4 Z-score

H Qin

10/26/2018

```r
rm(list=ls())
debug = 1
library(igraph)
```

```
##
## Attaching package: 'igraph'

## The following objects are masked from 'package:stats':
##
##     decompose, spectrum

## The following object is masked from 'package:base':
##
##     union
```

```r
pairs= read.csv("seats.csv", colClasses = c("character", "character"))
nodes = read.csv("students.csv", colClasses = c("character", "character"))
```

Order state,

```r
newpairs = data.frame( t(apply(pairs, 1, sort)) ); #oder id1 and id2
cbind( newpairs, pairs)[1:10,]
```

```
##            X1        X2 person1 person2
## 1       Chris       Tes   Chris     Tes
## 2      Isabel       Tes     Tes  Isabel
## 3      Airuna     Caleb  Airuna   Caleb
## 4      Airuna     Sudha  Airuna   Sudha
## 5       Foram     Sudha   Sudha   Foram
## 6       Foram   Jasmine   Foram Jasmine
## 7       Caleb    Edward  Edward   Caleb
## NA       <NA>      <NA>    <NA>    <NA>
## NA.1     <NA>      <NA>    <NA>    <NA>
## NA.2     <NA>      <NA>    <NA>    <NA>
```

convert long names to 2 letter

```r
pairs = newpairs
names(pairs) = c("name1", "name2")
pairs$cat1 = nodes$Gender[match( pairs$name1, nodes$Names)]
pairs$cat2 = nodes$Gender[match( pairs$name2, nodes$Names )]

tags = t(apply(pairs[,c("cat1", "cat2")], 1, sort))
pairs$tag = paste( tags[,1], tags[,2], sep='_')
F.obs = data.frame( table(pairs$tag))
```

```r
names(F.obs) = c("tag", "freq")
F.obs
```

```
##   tag freq
## 1 F_F    3
## 2 F_M    2
## 3 M_M    2
```

load MS02 null networks

```r
ms02files = list.files(path='MS02')
F.ms02 = data.frame(matrix(data=NA, nrow=1, ncol=3))
names(F.ms02) = c('tag', 'freq', 'file')

# file = "_ms02_seats.1.csv" #debug

for (file in ms02files ){
  if ( debug > 0 ) { print(file) }
  ms02_pairs= read.csv(paste("MS02/", file, sep=''))
  ms02_pairs = ms02_pairs[,1:2]
  ms02_pairs$cat1 = nodes$Gender[match( ms02_pairs$id1, nodes$Names)]
  ms02_pairs$cat2 = nodes$Gender[match( ms02_pairs$id2, nodes$Names)]

  tags2 = t(apply(ms02_pairs[,c("cat1", "cat2")], 1, sort))
  ms02_pairs$tag = paste( tags2[,1], tags2[,2], sep='_')
  F.ms02current = data.frame( table(ms02_pairs$tag))
  F.ms02current$file = file
  names(F.ms02current) = c('tag', 'freq', 'file')
  F.ms02 =  data.frame( rbind(F.ms02, data.frame(F.ms02current)) )
}
```

```
## [1] "_ms02_seats.1.csv"
## [1] "_ms02_seats.10.csv"
## [1] "_ms02_seats.11.csv"
## [1] "_ms02_seats.12.csv"
## [1] "_ms02_seats.13.csv"
## [1] "_ms02_seats.14.csv"
## [1] "_ms02_seats.15.csv"
## [1] "_ms02_seats.16.csv"
## [1] "_ms02_seats.17.csv"
## [1] "_ms02_seats.18.csv"
## [1] "_ms02_seats.19.csv"
## [1] "_ms02_seats.2.csv"
## [1] "_ms02_seats.20.csv"
## [1] "_ms02_seats.21.csv"
## [1] "_ms02_seats.22.csv"
## [1] "_ms02_seats.23.csv"
## [1] "_ms02_seats.24.csv"
## [1] "_ms02_seats.25.csv"
## [1] "_ms02_seats.26.csv"
## [1] "_ms02_seats.27.csv"
## [1] "_ms02_seats.28.csv"
```

```
## [1] "_ms02_seats.29.csv"
## [1] "_ms02_seats.3.csv"
## [1] "_ms02_seats.30.csv"
## [1] "_ms02_seats.31.csv"
## [1] "_ms02_seats.32.csv"
## [1] "_ms02_seats.33.csv"
## [1] "_ms02_seats.34.csv"
## [1] "_ms02_seats.35.csv"
## [1] "_ms02_seats.36.csv"
## [1] "_ms02_seats.37.csv"
## [1] "_ms02_seats.38.csv"
## [1] "_ms02_seats.39.csv"
## [1] "_ms02_seats.4.csv"
## [1] "_ms02_seats.40.csv"
## [1] "_ms02_seats.41.csv"
## [1] "_ms02_seats.42.csv"
## [1] "_ms02_seats.43.csv"
## [1] "_ms02_seats.44.csv"
## [1] "_ms02_seats.45.csv"
## [1] "_ms02_seats.46.csv"
## [1] "_ms02_seats.47.csv"
## [1] "_ms02_seats.48.csv"
## [1] "_ms02_seats.49.csv"
## [1] "_ms02_seats.5.csv"
## [1] "_ms02_seats.50.csv"
## [1] "_ms02_seats.6.csv"
## [1] "_ms02_seats.7.csv"
## [1] "_ms02_seats.8.csv"
## [1] "_ms02_seats.9.csv"
```

```r
F.ms02 = F.ms02[ !is.na(F.ms02$tag), ]
```

