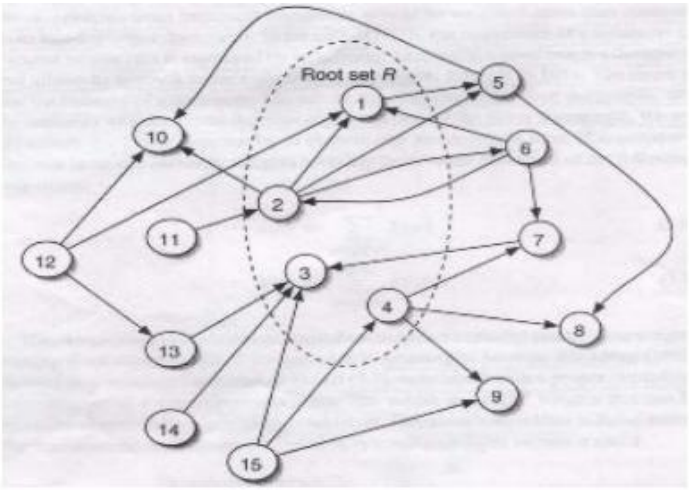
DA Assignment 8: Social Network Analysis

Jingyi Lin (jil173) & Yichao Chen (yic85)

1. Apply the HITS algorithm to the following network (there is no one correct answer)



Root Set R={1,2,3,4}
Extend it to form the base set S

HITS algorithm is an evaluation of the popularity of a search result. The popularity can be got from calculating the hub score and authority score of each page. The hub score is the sum of its link-out pages, and the authority score is the sum of its link-in pages. In the above graph, the root set R includes 1, 2, 3, 4, and by extending, the base set S includes pages from 1 to 15.

First, set initial value of both hub score (h(i)) and authority score (a(i)) to be 1.

$$h(i) = 1$$
$$a(i) = 1$$

Second, apply the calculation equations of hub score and authority score.

$$h(i) = \sum_{(i,j)\in E} a(j)$$

$$a(i) = \sum_{(j,i)\in E} h(j)$$

Third, to iterative the result for many times to get the highest reliable output, the normalizations of the hub score and authority score are necessary.

$$h\,(i) = h\,(i)/|h(i)|$$
$$a\,(i) = a\,(i)/|a(i)|$$

The adjacent matrix is:

0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
1	0	0	0	1	1	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	1	1	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	1	0	0	0	0	0
1	1	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	1	0	0	1	0	0
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	1	0	0	0	0	1	0	0	0	0	0	0

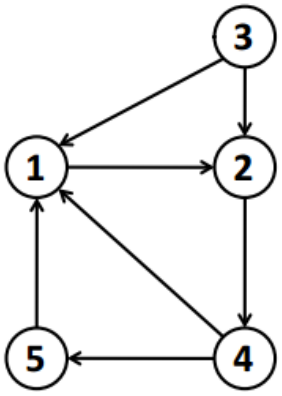
We use R to analysis the orders of authority scores and hub scores.

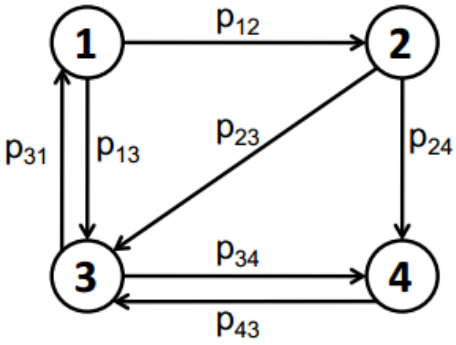
```
> test <- read.table("C:/Users/JYee/Desktop/matrix.txt", header = FALSE)
>
> library(igraph)
>
> matrix <- as.matrix(test)
> matrix
      V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15
 [1,]  0  0  0  0  1  0  0  0  0   0   0   0   0   0   0
 [2,]  1  0  0  0  1  1  0  0  0   1   0   0   0   0   0
 [3,]  0  0  0  0  0  0  0  0  0   0   0   0   0   0   0
 [4,]  0  0  0  0  0  0  1  1  1   0   0   0   0   0   0
 [5,]  0  0  0  0  0  0  0  1  0   1   0   0   0   0   0
 [6,]  1  1  0  0  0  0  1  0  0   0   0   0   0   0   0
 [7,]  0  0  1  0  0  0  0  0  0   0   0   0   0   0   0
 [8,]  0  0  0  0  0  0  0  0  0   0   0   0   0   0   0
 [9,]  0  0  0  0  0  0  0  0  0   0   0   0   0   0   0
[10,]  0  0  0  0  0  0  0  0  0   0   0   0   0   0   0
[11,]  0  1  0  0  0  0  0  0  0   0   0   0   0   0   0
[12,]  1  0  0  0  0  0  0  0  0   1   0   0   1   0   0
[13,]  0  0  1  0  0  0  0  0  0   0   0   0   0   0   0
[14,]  0  0  1  0  0  0  0  0  0   0   0   0   0   0   0
[15,]  0  0  1  1  0  0  0  0  1   0   0   0   0   0   0
>
> g <- graph.adjacency(matrix, mode = "directed")
```

```
> authority <- authority.score(g)$vector
>
>
>
> Authority_Order <- order(authority, decreasing = TRUE)
> Authority_Order
 [1]  1 10  5  6  7 13  8  2  9  3  4 11 12 14 15
>
> hub <- hub.score(g)$vector
> Hub_Order <- order(hub, decreasing = TRUE)
> Hub_Order
 [1]  2 12  6  5  4  1 11 15 13  7 14  3  8  9 10
```

From the above output we can see that the both orders of authority scores and hub scores. For the authority, 1 has the highest score, and for the hub, 2 has the highest score.

2. Find the Hubs and Authorities of the graphs below given by HITS. Are the results consistent with the notions of Hubs and Authorities?





In the first graph:
Hubs: 1, 2, 3, 4, 5
Authorities: 1, 2, 4, 5

In the second graph:
Hubs: 1, 2, 3, 4
Authorities: 1, 2, 3, 4

Refer to the node notion, it is hard to define exactly whether a node is an authority or a hub without some calculations.