Assignment 1

Huang Yingshan, Kahlbacher Fabian, Wattenberg Yannick, Aggeler Samuel

10/17/2021

Task 1

1)

The probability for observing a tie is equal to the amount of present ties divided by the number of total possible ties. Let $n = \#\text{of nodes}$ then:

$p = \frac{\#\text{of ties}}{\#\text{of possible ties}} = \frac{m}{n \cdot (n-1)} = \frac{174}{1190} \approx 0.146$

2)

The probability of observing a specific type of dyad (mutual, asymmetric and null) is the number of dyads of that type divided by the number of total possible dyads. Let $T = \binom{n}{2} = \#\text{of dyads}$:

$p_M = \frac{M}{T} = \frac{52}{595} \approx 0.087$

$p_A = \frac{A}{T} = \frac{70}{595} \approx 0.118$

$p_N = \frac{N}{T} = \frac{473}{595} \approx 0.795$

3)

Assuming tie independence with $p = 0.146$ as the probability that a tie exists we get the following values:

$p'_M = P[x_{ij} = 1 \cap x_{ji} = 1] \overset{\text{tie ind.}}{=} P[x_{ij} = 1] \cdot P[x_{ji} = 1] = p^2 \approx 0.021$

$p'_A = P[x_{ij} = 1 \cap x_{ji} = 0] + P[x_{ij} = 0 \cap x_{ji} = 1] \overset{\text{tie ind.}}{=} 2 \cdot P[x_{ij} = 1] \cdot P[x_{ji} = 0] = 2 \cdot p \cdot (1-p) \approx 0.25$

$p'_N = P[x_{ij} = 0 \cap x_{ji} = 0] \overset{\text{tie ind.}}{=} P[x_{ij} = 0] \cdot P[x_{ji} = 0] = (1-p)^2 \approx 0.729$

4)

The probability of observing mutual dyads in the network is four times higher than what we would expect with tie independence. Further the probability of observing asymmetric dyads is more than two times higher when assuming tie independence, for null dyads the probability of observation is 6.6% lower when assuming tie independence. Given the differences in probability for mutual and asymmetric dyads we would argue that assuming tie independence would not be reasonable.

```
set.seed(1908)
library(sna)
```

```
## Loading required package: statnet.common
```

```
##
## Attaching package: 'statnet.common'
```

```
## The following objects are masked from 'package:base':
##
##     attr, order
```

```
## Loading required package: network
```

```
##
## 'network' 1.17.1 (2021-06-12), part of the Statnet Project
## * 'news(package="network")' for changes since last version
## * 'citation("network")' for citation information
## * 'https://statnet.org' for help, support, and other information
```

```
## sna: Tools for Social Network Analysis
## Version 2.6 created on 2020-10-5.
## copyright (c) 2005, Carter T. Butts, University of California-Irvine
##  For citation information, type citation("sna").
##  Type help(package="sna") to get started.
```

```
library(network)
obsMat <- as.matrix(read.csv("matrix.csv", header=F)) #header needs to be false, look at observed value
cugRec1 <- cug.test(obsMat, grecip, cmode = "edges", reps = 3000)
cugRec2 <- cug.test(obsMat, grecip, cmode = "dyad.census", reps = 3000)
cugTrans <- cug.test(obsMat, gtrans, cmode = "dyad.census",  reps = 3000)
cugRec1
```

```
##
## Univariate Conditional Uniform Graph Test
##
## Conditioning Method: edges
## Graph Type: digraph
## Diagonal Used: FALSE
## Replications: 3000
##
## Observed Value: 0.8823529
## Pr(X>=Obs): 0
## Pr(X<=Obs): 1
```

```
cugRec2
```

