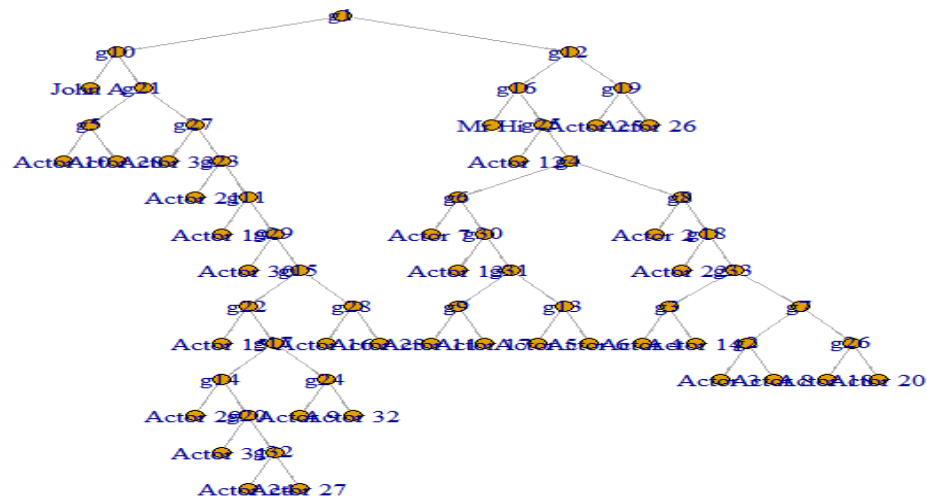Report

1. A) **For karate Dataset, Deleting 5%** of dataset, we get
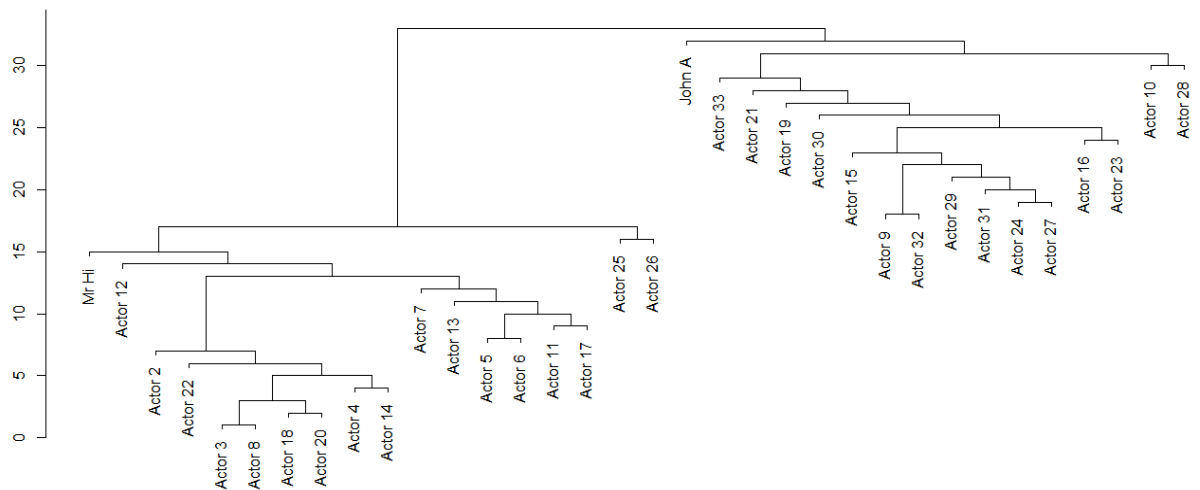
```
> noise_karate <- remove.edges(karate,5)
> E(karate-noise_karate)
+ 4/4 edges from 83b7e76 (vertex names):
[1] Actor 3 --Actor 33 Actor 15--Actor 33 Actor 23--Actor 33 Actor 31--Actor 33
```

