INFX 576: Problem Set 2 - Graph Level Indices and CUG Tests*

Jay Chauhan

Due: Thursday, January 26, 2017

Collaborators: Avanti Chande, Gossudin Siddiqi

Instructions:

Before beginning this assignment, please ensure you have access to R and RStudio.

1. Download the `problemset2.Rmd` file from Canvas. You will also need the `problemset2_data.Rdata` file which contains the three different network datasets needed for this assignment.
2. Replace the "Insert Your Name Here" text in the `author:` field with your own full name. Any collaborators must be listed on the top of your assignment.
3. Be sure to include well-documented (e.g. commented) code chucks, figures and clearly written text chunk explanations as necessary. Any figures should be clearly labeled and appropriately referenced within the text.
4. Collaboration on problem sets is acceptable, and even encouraged, but each student must turn in an individual write-up in his or her own words and his or her own work. The names of all collaborators must be listed on each assignment. Do not copy-and-paste from other students' responses or code.
5. When you have completed the assignment and have **checked** that your code both runs in the Console and knits correctly when you click `Knit PDF`, rename the R Markdown file to `YourLastName_YourFirstName_ps2.Rmd`, knit a PDF and submit the PDF file on Canvas.

Setup:

In this problem set you will need, at minimum, the following R packages.

```r
# Load standard libraries
library(statnet)

# Load data
load("problemset2_data.Rdata")
ls() # Print objects in workspace to see what is available
```

```
## [1] "kaptail.ins" "mids_1993"   "sampson"
```

Problem 1: Graph-Level Indices

Consider the Sampson monk data[1]. Sampson collected various relationships between several monks at a monastery. Suppose we divide the types of social ties into positive and negative relationship types as follows:

- Positive Relationships: Esteem, Influence, LikeT1, LikeT2, LikeT3, and Praise
- Negative Relationships: Disesteem, NegInfluence, Dislike, and Blame

Using a vector permutation test, evaluate the questions below.

*Problems originally written by C.T. Butts (2009)

[1] F. S. Sampson. A novitiate in a period of change: An experimental and case study of social relationships. PhD thesis, Cornell University. 1968.

(a) Are positive relations more reciprocal (relative to density) than negative ones?

```
perm.ver.test<-function(x,niter=5000){  #Define a simple test function
  is_recip<-c(1,0,1,0,1,1,1,0,1,0)
  c.obs<-sum(x[is_recip==1])-sum(x[is_recip==0])
  c.rep<-vector()
  c.two_sided<-vector()
  for(i in 1:niter){
    is_recip_new<-sample(is_recip)
    c.rep[i]<-sum(x[is_recip_new==1])-sum(x[is_recip_new==0])
    }
  cat("Vector Permutation Test:\n\tObserved Difference in reciprocity: ",
      c.obs,"\tReplicate quantiles (niter=",niter,")\n",sep="")
  cat("\t\tPr(rho>=obs):",mean(c.rep>=c.obs),"\n")
  cat("\t\tPr(rho<=obs):",mean(c.rep<=c.obs),"\n")
  cat("\t\tPr(|rho|>=|obs|):",mean(abs(c.rep)>=abs(c.obs)),"\n")
  invisible(list(obs=c.obs, rep=c.rep, two_sided_prob=mean(abs(c.rep)>=abs(c.obs))))
}
```

```
#permutation vector test of the reciprocity(relative to density)
x<-grecip(sampson,measure="edgewise.lrr")
perm.ver.test(x)
```

```
## Vector Permutation Test:
##  Observed Difference in reciprocity: 3.059769    Replicate quantiles (niter=5000)
##      Pr(rho>=obs): 0.0322
##      Pr(rho<=obs): 0.9732
##      Pr(|rho|>=|obs|): 0.0322
```

- Null hypothesis: positive relations are not more reciprocal than negative relations • By the above findings for confidence of 95% we find that the p value is less than the critical value, and we reject the null hypothesis. Thus we conclude that the reciprocity(relative to density) of positivie realtions are higher than the negative relations.