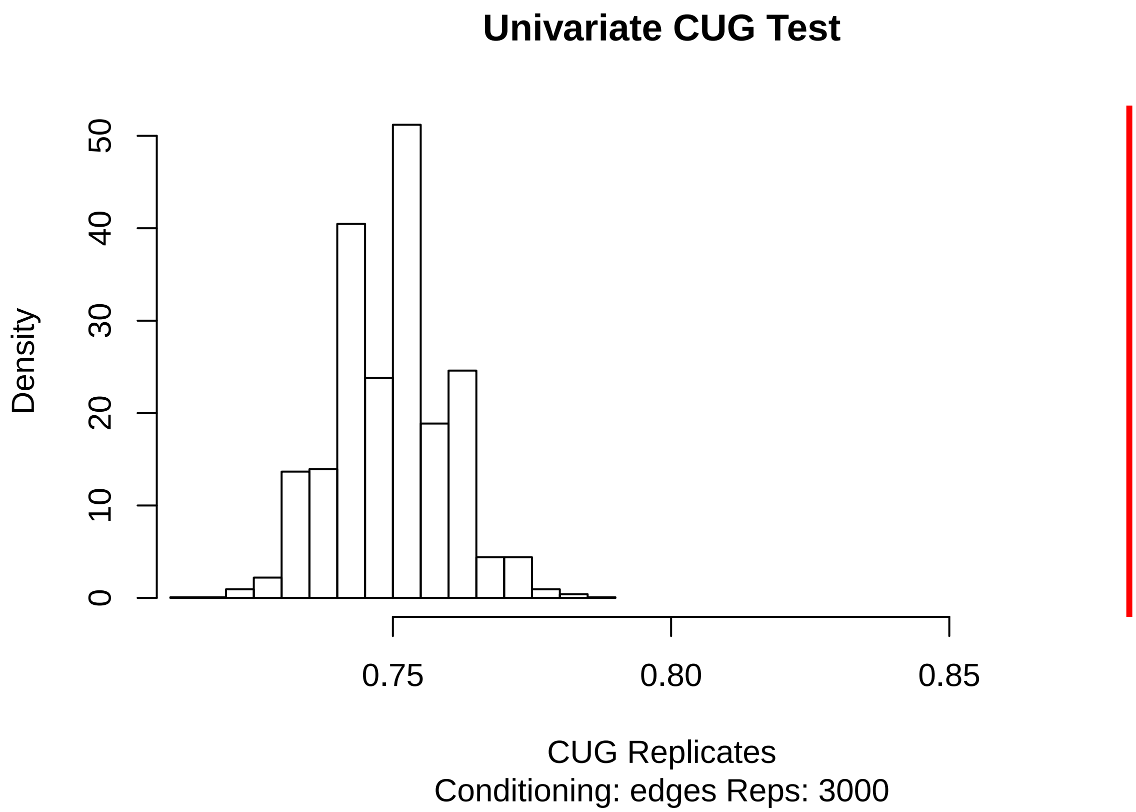
```
##
## Univariate Conditional Uniform Graph Test
##
## Conditioning Method: dyad.census
## Graph Type: digraph
## Diagonal Used: FALSE
## Replications: 3000
##
## Observed Value: 0.8823529
## Pr(X>=Obs): 1
## Pr(X<=Obs): 1
```

```
cugTrans
```