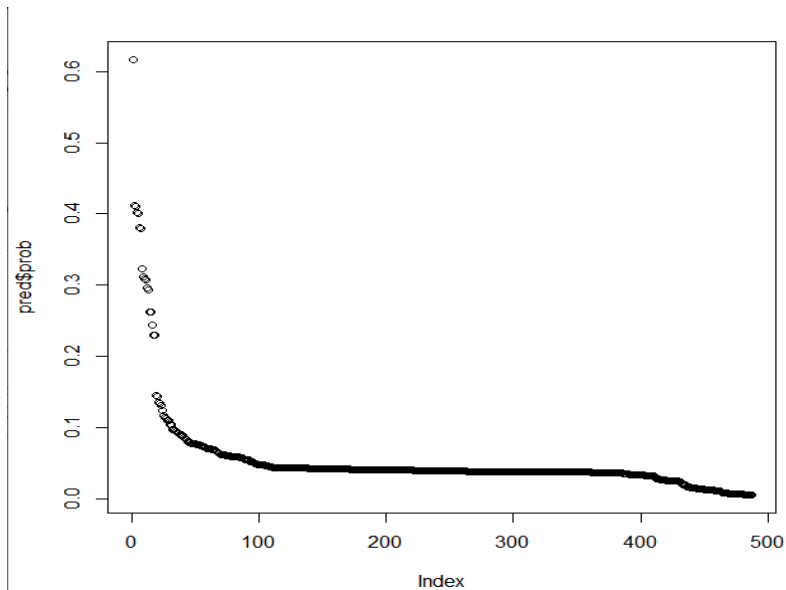Following 4 edges are removed from the dataset.

Hierarchical random graph plot



Dendrogram for Hierarchical random graph with vertices



Prediction Probability graph

From the following image we can see that we are able to predict 3 of the 4 deleted karate dataset edges for 5% noisy dataset in the top 10 edges.

```
> karate_pre_edge <- edges.prediction(noise_karate,noise_karate_nrg)
> karate_pre_edge$edges[1:10,]
        [,1] [,2]
[1,]     1   17
[2,]    25   34
[3,]    26   34
[4,]    27   33
[5,]    28   33
[6,]    31   33
[7,]    29   33
[8,]     2   13
[9,]    23   33
[10,]   15   33
> E(karate-noise_karate)
+ 4/4 edges from a9a6f8b (vertex names):
[1] Actor 3 --Actor 33 Actor 15--Actor 33 Actor 23--Actor 33 Actor 31--Actor 33
>
```

We can see that three edges 1
a)  5 - -33 is present at 10$^{th}$ position.
b)  23 - -33 is present at 9$^{th}$ position.
c)  31 - - 33 is present at 6$^{th}$ position.
    3 of the 4 deleted edges were predicted in top 10 predicted edges for 5% noisy dataset.

