

HW3

Zihan Lin

2024-10-24

```
library(sna)
library(network)
library(networkdata)
library(tidyverse)
library(ergm)
data(coldwar)
data(hansell)
```

Question 1

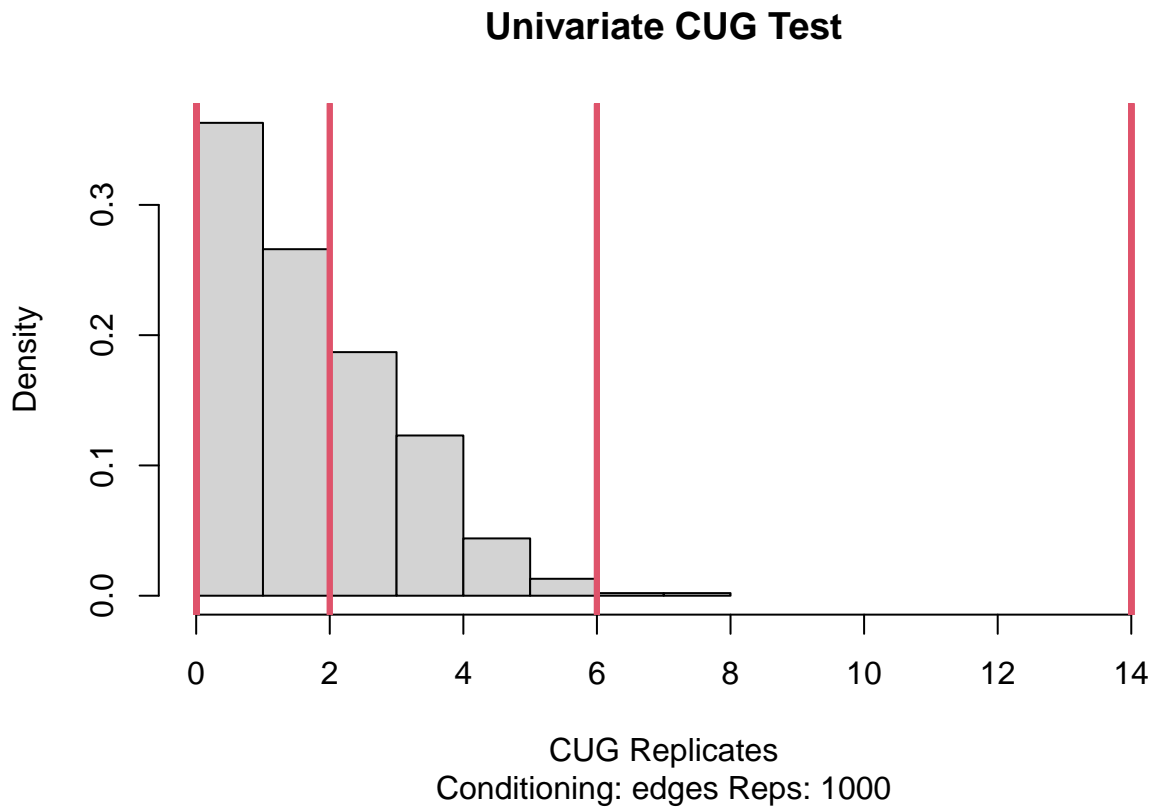
```
coldwarcc <- coldwar$cc[,1]
coldwarcc[coldwarcc < 0] <- 0
coldwarcc[is.na(coldwarcc)] <- 0

network_cw <- network(coldwarcc, directed = FALSE)
degree_cw <- sna::centralization(dat = network_cw, FUN = sna::degree)
degree_cw
```

```
## [1] 0.09423077
```

since the value of degree centralization is 0.0942308, the null hypothesis cannot be rejected, which means it is a simple random graph model.

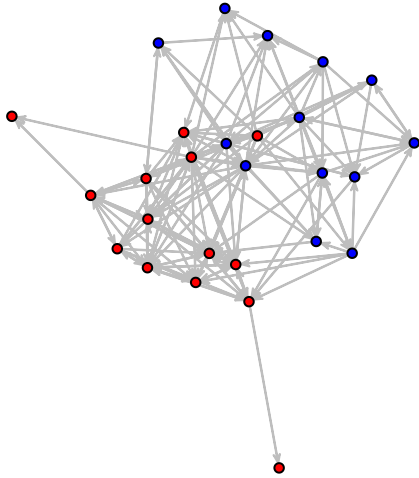
```
cug_test <- cug.test(dat = network_cw, FUN = degree, cmode = "edges", reps = 1000)
plot(cug_test)
```



