

Homework 1

Jing Leng

September 17, 2014

```
# problem 2
# a).
vote <- read.table("Pairwise.dat")
rowSums(vote)
# b).
simmat <- matrix(numeric(0), 4, 4) ## used to store the pairwise similarity
for (i in 1:4) {
  for (j in 1:4) {
    simmat[i, j] = (sum(vote[i,] * vote[j, ]) + sum( (1 - vote[i,]) * (1 - vote[j,]))) / (30) ## simple
  }
}
# c).
place <- c(NA, 0, NA, 1)
place[1] = simmat[1, 2]/(simmat[1, 4] + simmat[1,2])
place[3] = simmat[3, 2]/(simmat[3, 2] + simmat[3, 4])

# problem 3
# a).
rm(list=ls())
load('FrontRange.RData')
stainfo = FR$info
which(stainfo$elev == max(stainfo$elev))
which(stainfo$elev == min(stainfo$elev))
plot(FR$time[[36]], FR$precip[[36]], main = "rain fall records in station of highest elevation")
plot(FR$time[[27]], FR$precip[[27]], main = "rain fall records in station of lowest elevation")
# b).
sum(FR$precip[[36]] >= 10) / length(FR$precip[[3]])
sum(FR$precip[[27]] >= 10) / length(FR$precip[[3]])
```

1.

a). Use one $n \times 1$ (n = number of users) matrix to record the users. The data is name of user (character). -A
Use one $n \times n$ binary matrix to represent friend relations. Each column represents a user, each row represents a user. Each entry represents the relation between two user, 1 in i th column and j th row means that $user_i$ and $user_j$ are friends, 0 means they are not friends. -B

Use one $m \times 3$ (m = number of messages) matrix to record all the messages. Each row is a message, and the columns are the content (character), date (numeric), and the user that send the message (user index, numeric). -C
Use one $m \times n$ binary matrix to represent the “likes” of users. Each column represents a user, each row represents a message. 1 in i th row and j th column means that $user_j$ liked $message_i$. -D

b). Take the submatrix of D where the columns are only the message on the date, 9/10/2014. Sum up all the rows and select the largest one. The column number of the largest sum is the index of the most liked message.
c). Count the number of times a word is used in messages. The most frequently used word has the largest count.

d). Use the name to look up the index of the user in A. Then use table B to look up “Mary Poppins”’s friends. Then use table C to select the messages that are send by Mary’s friends. Take the submatrix to D where user is Mary Poppins and messages are those sent by her friends, count the number when a message is liked and are sent by the same user.

2.

- a). Senator 3 voted “yes” the most (18 times). Senator 1 and 2 voted “no” the most (14 times).
 b). Measurement of similarity: simple match coefficient:

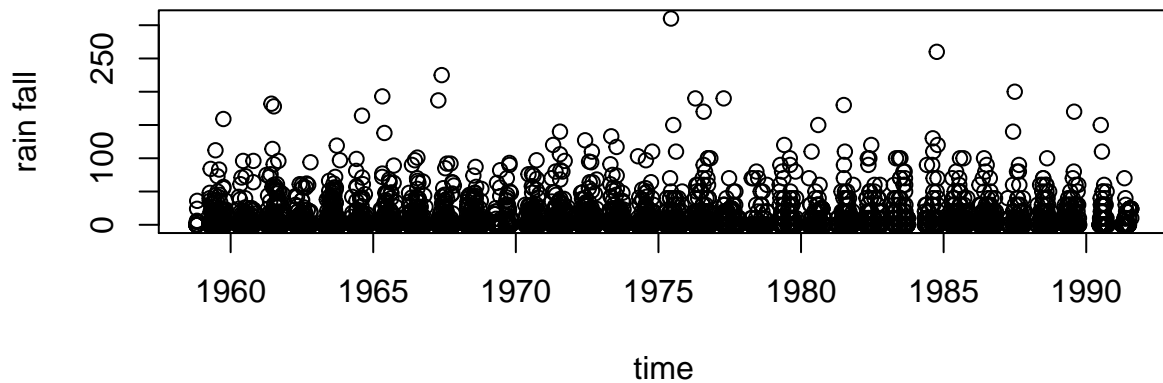
$$\frac{M_{00} + M_{11}}{M_{00} + M_{01} + M_{10} + M_{11}}$$

Senator 2 and Senator 4 has the least similarity. c). Place the least similar pair of senators on the two end of a line. Then based on each other senator’s similarity of them, place them on the middle.

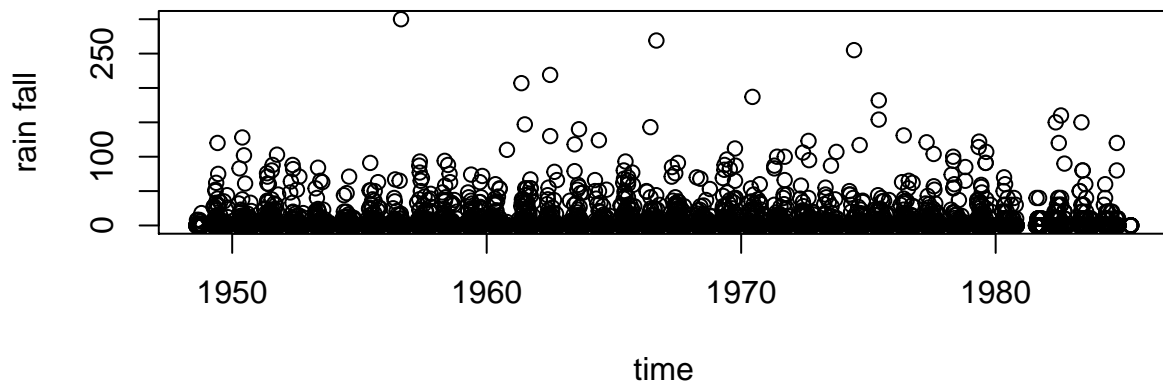
3.

- a). The 36th station has the highest elevation, whereas the 27th station has the lowest.

rain fall records in station of highest elevation



rain fall records in station of lowest elevation



.419 for the 36th station, and 0.215 for the 27th station.

b).