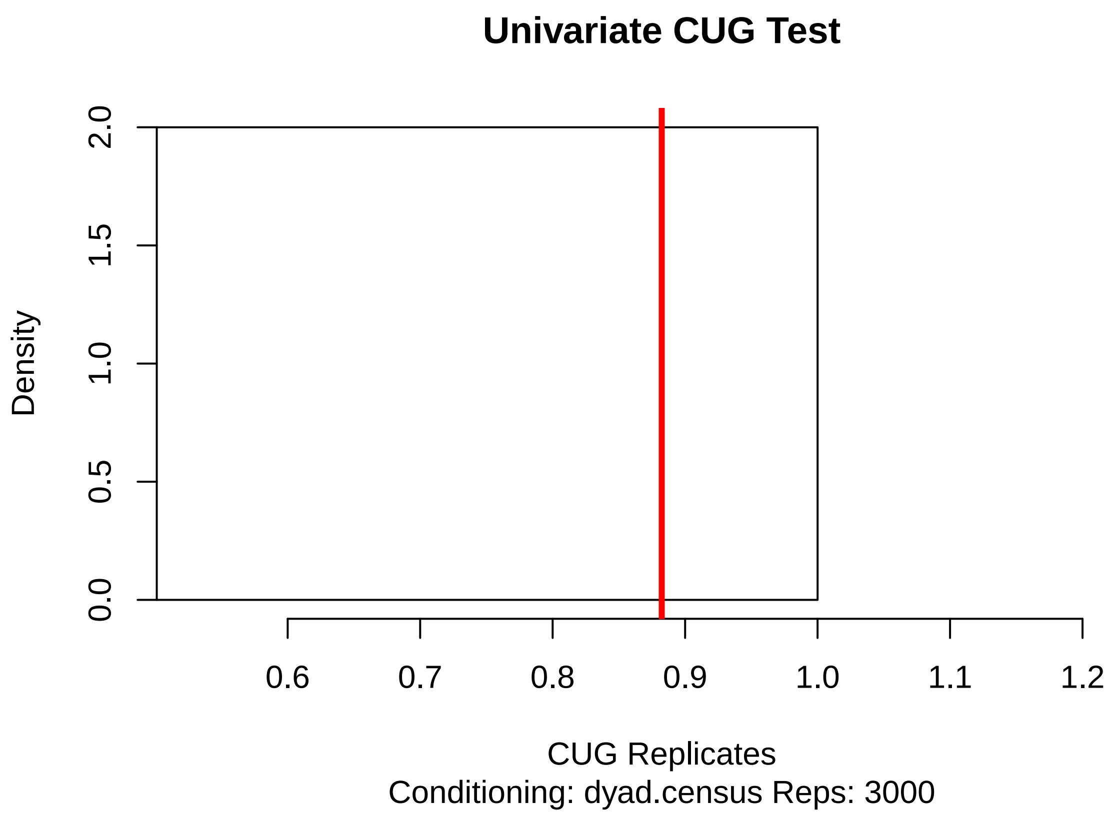
```
##
## Univariate Conditional Uniform Graph Test
##
## Conditioning Method: dyad.census
## Graph Type: digraph
```

```
## Diagonal Used: FALSE
## Replications: 3000
##
## Observed Value: 0.3689482
## Pr(X>=Obs): 0
## Pr(X<=Obs): 1
```

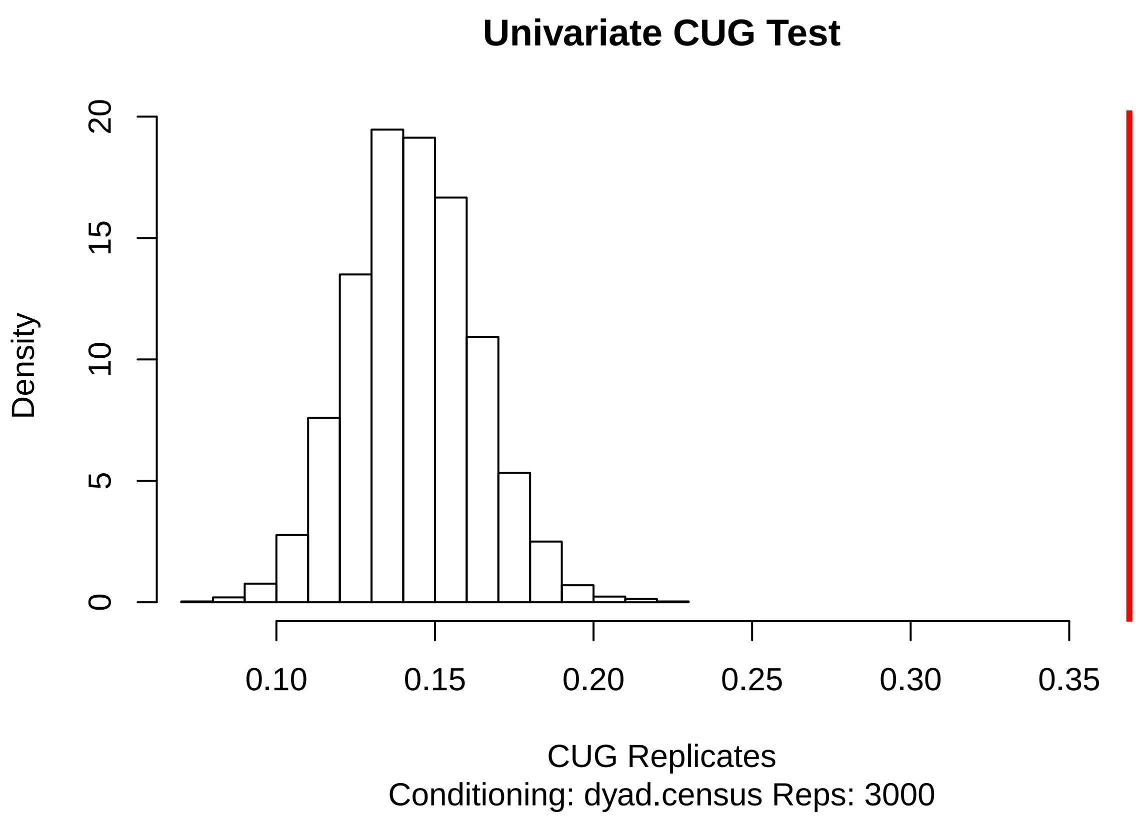
```r
plot(cugRec1)
```




```r
plot(cugRec2)
```




```
plot(cugTrans)
```



```
# look for what we're testing for with grecip (testing for both cases at once)
#need to rephrase the hypothesis
```