Initialize the Z-score matrix

```r
unique_tags = unique( c(as.character(F.obs$tag), as.character(F.ms02$tag)))
Zs = data.frame(unique_tags)
names(Zs) = c('tag')
Zs$freq = F.obs$freq[ match( Zs$tag , F.obs$tag) ]
Zs$freq[is.na(Zs$freq)] = 0;
```

calculate Z-score

```r
for (i in 1 : length(Zs$tag)) {
#i = 2
  sub = F.ms02[ F.ms02$tag == Zs$tag[i], ]
  if(debug>0 ){
     print( paste( Zs$tag[i], "mean:", mean(sub$freq), "sd:", sd(sub$freq) ))
  }
  Zs$Z[i] = Zs$freq[i] - mean(sub$freq) / sd(sub$freq)
}
```

```
## [1] "F_F mean: 2.08 sd: 0.528378343111627"
## [1] "F_M mean: 3.84 sd: 1.05675668622325"
## [1] "M_M mean: 1.2 sd: 0.404519917477945"
```

split the tags

```
#tmp = as.vector(unlist(strsplit(as.character(Zs$tag), split="_")))
#tmp2 = data.frame( matrix( tmp, nrow=10, ncol=2, byrow = T) )
#names(tmp2) = c('c1', 'c2')
#Zs = cbind( Zs, tmp2)
```