Adding top 5 predicted edges to the graph and the connection for the top 5 edges are shown in red in the graph below:
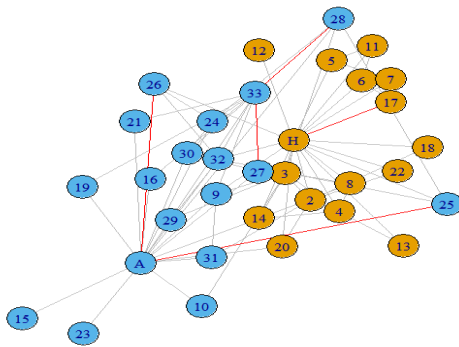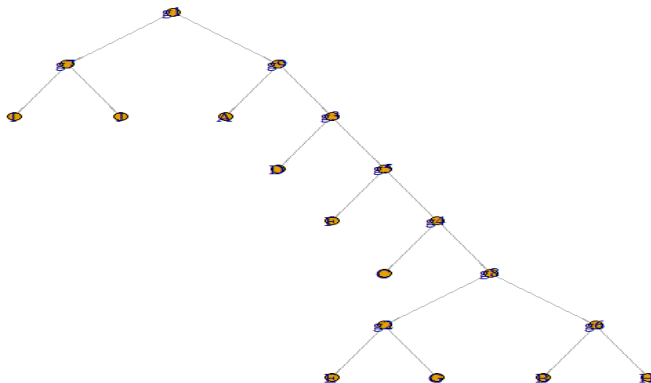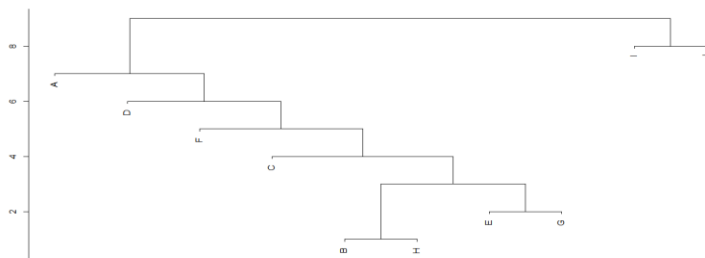
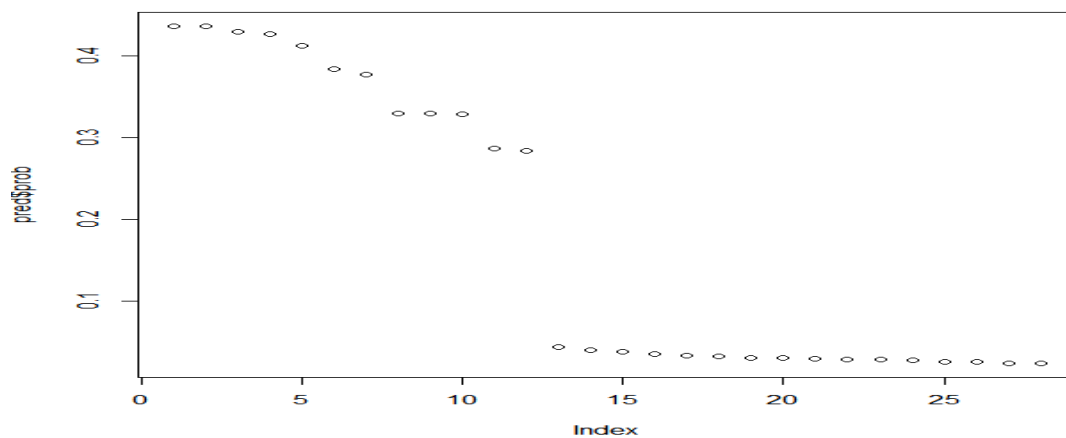b) For **kite dataset with 5% edge deletion**

Deleted edge is

```
> y_diffrn
+ 1/18 edge from 6b7ddad (vertex names):
[1] I--J
```

Hierarchical random graph plot is as follows for 5% noisy dataset.



Dendrogram for dataset:
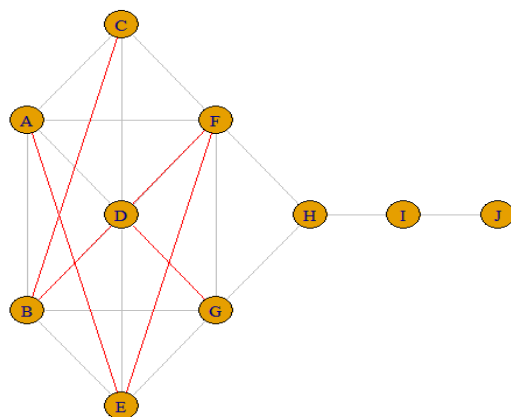
Due to small dataset size of kite ,we are not able to predict the deleted edge

```
> yeast_pre_edge$edges[1:10,]
       [,1] [,2]
[1,]    2    6
[2,]    4    7
[3,]    2    3
[4,]    5    6
[5,]    1    5
[6,]    3    7
[7,]    1    7
[8,]    3    5
[9,]    4    8
[10,]    2    8
> y_diffrn
+ 1/18 edge from 6b7ddad (vertex names):
[1] I--J
> |
```