(b) Are positive relations more transitive (relative to density) than negative ones?

```
#get transitivity(realtive to density)
x<-log(gtrans(sampson)/gden(sampson))
perm.ver.test(x)
```

```
## Vector Permutation Test:
##  Observed Difference in reciprocity: 3.421488    Replicate quantiles (niter=5000)
##      Pr(rho>=obs): 0.029
##      Pr(rho<=obs): 0.9766
##      Pr(|rho|>=|obs|): 0.029
```

- Null hypothesis: positive relations are not more transitive than negative relations • By the above findings for confidence of 95% we find that the p value is less than the critical value, and we reject the null hypothesis. Thus we conclude that the transitivity(relative to density) of positivie realtions are higher than the negative relations. ##### (c) Discuss the findings from part (a) and part (b). • Evidence suggests that the in positive network, the relationship between the monks tend to be reciprocated more than in the negative network. This suggests that reciporcity is higher inm monks having positive relations, while reciprocity is lower in monks having negative relations, and these relations tend to be one-sided. Also, when monks have positive relations with each other, there is a high chances that those monks will have many mutual relations with other monks. That is if monk1 has a positive relation with monk2 and monk2 has a positive relation with monk2, then there is a significant chance that monk1 and monk3 will also have a positive relation. And if two monks have a negative relation with each other, there is less chance that a third monk will also have a negative relation.

Problem 2: Random Graphs

(a) Generating Random Graphs

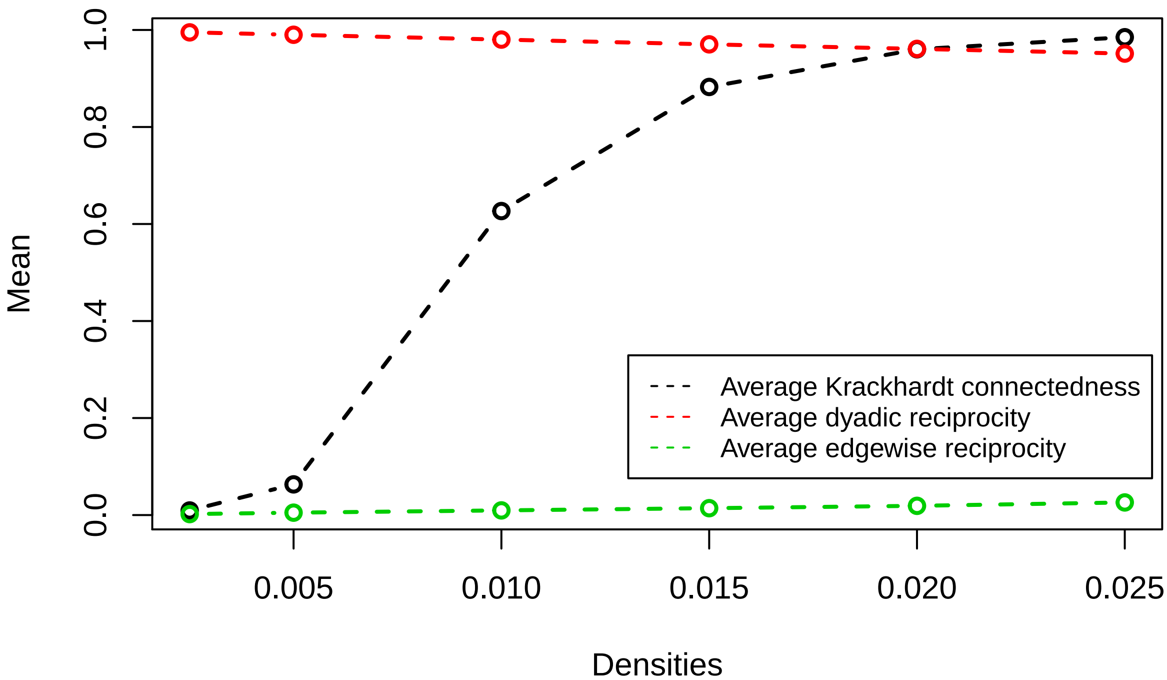
Generate 100-node random directed graphs with expected densities of 0.0025, 0.005, 0.01, 0.015, 0.02, and 0.025, with at least 500 graphs per sample. Remember the `rgraph` function can draw more than one graph at a time. Plot the average Krackhardt connectedness, dyadic reciprocity, and edgewise reciprocity as a function of expected density. Use these to describe the baseline effect of increasing density on network structure.

```
#create random graphs of given densities
den<-c(0.0025,0.005,0.01,0.015,0.02,0.025)
x1<-rgraph(n = 100, m = 500, tprob = den[1])
x2<-rgraph(n = 100, m = 500, tprob = den[2])
x3<-rgraph(n = 100, m = 500, tprob = den[3])
x4<-rgraph(n = 100, m = 500, tprob = den[4])
x5<-rgraph(n = 100, m = 500, tprob = den[5])
x6<-rgraph(n = 100, m = 500, tprob = den[6])
```

```
conn_mean<-c()
dyad_mean<-c()
edge_dyad_mean<-c()

#calcluate connectedness of all the random graphs
conn_mean[1]<-mean(connectedness(x1))
conn_mean[2]<-mean(connectedness(x2))
conn_mean[3]<-mean(connectedness(x3))
conn_mean[4]<-mean(connectedness(x4))
conn_mean[5]<-mean(connectedness(x5))
conn_mean[6]<-mean(connectedness(x6))

#calcluate edgewise reciprocity of all the random graphs
dyad_mean[1]<-mean(grecip(x1))
dyad_mean[2]<-mean(grecip(x2))
dyad_mean[3]<-mean(grecip(x3))
dyad_mean[4]<-mean(grecip(x4))
dyad_mean[5]<-mean(grecip(x5))
dyad_mean[6]<-mean(grecip(x6))

#calcluate dyadic reciprocity of all the random graphs
edge_dyad_mean[1]<-mean(grecip(x1,measure = "edgewise"))
edge_dyad_mean[2]<-mean(grecip(x2,measure = "edgewise"))
edge_dyad_mean[3]<-mean(grecip(x3,measure = "edgewise"))
edge_dyad_mean[4]<-mean(grecip(x4,measure = "edgewise"))
edge_dyad_mean[5]<-mean(grecip(x5,measure = "edgewise"))
edge_dyad_mean[6]<-mean(grecip(x6,measure = "edgewise"))
```

```
#plotting the means of connectedness, edgewise reciprocity and dyadic reciprocity against increaseing d
plot(den,conn_mean,type = "b", col=1, lty=2, lwd=2, xlab="Densities", ylab="Mean")
points(den,dyad_mean, type="b",col=2, lty=2, lwd=2)
points(den,edge_dyad_mean, type="b",col=3, lty=2, lwd=2)
```

```
legend(x="bottomright", inset = c(0.01,0.1), legend=c("Average Krackhardt connectedness","Average dyadi
```