5)

5.1)

The hypothesis of the first test is testing whether there is more or less reciprocity then density would suggest. The test is conditioned on the edges that means on the network density or number of edges in the network or $U|p$ where p denotes the network density.

The hypothesis of the second test is the same as for the first test but here we condition on the dyad census. Conditioning on the dyad census here means we condition on the count mutual, asymmetric, and null dyads or $U|MAN$

The hypothesis of the last test is that the amount of transitive relations (ties) that is observed is higher or lower than the amount expected conditioned on the dyad census.

5.2)

We think that the second test doesn't make sense because we are testing for the reciprocity of ties conditioned on dyad census. But the dyad census contains the proportion of mutual ties which is equal to the reciprocity of the network by definition. This is also shown in the test results showing we get exactly the value that we observed.

5.3)

For the first hypothesis we can see that there is a strong indication that our network has more reciprocity then the density of our graph would let assume.

For the second hypothesis there is no significant indication that our graph has more reciprocity then the dyad census would let you assume.

For the third hypothesis there is also significant indication that our network has more transitivity then the dyad census would let you assume.

Task 2

1)

```
library(sna)
library(network)
```

```
set.seed(1908)
trade2006 <- as.matrix(read.csv("trade1.csv", header=FALSE))
trade2007 <- as.matrix(read.csv("trade2.csv", header=FALSE))
attr <- read.csv("attr.csv")

permutations <- 5000

res_netlm <- netlm(y=trade2007, x=trade2006, reps=permutations, nullhyp="qapspp", intercept=T)
res_netlm$names <- c("intercept", "trade_06")

#summary(res_netlm)

res_nice <- cbind(res_netlm$coefficients, res_netlm$pgreqabs)
```

```r
colnames(res_nice) <- c("Est.", "p-value")
rownames(res_nice) <- res_netlm$names
res_nice
```

```
##                 Est. p-value
## intercept 2.1000241   0e+00
## trade_06  0.5646105   2e-04
```

The fitted model is: $X = 2.1 + 0.5646 * trade\_06 + \epsilon$ where trade_06 is the adjacency matrix containing the amount of goods traded by company i to company j in 2006. The residual permutation method double semi-partialing (Dekker et al., 2007) was used to obtain empirical p-values. The intercept and slope (trade_06) are statistically significant on the significance level $alpha = 0.05$. Therefore, the data indicate that both parameters are different from 0. The relationship is positive which means that for each million CHF which is traded in 2006 the amount of goods traded in 2007 increases by 0.5646 million CHF.

2)

```r
# Hp 1
sector <- attr[,2]
same_sector <- outer(sector,sector,"==")*1

# Hp 2
sector_sender_private <- matrix(sector,18,18,byrow=FALSE)
```

