Homework example

SOBA221 - Social Network teaching team

28/05/2020

RESEARCH QUESTIONS

RQ1*: *Is there sport-based homophily? (Are sporty people more likely to be friends?)

RQ2*: *Do discrepancies in reports of giving and getting study-related information reveal status inequality among students?

```
# Data loading
load('ActorVariables.RData')
load('NetworkVariables.RData')
load('TwoModeVariables.RData')
```

```
# Packages
library(sna)
library(igraph)
library(ggplot2)
```

Selection of the variables that I will be using

```
# RQ1
sport <- ActorVariables$sport_time # Sport time
friend <- NetworkVariables$friends_out_net # Out-friendship network
# Co-factors: gender
gender <- ActorVariables$gender

# RQ2
studyin <- NetworkVariables$study_info_in_net # Study info in
studyout <- NetworkVariables$study_info_out_net # Study info out
pop <- NetworkVariables$popular_net # Popularity
```

RQ 1

Descriptive statistics

In this stage, check the codebook carefully!

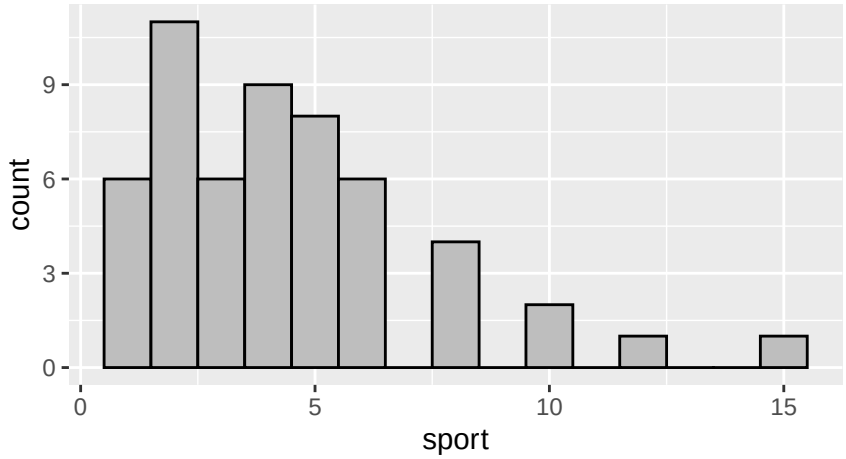
```
sport[sport == -1] <- NA # Prefer not answering (-1) to NA (missing)
sum_sport <- c(mean(sport,na.rm=TRUE),
               sd(sport,na.rm=TRUE),
               range(sport,na.rm=TRUE),
               sum(!is.na(sport)))
```

```r
names(sum_sport) <- c('mean','sd','min','max','n')
sum_sport
```

```
##      mean        sd       min       max         n
##  4.388889  2.910078  1.000000 15.000000 54.000000
```

```r
ggplot()+
  geom_histogram(aes(x=sport),binwidth=1,color='black',fill='grey')
```

```
## Warning: Removed 13 rows containing non-finite values (stat_bin).
```



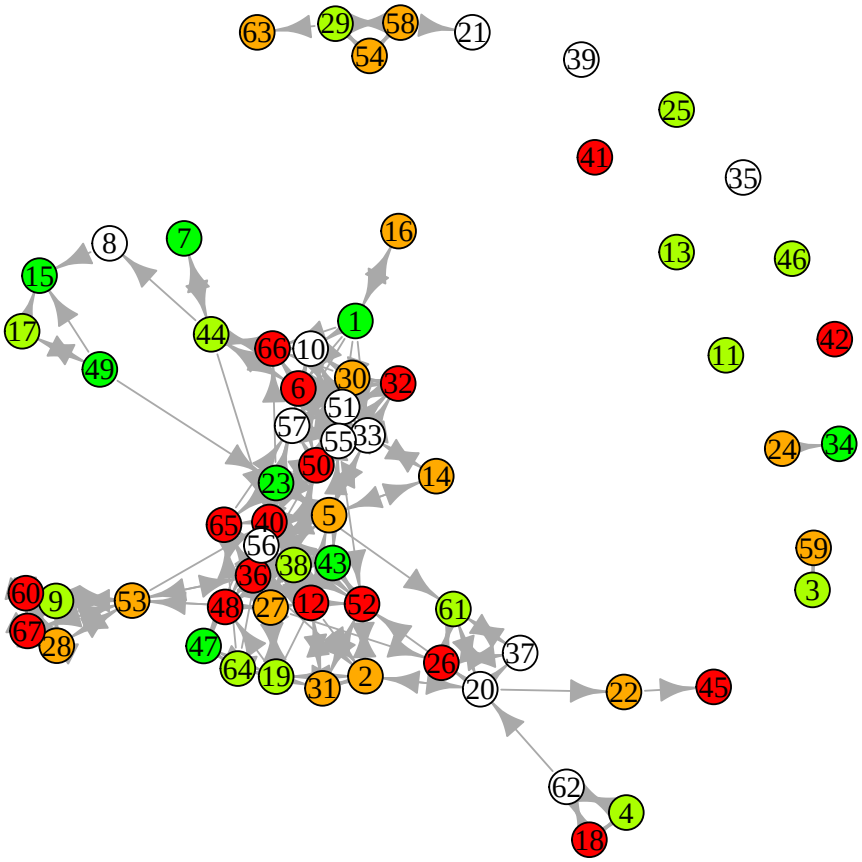

```r
c(density=gden(friend)*100,
  reciprocity=grecip(friend,measure='edgewise')*100,
  transitivity=gtrans(friend)*100)
```

```
##         density reciprocity.Mut    transitivity
##        5.517865       68.852459       44.899738
```

```r
diag(friend) <- NA # I removed the diagonal
rownames(friend) <- 1:nrow(friend)
colnames(friend) <- 1:ncol(friend)

gradient <- colorRampPalette(c('red','yellow','green'))
plot(graph_from_adjacency_matrix(friend,mode='directed'),
     layout=layout_nicely,
     vertex.size = 8.5,
     vertex.color=ifelse(is.na(sport),'white',
                         ifelse(sport %in% 1:2,gradient(4)[1],
                                ifelse(sport %in% 3:4,gradient(4)[2],
                                       ifelse(sport %in% 5:6,gradient(4)[3],
                                              gradient(4)[4])))),
     vertex.label.color='black',
     main='Friendship network coloured by how sporty each student is')
```