- As we can see, with the increase in density of the graph, the number of weak connections across the graph increases, that is Krackhardt connectedness increases with the increase in connectivity.
- Increase in the density mean that the number of edges across the graph increases, which leads to increase in the number of asymmetric and mutual dyads and a corresponding decrease in the number of null dyads. Since the dyadic reciprocity considers null dyads to be reciprocal as well, it leads to a slight decrease in the mean dyadic reciprocity.
- With the increase in the density the number of assymetroc and mutual dyads increases. Now, since the edgewise reciprocity doesn't consider null dyads to be reciprocal and only considers mututal dyads to be reciprocal, there is only a slight increase in mean edgewise reciprocity. Although we can infer that, with the increase or decrease in the probability of any given edge to be mutual, the edgwise reciprocity can also increase or decrease respecitvely.

(b) Comparing GLIs

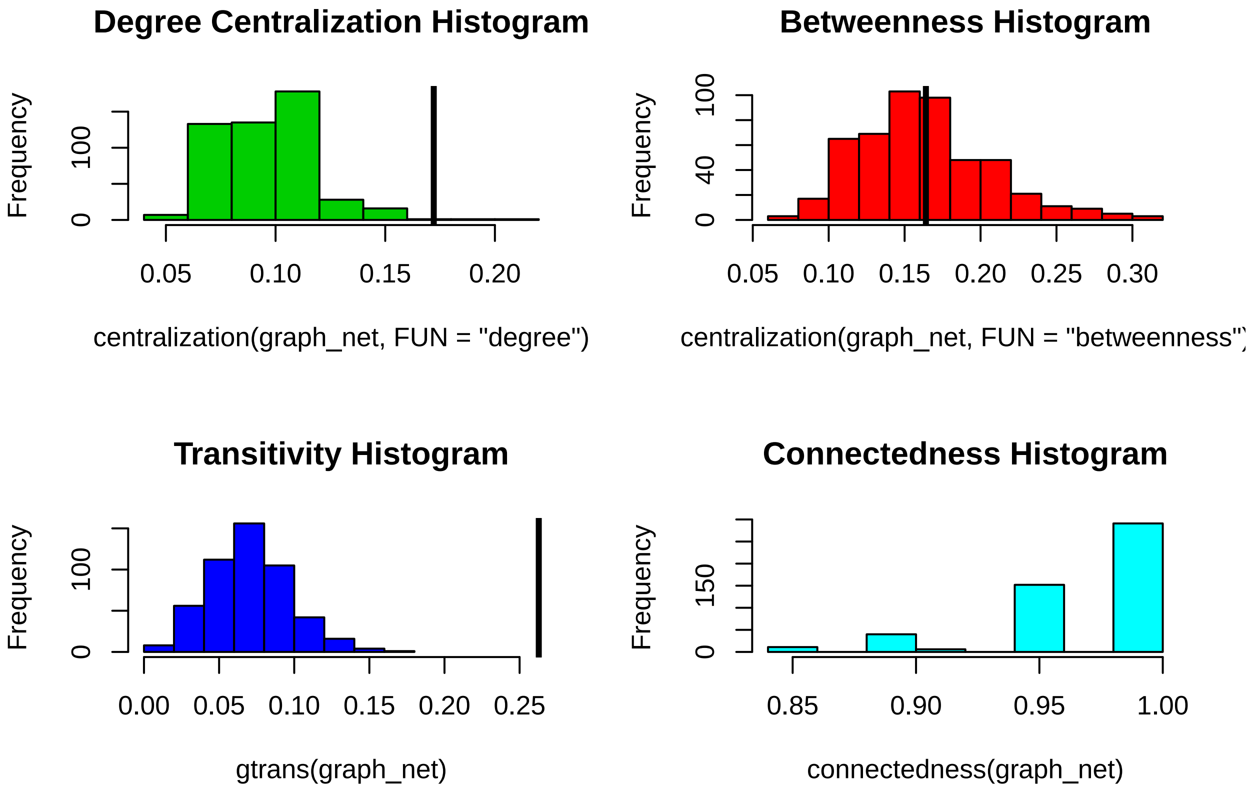
In this problem we will use the well-known social network dataset, collected by Bruce Kapferer in Zambia from June 1965 to August 1965, involves interactions among workers in a tailor shop as observed by Kapferer himself.[2] Here, an interaction is defined by Kapferer as "continuous uninterrupted social activity involving the participation of at least two persons"; only transactions that were relatively frequent are recorded.

Generate 500 random directed graphs whose dyad census is the same as that of `kaptail.ins`. Plot histograms for total degree centralization, betweenness centralization, transitivity, and Krackhardt connectedness from this random sample. On your plot mark the observed values of these statistics (from the `kaptail.ins` data) using a vertial line You mgith find the `abline` function helpful here. Try modifying the `lwd` argument to the plot function to make the vertical line stand out. How do the replicated graphs compare to the observed data.

```
#retrive the dyad census of kaptail network