Question 2

a)

```
plot(hansell,  
     vertex.col = ifelse(get.vertex.attribute(hansell, "sex") == "male", "blue", "red"),  
     edge.col = "grey")
```



based on the plot, there is a general preference for same-sex friendship ties, it is easy to see that blue nodes are predominantly connected the other blue nodes, same for red nodes which represents girls.

b)

```
fit.er <- ergm(hansell ~ edges)

## Starting maximum pseudolikelihood estimation (MPLE):
## Obtaining the responsible dyads.
## Evaluating the predictor and response matrix.
## Maximizing the pseudolikelihood.
## Finished MPLE.
## Evaluating log-likelihood at the estimate.
summary(fit.er)

## Call:
## ergm(formula = hansell ~ edges)
##
## Maximum Likelihood Results:
##
##      Estimate Std. Error MCMC % z value Pr(>|z|)
## edges -1.24454    0.09058      0 -13.74  <1e-04 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      Null Deviance: 973.2  on 702  degrees of freedom
##      Residual Deviance: 746.2  on 701  degrees of freedom
##
## AIC: 748.2  BIC: 752.8  (Smaller is better. MC Std. Err. = 0)
log_odds_b <- -1.24454
cat("the log-odds of a tie between two randomly chosen students is", log_odds_b)

## the log-odds of a tie between two randomly chosen students is -1.24454
probability_b <- exp(log_odds_b) / (1 + exp(log_odds_b))
cat("the MLE of the probability of a tie between two randomly chosen students is",
    probability_b)

## the MLE of the probability of a tie between two randomly chosen students is 0.2236467
hansell

## Network attributes:
##   vertices = 27
##   directed = TRUE
##   hyper = FALSE
##   loops = FALSE
##   multiple = FALSE
##   total edges= 157
##   missing edges= 0
##   non-missing edges= 157
##
## Vertex attribute names:
##   sex
##
## No edge attributes
```

```

vertices <- 27
tot_edges <- 157
density_b <- tot_edges / (vertices * (vertices - 1) / 2)
cat("the density of the network is", density_b)

## the density of the network is 0.4472934

# H_0: The null hypothesis is that all networks among the students are equally likely
# H_1: The alternative hypothesis is that the Erdos-Renyi model fits the data better

Null_Deviance <- 973.2
Residual_Deviance <- 746.2
p_value_b <- 1 - pchisq(Null_Deviance - Residual_Deviance, df = 1)
p_value_b

## [1] 0

```

the MLE probability under the Erdos-Renyi model is lower than the actual observed density of ties in the network (MLE = 0.2236467, density = 0.4472934) and the p-value is 0, suggesting that the Erdos-Renyi model might underestimate the number of ties in this specific network.

c)

```
fit.homo <- ergm(hansell~edges + nodematch("sex"))
```

```
## Starting maximum pseudolikelihood estimation (MPLE):
```

```
## Obtaining the responsible dyads.
```

```
## Evaluating the predictor and response matrix.
```

```
## Maximizing the pseudolikelihood.
```

```
## Finished MPLE.
```

```
## Evaluating log-likelihood at the estimate.
```

```
summary(fit.homo)
```

```
## Call:
```

```
## ergm(formula = hansell ~ edges + nodematch("sex"))
```

```
##
```

```
## Maximum Likelihood Results:
```

```
##
```

	Estimate	Std. Error	MCMC %	z value	Pr(> z)
edges	-1.9841	0.1608	0	-12.341	<1e-04 ***
nodematch.sex	1.2954	0.1979	0	6.547	<1e-04 ***

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Null Deviance: 973.2 on 702 degrees of freedom
```

```
## Residual Deviance: 699.1 on 700 degrees of freedom
```

```
##
```

```
## AIC: 703.1 BIC: 712.2 (Smaller is better. MC Std. Err. = 0)
```

since p-value for nodematch.sex is <1e-04, which is highly significant result that indicates strong evidence of homophily based on sex

```

log_odds_c <- -1.9841 + 1.2954
cat("the log-odds of a tie between two students of the same sex is", log_odds_c)

## the log-odds of a tie between two students of the same sex is -0.6887
probability_c <- exp(log_odds_c) / (1 + exp(log_odds_c))
cat("the MLE of the probability of a tie between two students of the same sex is",
    probability_c)