Adding top 5 predicted edges to the graph and the connection for the top 5 edges are shown in red in the graph below:
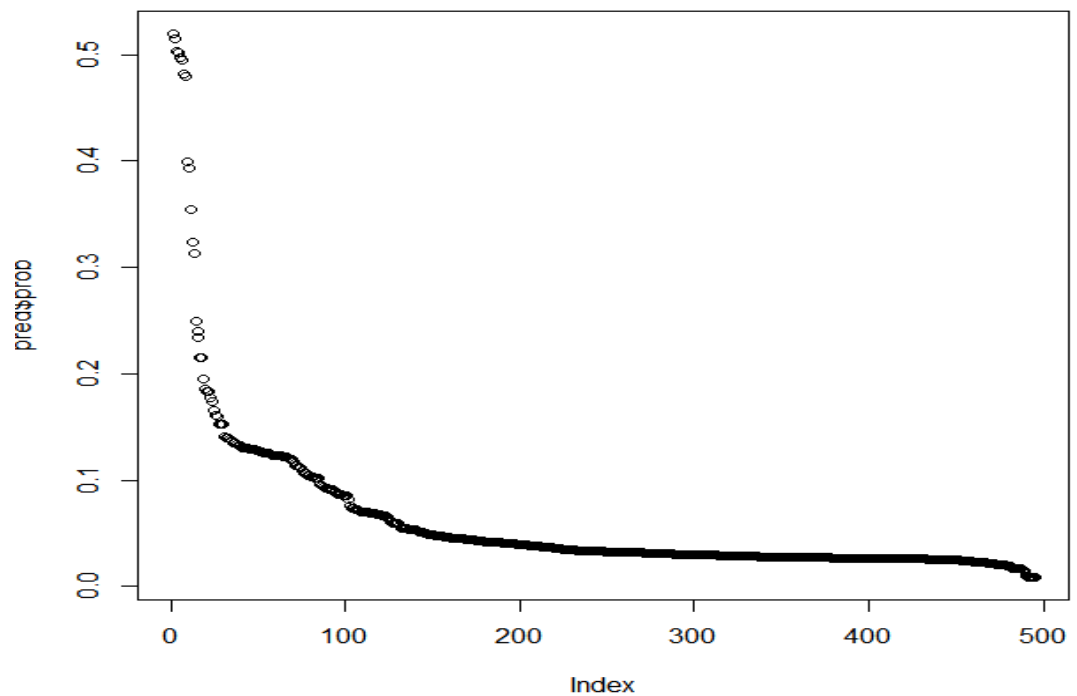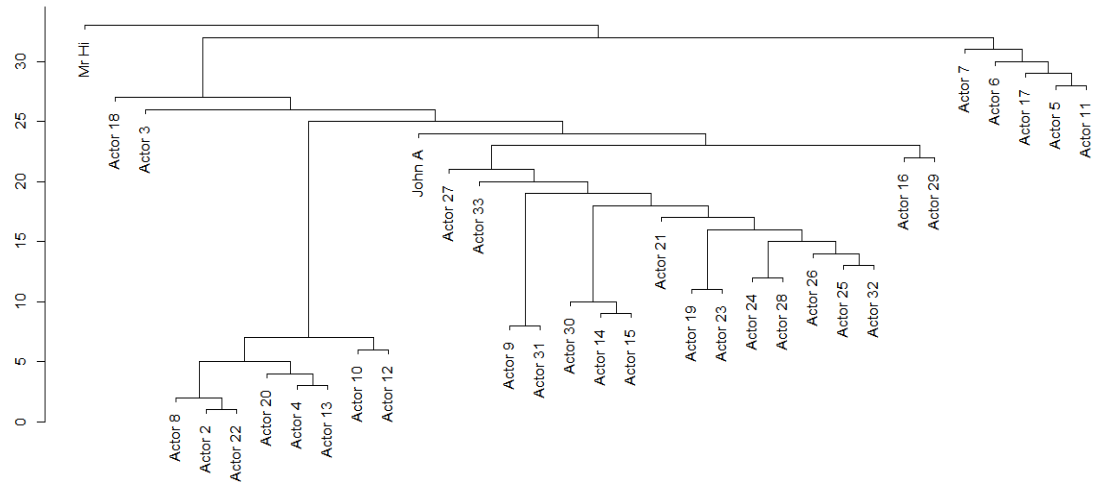
C)
## For **15% deleted Dataset of karate**