kaptail_census<-dyad.census(kaptail.ins)
```

[2] Kapferer B. (1972). Strategy and transaction in an African factory. Manchester: Manchester University Press.

```
#create random graphs using kaptail dyad census
graph_net<-rguman(n=500,nv=network.size(kaptail.ins),method="exact",mut=kaptail_census[1],asym=kaptail_

#caluculate degree centrailization, betweenness, transitivity and connectedness of kaptail network
kap_deg_cent<-centralization(kaptail.ins, FUN="degree")
kap_bet_cent<-centralization(kaptail.ins,FUN = "betweenness")
kap_trans<-gtrans(kaptail.ins)
kap_conn<-connectedness(kaptail.ins)

par(mfrow=c(2,2))

#plot the histogram of entrailization, betweenness, transitivity and connectedness of the random graphs
hist(centralization(graph_net, FUN = "degree"), col=3, main = "Degree Centralization Histogram")
abline(v=kap_deg_cent,lwd=3)

hist(centralization(graph_net,FUN = "betweenness"),col=2, main="Betweenness Histogram")
abline(v=kap_bet_cent,lwd=3)

hist(gtrans(graph_net),col=4, xlim=c(0,kap_trans), main="Transitivity Histogram")
abline(v=kap_trans,lwd=3)

hist(connectedness(graph_net),col=5, main="Connectedness Histogram")
abline(v=kap_conn,lwd=3)
```



- As we can see from the above graphs, degree of centralization of the kaptail network is more than that of random graph having the same dyad census, which means that the network is much more concentrated on any one vertex in the kaptail network as compared to random graph having the same dyad census.
- The betweenness of the kaptail network is almost the same as that of the generated random graph having

the same dyad census

- The transitivity of the kaptail network is much higher than that of a random graph having the same dyad census, and we could say that the interaction between the tailors is highly transitive.

- The Krackhaardt connectedness of the Kaptail network is much less than that of a random graph with the same dyad census. Here the Kaptail network is more centralized than the random graph, and less hierarchial, which means that it is less weakly connected than the rondom graph.