3)

```r
zm <- list(same_sector, sector_sender_private)
res_netlm <- netlm(trade2007, zm, reps=permutations, nullhyp="qapspp")
res_netlm$names <- c("intercept", "same_sector", "sector_sender_private")
summary(res_netlm)
```

```
##
## OLS Network Model
##
## Residuals:
##         0%        25%        50%        75%       100%
## -3.8134237 -1.4197348 -0.1565792  1.2402652  5.7000309
##
## Coefficients:
##                       Estimate   Pr(<=b) Pr(>=b) Pr(>=|b|)
## intercept              3.8099691 1.0000  0.0000  0.0000
## same_sector           -1.1002344 0.0028  0.9972  0.0028
## sector_sender_private  0.1434546 0.6336  0.3664  0.6886
##
## Residual standard error: 1.907 on 303 degrees of freedom
## Multiple R-squared: 0.0808   Adjusted R-squared: 0.07473
## F-statistic: 13.32 on 2 and 303 degrees of freedom, p-value: 2.864e-06
##
##
## Test Diagnostics:
##
##  Null Hypothesis: qapspp
##  Replications: 5000
##  Coefficient Distribution Summary:
```

```
##
##        intercept same_sector sector_sender_private
## Min     -0.59091    -4.78141              -3.86412
## 1stQ     1.82437    -0.74722              -0.75178
## Median   2.50062     0.02702               0.09409
## Mean     2.81475    -0.01884               0.03545
## 3rdQ     3.52118     0.82489               0.84962
## Max     16.15295     3.48752               2.89779
```

The parameter "same_sector" is significantly different from 0 at the significance level $\alpha = 0.05$. The parameter indicates that a company which sells goods to another company in the same sector sells 1.1 million CHF less than to another company which is not in the same sector, holding all other variables constant. Thus, the data does not support Hp. 1.

The paramter "sector_sender_private" is not significantly different from 0 at the significance level $\alpha = 0.05$. It indicates that companies in the private sector sell 0.14 million CHF more goods than companies which are in the public sector but this relationship is not statistically significant. Therefore, the data does not support Hp 2.

4)

Hp. 3 (size of companies): If the number of employees of a company increases then the amounts of goods sold increases too.

Hp. 4 (region): The amounts of goods sold by a company is higher if the company which sells is in region A or B than the amount of goods sold by a company which is in region C.

```r
# Hp 3.
size_sender <- matrix(attr$size,18,18,byrow=FALSE)

# Hp.4
region_AB <- (attr$region == 1 | attr$region == 2)*1
region_AB_sender <- matrix(region_AB,18,18,byrow=FALSE)
```

5)

```r
zm <- list(same_sector, sector_sender_private, size_sender, region_AB_sender)
res_netlm <- netlm(trade2007, zm, reps=permutations, nullhyp="qapspp")
res_netlm$names <- c(
  "intercept",
  "same_sector",
  "sector_sender_private",
  "size_sender",
  "region_AB_sender"
)
summary(res_netlm)
```

```
##
## OLS Network Model
##
## Residuals:
##          0%         25%         50%         75%        100%
## -3.80566085 -1.43468230 -0.07047998  1.16803731  5.25321048
##
## Coefficients:
##                        Estimate    Pr(<=b) Pr(>=b) Pr(>=|b|)
```

```
## intercept              2.74003978 1.0000  0.0000  0.0000
## same_sector           -1.10023438 0.0028  0.9972  0.0028
## sector_sender_private -0.05618166 0.4020  0.5980  0.8540
## size_sender            0.03889102 0.9956  0.0044  0.0074
## region_AB_sender      -0.01383804 0.4782  0.5218  0.9362
##
## Residual standard error: 1.895 on 301 degrees of freedom
## Multiple R-squared: 0.09807  Adjusted R-squared: 0.08609
## F-statistic: 8.183 on 4 and 301 degrees of freedom, p-value: 2.823e-06
##
##
## Test Diagnostics:
##
##  Null Hypothesis: qapspp
##  Replications: 5000
##  Coefficient Distribution Summary:
##
##        intercept same_sector sector_sender_private size_sender region_AB_sender
## Min    -4.742193   -4.811798             -3.249439   -3.098245        -2.099121
## 1stQ   -0.695143   -0.768736             -0.584949   -0.637282        -0.539179
## Median  0.174975    0.025341              0.064552    0.041578        -0.017382
## Mean    0.195040   -0.022846              0.005672    0.018552        -0.012559
## 3rdQ    1.060384    0.830414              0.615833    0.662985         0.484512
## Max     4.633423    3.438551              2.662971    2.909934         2.184282
```

The parameter "size_sender" is significantly different from 0 at the significance level $\alpha = 0.05$. The parameter indicates that for each additional employee the company sells 0.039 million CHF more goods. The data supports Hp 3.

The parameter "region_AB_sender" indicates that companies which operate in the region A or B sell 0.014 million CHF less goods than companies which operate in region C. Since the parameter is not statistically significant from 0 at the significance level $\alpha = 0.05$ the data does not support Hp 4.