Friendship network coloured by how sporty each student is




```
# Gender
gender <- ifelse(gender == 1,'male','female')
gender <- as.factor(gender) # Defined as a factor
summary(gender)
```

```
## female   male
##     46     21
```

Remember that for QAP-family analyses, variables must be matrices (not vectors)

```
# Sport and gender as matrix
sport_diff <- abs(outer(sport,sport,'-')) # Absolute difference
gender_same <- 1*outer(gender,gender,'==') # Same gender
```

Analysis

Is there an association between friendship and sportiness?

```
# Matrix correlation
gcor(friend,sport_diff)
```

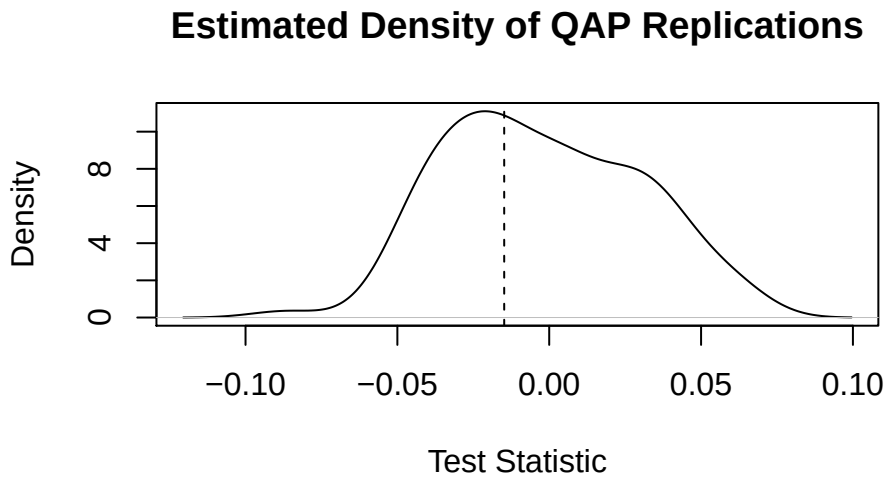
```
## [1] -0.01482314
```

Is it significant?

```
qap_corr <- qaptest(list(friend,sport_diff),
                    gcor,g1=1,g2=2,reps=100)
summary(qap_corr)
```

```
##
## QAP Test Results
##
## Estimated p-values:
##  p(f(perm) >= f(d)): 0.59
##  p(f(perm) <= f(d)): 0.42
##
## Test Diagnostics:
##  Test Value (f(d)): -0.01482314
##  Replications: 100
##  Distribution Summary:
##      Min:     -0.08641616
##      1stQ:    -0.02849747
##      Med:     -0.005166018
##      Mean:    -0.002222553
##      3rdQ:    0.0243975
##      Max:     0.06545579
```

```
plot(qap_corr)
```



What if we adjust for reciprocity and same gender?

```
# QAP regression (friend as a function of reciprocity, gender and sport homophily)
qap_mr <- netlogit(y=friend,x=list(t(friend),gender_same,sport_diff),
```

```
                  nullhyp='qap',reps=100)
print(qap_mr)
```

```
##
## Network Logit Model
##
## Coefficients:
##             Estimate     Exp(b)       Pr(<=b) Pr(>=b) Pr(>=|b|)
## (intercept) -4.437653801   0.01182365 0.00    1.00    0.00
## x1           5.078691854 160.56387750 1.00    0.00    0.00
## x2           0.417416116   1.51803406 0.95    0.05    0.11
## x3           0.005307218   1.00532133 0.62    0.38    0.89
##
## Goodness of Fit Statistics:
##
## Null deviance: 3967.574 on 2862 degrees of freedom
## Residual deviance: 610.0099 on 2858 degrees of freedom
## Chi-Squared test of fit improvement:
##   3357.565 on 4 degrees of freedom, p-value 0
## AIC: 618.0099    BIC: 641.847
## Pseudo-R^2 Measures:
##  (Dn-Dr)/(Dn-Dr+dfn): 0.5398392
##  (Dn-Dr)/Dn: 0.8462512
```