Problem 3: Testing Structural Hypotheses

Consider the following set of propositions, which may or may not be true of given dataset. For each, do the following:

1. Identify a statistic (e.g. GLI) whose value should deviate from a random baseline if the proposition is true.
2. Identify the approporate baseline distribution to which the statistic should be compared.
3. Determine whether the proposition implies that the statistic should be greater or lower than its baseline distribution would indicate.
4. Conduct a conditional uniform graph test based on your conclusions in 1-3. In reporting your results, include appropriate summary output from the `cug.test` function as well as the resulting distributional plots. Based on the results, indicate whether the data appears to support or undermine the proposition in question. Be sure to justify your conclusion.

(a) In militarized interstate disputes, hostile acts are disproportionately likely to be responded to in kind.

- Statistic: Edgewise Reciprocity • Baseline Distribution: Size, edges • Direction of Deviation: Higher

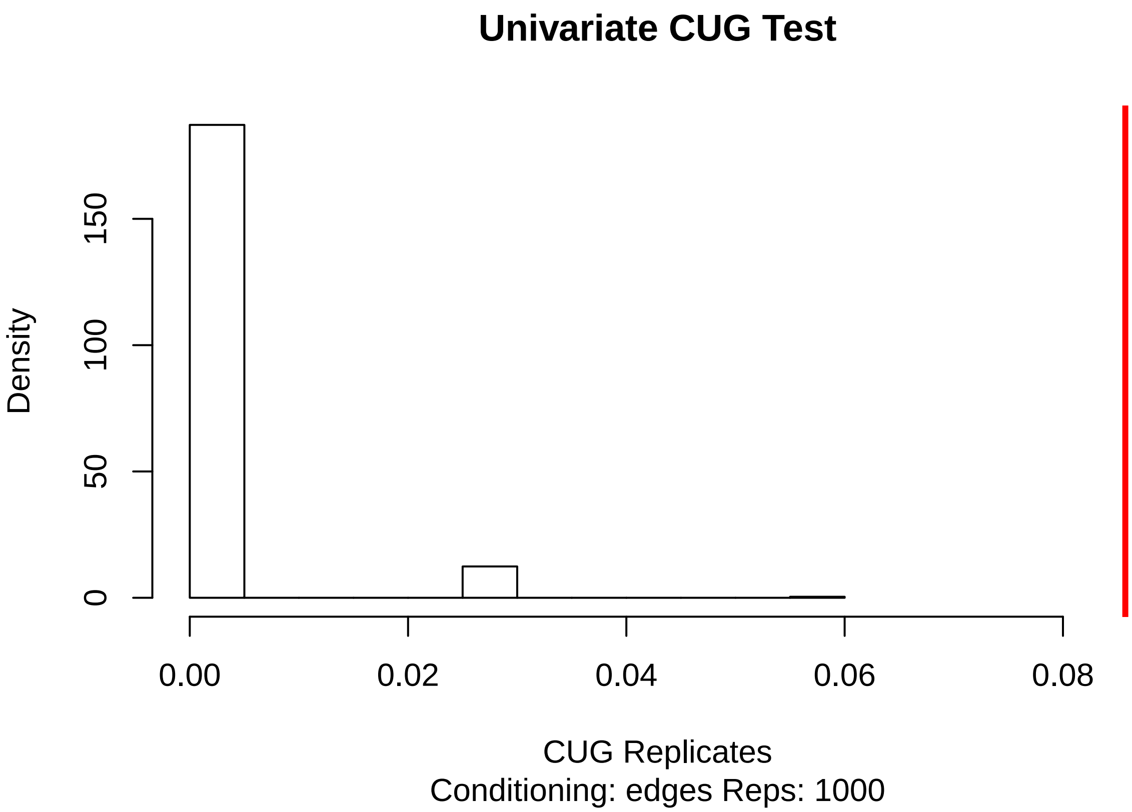
```r
#perform cugtetst of edgewise reciprocity using edge as baseline distribution
cug_1_edge<-cug.test(mids_1993, grecip, cmode = "edges", FUN.args = c(measure="edgewise"))
cug_1_edge
```

```
##
## Univariate Conditional Uniform Graph Test
##
## Conditioning Method: edges
## Graph Type: digraph
## Diagonal Used: FALSE
## Replications: 1000
##
## Observed Value: 0.08571429
## Pr(X>=Obs): 0
## Pr(X<=Obs): 1
```

```r
summary(cug_1_edge)
```

```
##           Length Class  Mode
## obs.stat    1    -none- numeric
## rep.stat 1000    -none- numeric
## mode        1    -none- character
## diag        1    -none- logical
## cmode       1    -none- character
## plteobs     1    -none- numeric
## pgteobs     1    -none- numeric
## reps        1    -none- numeric
```

```r
plot(cug_1_edge)
```