## the MLE of the probability of a tie between two students of the same sex is 0.3343223
probability_oppo <- exp(-1.9841) / (1 + exp(-1.9841))
cat("the MLE of the probability of a tie between two students of the opposite sex is",
    probability_oppo)

## the MLE of the probability of a tie between two students of the opposite sex is 0.1208825
Null_Deviance <- 973.2
Residual_Deviance <- 699.1
p_value_c <- 1 - pchisq(Null_Deviance - Residual_Deviance, df = 1)
p_value_c

## [1] 0

```

The p-value is extremely small (p-value = 0), indicating that the model with homophily based on sex fits significantly better than the Erdos-Renyi model.

d)

```
fit.diff.homo <- ergm(hansell~edges + nodematch("sex", diff=TRUE))

## Starting maximum pseudolikelihood estimation (MPLE):
## Obtaining the responsible dyads.
## Evaluating the predictor and response matrix.
## Maximizing the pseudolikelihood.
## Finished MPLE.
## Evaluating log-likelihood at the estimate.
summary(fit.diff.homo)

## Call:
## ergm(formula = hansell ~ edges + nodematch("sex", diff = TRUE))
##
## Maximum Likelihood Results:
##
##               Estimate Std. Error MCMC % z value Pr(>|z|)
## edges          -1.9841    0.1608     0 -12.341  <1e-04 ***
## nodematch.sex.female  1.4674    0.2221     0  6.607  <1e-04 ***
## nodematch.sex.male    1.0813    0.2389     0  4.526  <1e-04 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      Null Deviance: 973.2  on 702  degrees of freedom
## Residual Deviance: 696.4  on 699  degrees of freedom
##
## AIC: 702.4  BIC: 716  (Smaller is better. MC Std. Err. = 0)
beta_female <- 1.4674
beta_male <- 1.0813
se_female <- 0.2221
se_male <- 0.2389
```


e)

```
fit.mix <- ergm(hansell~nodemix("sex"))
```

```
## Starting maximum pseudolikelihood estimation (MPLE):
```

```
## Obtaining the responsible dyads.
```

```
## Evaluating the predictor and response matrix.
```

```
## Maximizing the pseudolikelihood.
```

```
## Finished MPLE.
```

```
## Evaluating log-likelihood at the estimate.
```

```
summary(fit.mix)
```

```
## Call:
```

```
## ergm(formula = hansell ~ nodemix("sex"))
```

```
##
```

```
## Maximum Likelihood Results:
```

```
##
```

	Estimate	Std. Error	MCMC %	z value	Pr(> z)
## mix.sex.male.female	-1.8374	0.2153	0	-8.533	<1e-04 ***
## mix.sex.female.male	-2.1493	0.2424	0	-8.867	<1e-04 ***
## mix.sex.male.male	-0.9029	0.1767	0	-5.109	<1e-04 ***

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
##      Null Deviance: 973.2  on 702  degrees of freedom
```

```
## Residual Deviance: 707.2  on 699  degrees of freedom
```

```
##
```

```
## AIC: 713.2  BIC: 726.9  (Smaller is better. MC Std. Err. = 0)
```

```

log_odds_e <- -1.8374
cat("the log-odds of a tie from a male to a female student is", log_odds_e)

## the log-odds of a tie from a male to a female student is -1.8374
probability_e <- exp(-2.1493) / (1 + exp(-2.1493))
cat("the MLE of the probability of a tie from a female to a male student is",
    probability_e)

## the MLE of the probability of a tie from a female to a male student is 0.1043967
sex_attr <- get.vertex.attribute(hansell, "sex")
edge_list <- as.matrix.network.edgelist(hansell)

female_to_male_count <- 0
for (i in 1:nrow(edge_list)) {
  from <- edge_list[i, 1]
  to <- edge_list[i, 2]

  if (sex_attr[from] == "female" && sex_attr[to] == "male") {
    female_to_male_count <- female_to_male_count + 1
  }
}

female_to_male_count / (13 * 14)

## [1] 0.1043956

```

they are almost the same, the difference between them is $-1.0491962 \times 10^{-6}$ (empirical frequency - MLE of the probability of a tie from a female to a male student)

```

Null_Deviance <- 973.2
Residual_Deviance <- 707.2
p_value_e <- 1 - pchisq(Null_Deviance - Residual_Deviance, df = 2)
p_value_e

## [1] 0

```

the p-value is very small, indicating that the full mixing model fits significantly better than the Erdos-Renyi model.