Following 12 edges are deleted

```
> E(karate-noise15_karate)
+ 12/12 edges from 6fbc888 (vertex names):
 [1] Mr Hi   --Actor 8  Actor 2 --Actor 31 Actor 2 --Actor 14 Actor 2 --Actor 3  Actor 3 --Actor 28
 [6] Actor 4 --Actor 8  Actor 6 --Actor 17 Actor 6 --Actor 11 Actor 9 --John A   Actor 10--John A
[11] Actor 19--John A   Actor 30--Actor 33
> |
```

Following dendrogram is obtained

We can see that 3 out of 12 deleted edges are in top 10 predicted edges

a) 19 – 34

b) 28 – 33

c) 30 -- 33

```
> E(karate-noise15_karate)
+ 12/12 edges from 6fbc888 (vertex names):
 [1] Mr Hi    --Actor 8  Actor 2 --Actor 31 Actor 2 --Actor 14 Actor 2 --Actor 3  Actor 3 --Actor 28
 [6] Actor 4 --Actor 8  Actor 6 --Actor 17 Actor 6 --Actor 11 Actor 9 --John A   Actor 10--John A
[11] Actor 19--John A   Actor 30--Actor 33
> kar_15_pre_edge$edges[1:10,]
      [,1] [,2]
 [1,]   19   34
 [2,]    1    8
 [3,]   28   33
 [4,]   26   34
 [5,]   29   33
 [6,]   25   34
 [7,]   30   33
 [8,]   27   33
 [9,]   26   33
[10,]   25   33
> |
```

Predicting top5 edges and adding them.
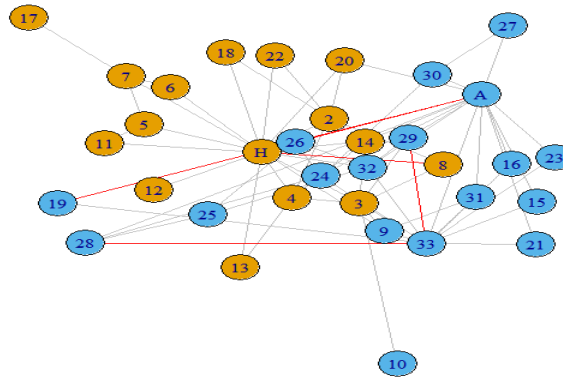


For **40% Dataset of karate**, we have

Following data is deleted

```
> diffrn
+ 31/78 edges from 4b458a1 (vertex names):
 [1] Actor 15--John A   Actor 16--Actor 33 Actor 16--John A   Actor 19--Actor 33 Actor 19--John A
 [6] Actor 20--John A   Actor 21--Actor 33 Actor 21--John A   Actor 23--Actor 33 Actor 23--John A
[11] Actor 24--Actor 26 Actor 24--Actor 28 Actor 24--Actor 30 Actor 24--Actor 33 Actor 24--John A
[16] Actor 25--Actor 26 Actor 25--Actor 28 Actor 25--Actor 32 Actor 26--Actor 32 Actor 27--Actor 30
[21] Actor 27--John A   Actor 28--John A   Actor 29--Actor 32 Actor 29--John A   Actor 30--Actor 33
[26] Actor 30--John A   Actor 31--Actor 33 Actor 31--John A   Actor 32--Actor 33 Actor 32--John A
[31] Actor 33--John A
> |
```