RQ 2

Are there discrepancies in the first place? Check if "get" network is same as transpose of the "give" network:

```
table(studyout,t(studyin))
```

```
##
## studyout    0    1
##        0 4107  146
##        1   79  157
```

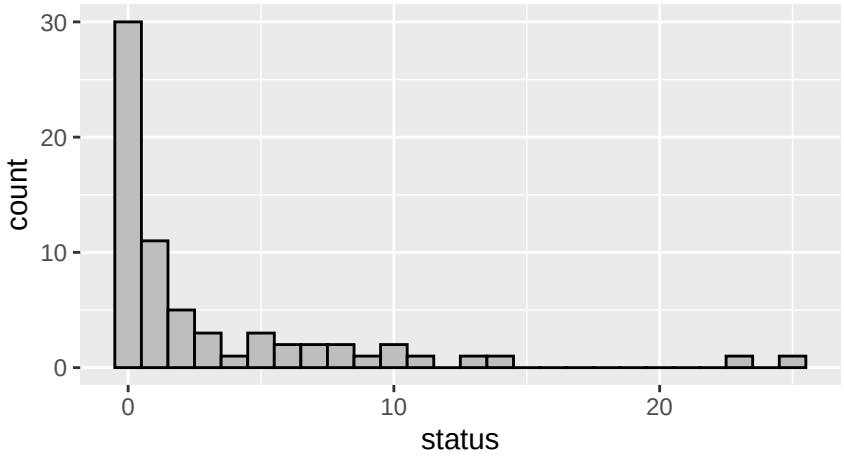
Apparently 146 "get" nominations are not confirmed by the giver, while another 79 "give" nominations are not confirmed by the recipient.

```
# Calculate difference matrix and take the absolute value:
discrepancy <- abs(studyout - t(studyin))
```

The question was if this goes together with status differences. So calculate status as the indegree of each student in the popularity network.

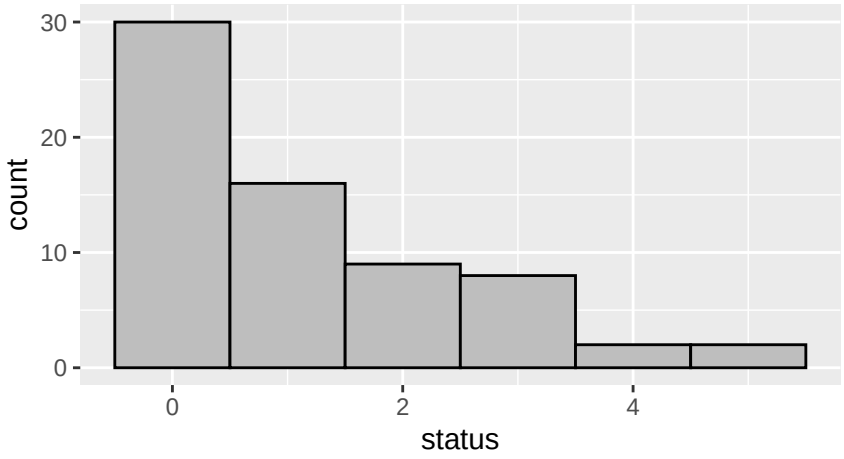
```
status <- sna::degree(pop,cmode="indegree")
```

```
ggplot()+
  geom_histogram(aes(x=status),binwidth=1,color='black',fill='grey')
```




```
# Because this variable in very right-skewed, I will take its square root
status <- sqrt(status)
```

```
ggplot()+
  geom_histogram(aes(x=status),binwidth=1,color='black',fill='grey')
```




```
# And calculate status difference (sender minus receiver):
status_diff <- abs(outer(status,status,"-"))
```

Analysis

Is there an association between discrepancy and status differences

```
gcor(discrepancy,status_diff)
```

```
## [1] -0.007693562
```

Is it significant?

```
qap_corr <- qaptest(list(discrepancy,status_diff),
                    gcor,g1=1,g2=2,reps=100)
summary(qap_corr)
```

```
##
## QAP Test Results
```

```
##
## Estimated p-values:
##  p(f(perm) >= f(d)): 0.61
##  p(f(perm) <= f(d)): 0.39
##
## Test Diagnostics:
##  Test Value (f(d)): -0.007693562
##  Replications: 100
##  Distribution Summary:
##       Min:     -0.0446336
##       1stQ:    -0.01518351
##       Med:     -0.0001169933
##       Mean:    -0.0006231299
##       3rdQ:    0.01232847
##       Max:     0.05050482
```

```
plot(qap_corr)
```