- Since the caluclated higer tail p-value is less than the critical 0.05, and there is a significant deviation from the basleine model in the plots, we reject the null hypothesis and conclude that in militarized interstate disputes, hostile acts are disproportionately likely to be responded to in kind.

```r
#perform cugtetst of edgewise reciprocity using size as baseline distribution
cug_1_size<-cug.test(mids_1993, grecip, cmode = "size", FUN.args = c(measure="edgewise"))
cug_1_size
```

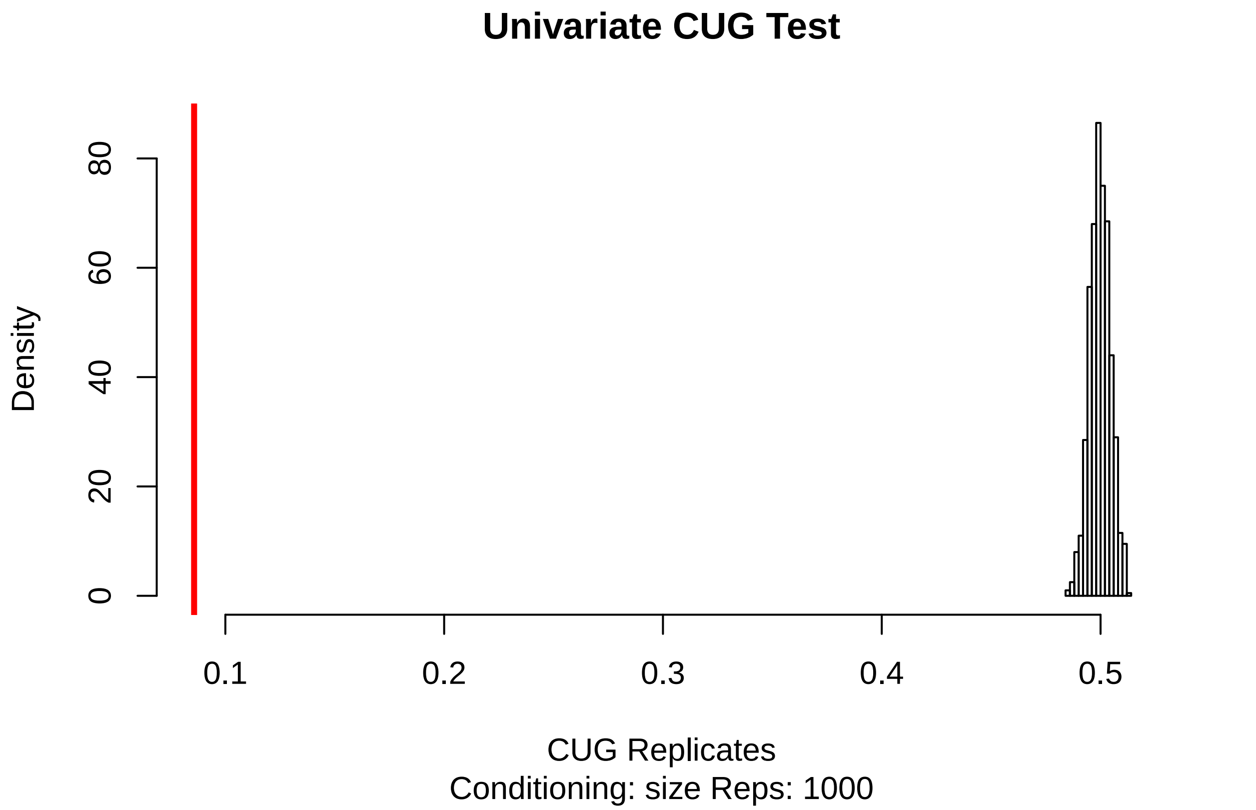
```
##
## Univariate Conditional Uniform Graph Test
##
## Conditioning Method: size
## Graph Type: digraph
## Diagonal Used: FALSE
## Replications: 1000
##
## Observed Value: 0.08571429
## Pr(X>=Obs): 1
## Pr(X<=Obs): 0
```

```r
summary(cug_1_size)
```

```
##          Length Class  Mode
## obs.stat    1   -none- numeric
## rep.stat 1000   -none- numeric
## mode        1   -none- character
## diag        1   -none- logical
```

```
## cmode        1    -none- character
## plteobs      1    -none- numeric
## pgteobs      1    -none- numeric
## reps         1    -none- numeric
```