Plotting the Dendrogram , we get
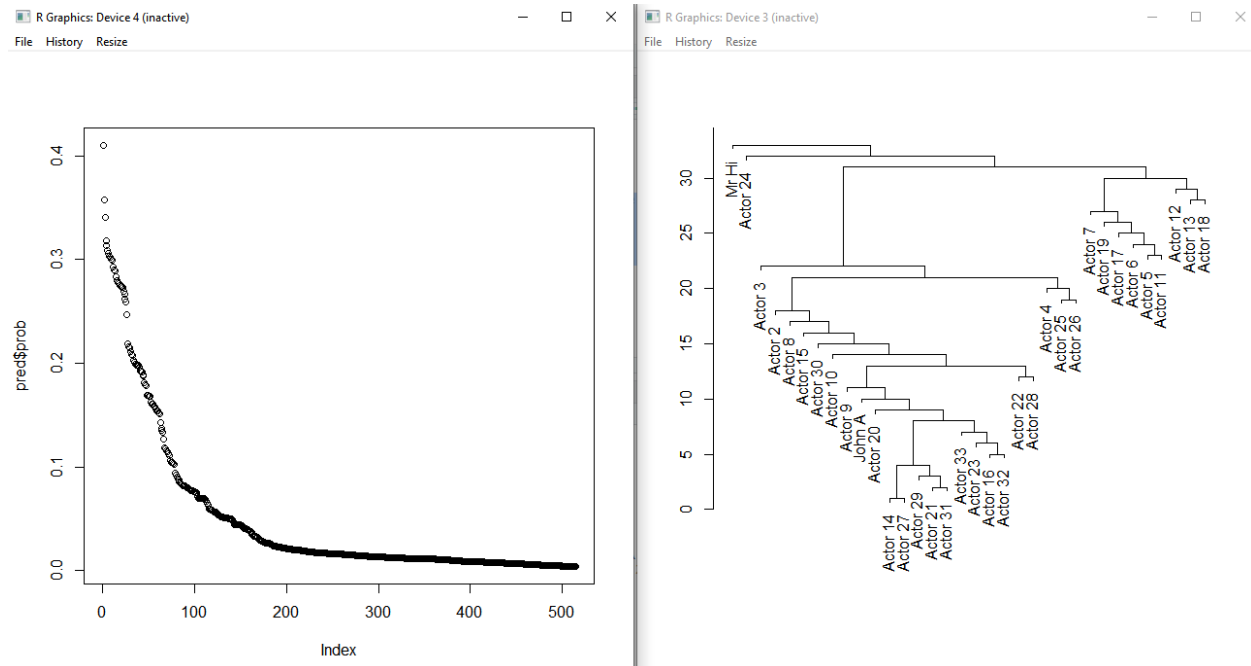
No deleted edges were found in top 10 predicted edges.

```
> diffrn
+ 31/78 edges from 4b458a1 (vertex names):
 [1] Actor 15--John A   Actor 16--Actor 33 Actor 16--John A   Actor 19--Actor 33 Actor 19--John A
 [6] Actor 20--John A   Actor 21--Actor 33 Actor 21--John A   Actor 23--Actor 33 Actor 23--John A
[11] Actor 24--Actor 26 Actor 24--Actor 28 Actor 24--Actor 30 Actor 24--Actor 33 Actor 24--John A
[16] Actor 25--Actor 26 Actor 25--Actor 28 Actor 25--Actor 32 Actor 26--Actor 32 Actor 27--Actor 30
[21] Actor 27--John A   Actor 28--John A   Actor 29--Actor 32 Actor 29--John A   Actor 30--Actor 33
[26] Actor 30--John A   Actor 31--Actor 33 Actor 31--John A   Actor 32--Actor 33 Actor 32--John A
[31] Actor 33--John A
> kar_40_pre_edge$edges[1:10,]
      [,1] [,2]
 [1,]    9   34
 [2,]    1   17
 [3,]    3   31
 [4,]    3   34
 [5,]    2   34
 [6,]    1   14
 [7,]    1   31
 [8,]    3   22
 [9,]    3   16
[10,]    3   23
> |
```
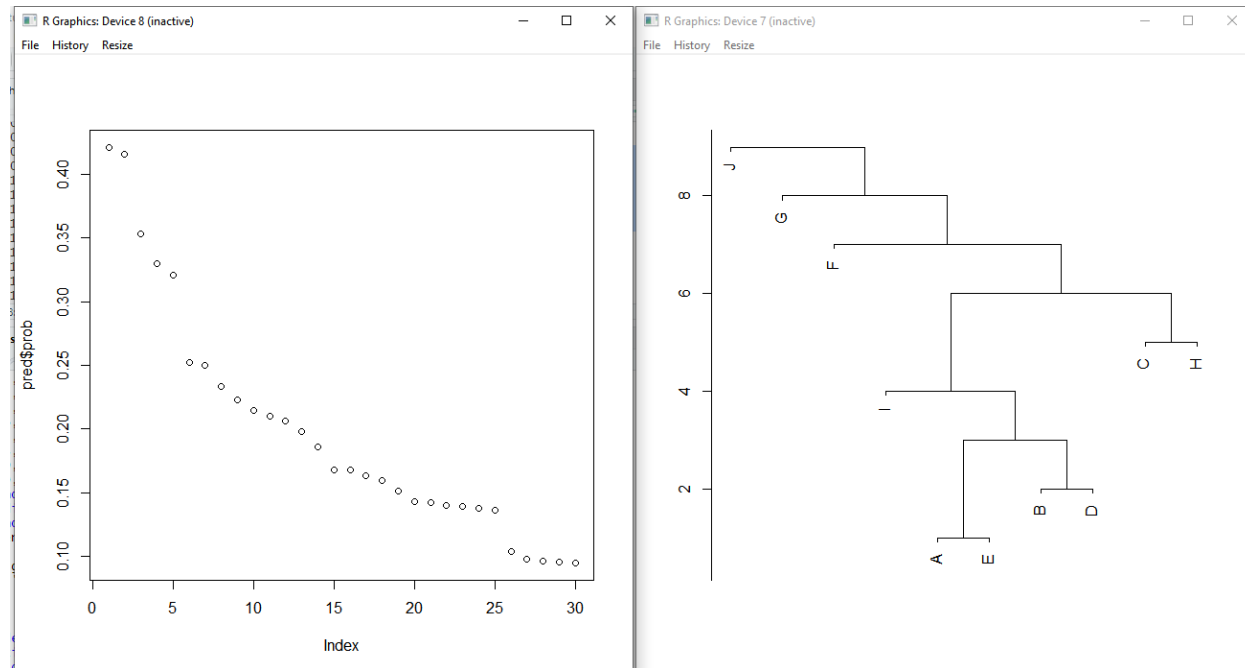
Predicting Top 5 edges and plotting them we get
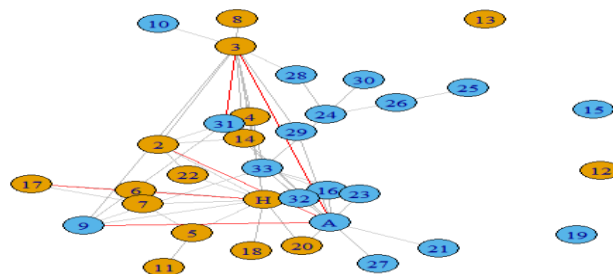