generate Z matrix

```
#unique categories
cats = sort( unique( nodes$Gender ))
Zmat = matrix(NA, ncol=length(cats), nrow=length(cats))
names(Zmat) = cats;
colnames(Zmat) = cats;
for (i in 1:length(cats)){#row
  for (  j in 1:length(cats)) { #column
    tmp = sort(c(cats[i], cats[j]))
    mytag = paste(tmp[1], tmp[2], sep="_")
    if (debug ) print (mytag)
    Zmat[i,j] = Zs$Z[ Zs$tag == mytag ]
  }
}
```

```
## [1] "F_F"
## [1] "F_M"
## [1] "F_M"
## [1] "M_M"
```
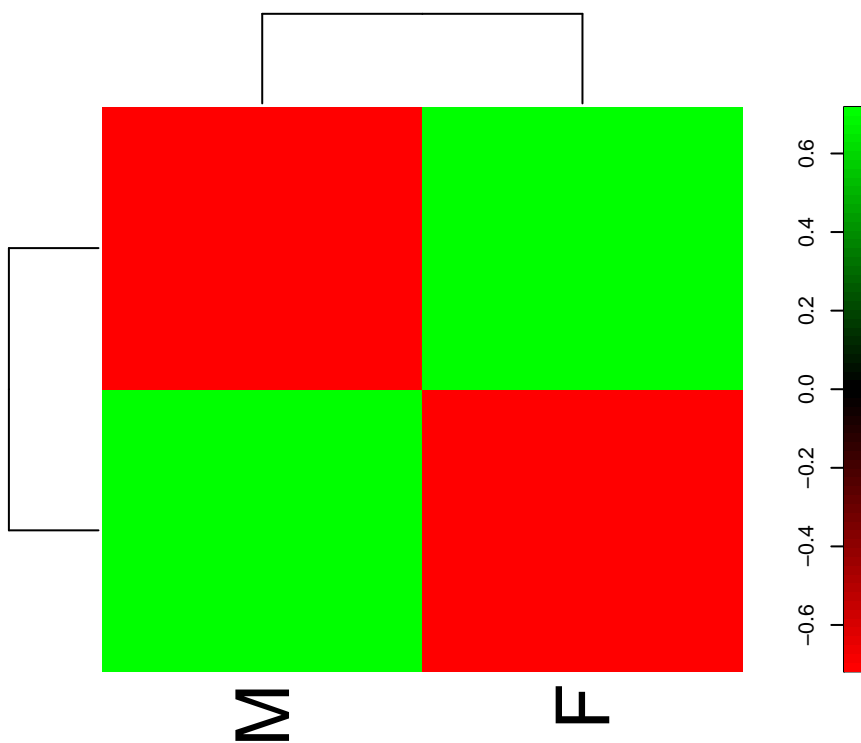
heatmap

```
library(Heatplus)
heatmap_2(Zmat,scale="none", legend=4, do.dendro=c(T,F), col=RGBColVec(64));
```
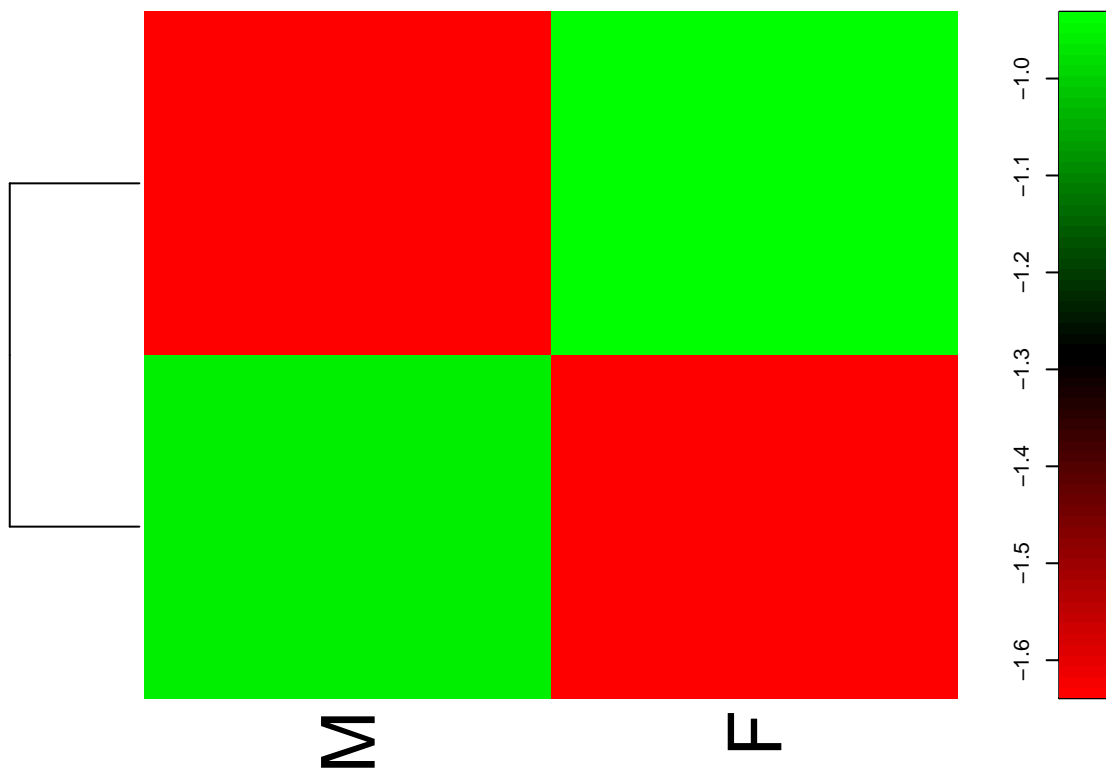
`heatmap_2(Zmat, legend=4, col=RGBColVec(64))`