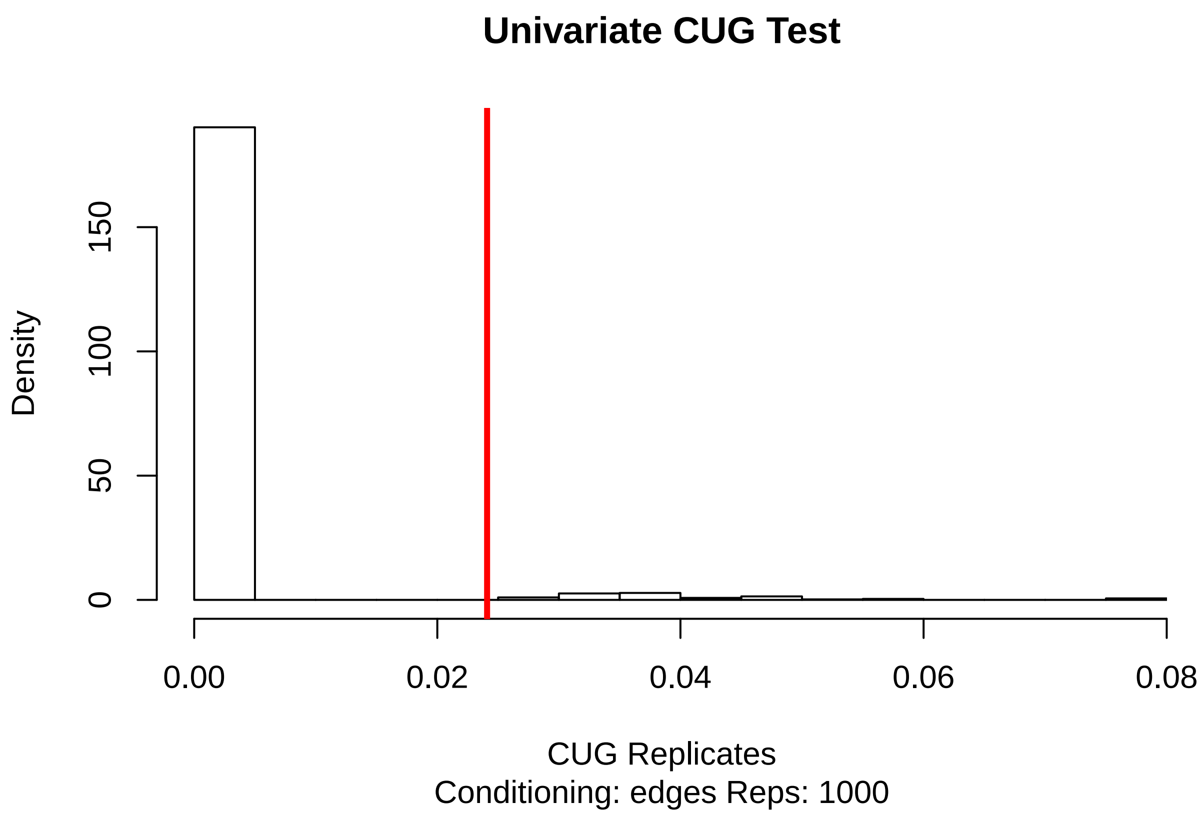
```
plot(cug_1_size)
```



(b) When engaging in disputes, nations behave in accordance with the notion that "the enemy of my enemy is not my enemy".

- Statistic: Transitivity • Baseline Distribution: edges • Direction of Deviation: Lower

```
#perform cugtetst of transitivty using edge as baseline distribution
cug_2_edge<-cug.test(mids_1993, gtrans, cmode="edges")
plot(cug_2_edge)
```




```
cug_2_edge
```

```
##
## Univariate Conditional Uniform Graph Test
##
## Conditioning Method: edges
## Graph Type: digraph
## Diagonal Used: FALSE
## Replications: 1000
##
## Observed Value: 0.02409639
## Pr(X>=Obs): 0.049
## Pr(X<=Obs): 0.951
```

```
summary(cug_2_edge)
```

```
##          Length Class  Mode
## obs.stat    1   -none- numeric
## rep.stat 1000   -none- numeric
## mode        1   -none- character
## diag        1   -none- logical
## cmode       1   -none- character
## plteobs     1   -none- numeric
## pgteobs     1   -none- numeric
## reps        1   -none- numeric
```