For **15% deleted kite dataset,** we get

Following 3 edges are deleted
```
> diffrn
+ 3/18 edges from 6b7ddad (vertex names):
[1] G--H H--I I--J
> |
```
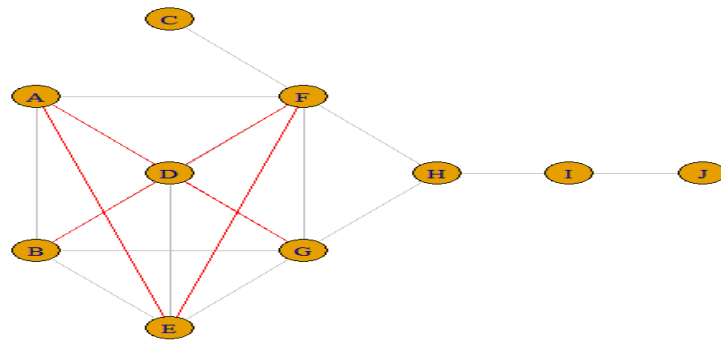Following Dendrogram is obtained:



None of the deleted edges were predicted
```
> diffrn
+ 3/18 edges from 6b7ddad (vertex names):
[1] G--H H--I I--J
> kite_15_pre_edge$edges[1:10,]
      [,1] [,2]
[1,]    5    6
[2,]    2    6
[3,]    4    7
[4,]    1    7
[5,]    1    5
[6,]    6    9
[7,]    2    8
[8,]    4    8
[9,]    5    8
[10,]   6   10
```
Adding top 5 predicted edges is shown below:

We can conclude that as the network datasize decreases it becomes difficult to find the deleted edges as the predicted edges. As is seen in case of kite data
Also we can conclude that as the percent of deleted data increases in the graph the accuracy of the prediction decreases.