- Since the calculated p-value of the lower tail is greater than the critical value of 0.05, and the plots show that there is not a noteworthy departure from the baseline mode, we don't have sufficient evidence to reject the null hypothese, we conclude that when engaging in disputes, nations don't behave in accordance with the notion that "the enemy of my enemy is not my enemy"

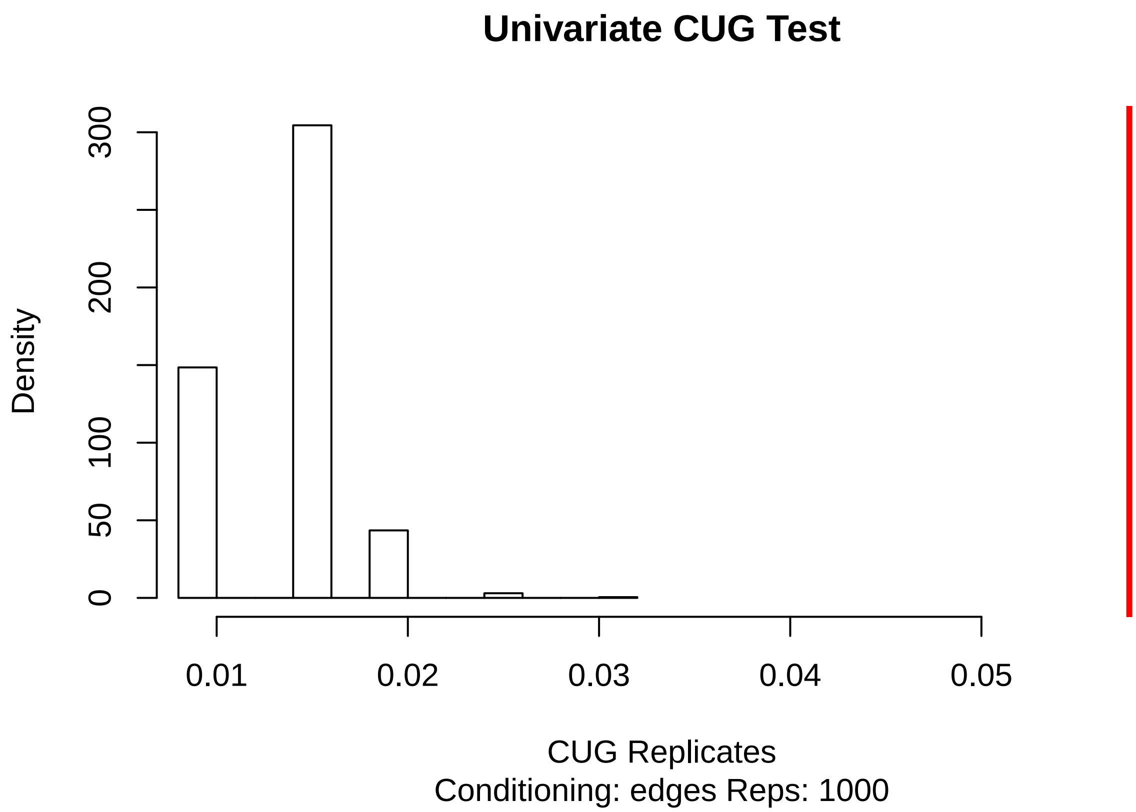
(c) Given the number of disputes at any given time, as small number of nations will receive a disproportionate share of aggressive acts.

- Statistic: Indegree
- Baseline Distribution: Size, edges
- Direction of Deviation: Higher

```
#perform cugtetst of indegree centralization using edge as baseline distribution
cug_3_edge<-cug.test(mids_1993,centralization,cmode="edges",FUN.args=c(FUN="degree",cmode="indegree"))
```

```
## Warning in FUN.args$mode <- mode: Coercing LHS to a list
```

```
plot(cug_3_edge)
```




```
cug_3_edge
```

```
##
## Univariate Conditional Uniform Graph Test
##
## Conditioning Method: edges
## Graph Type: digraph
## Diagonal Used: FALSE
## Replications: 1000
##
## Observed Value: 0.05773557
## Pr(X>=Obs): 0
## Pr(X<=Obs): 1
```

```
summary(cug_3_edge)
```

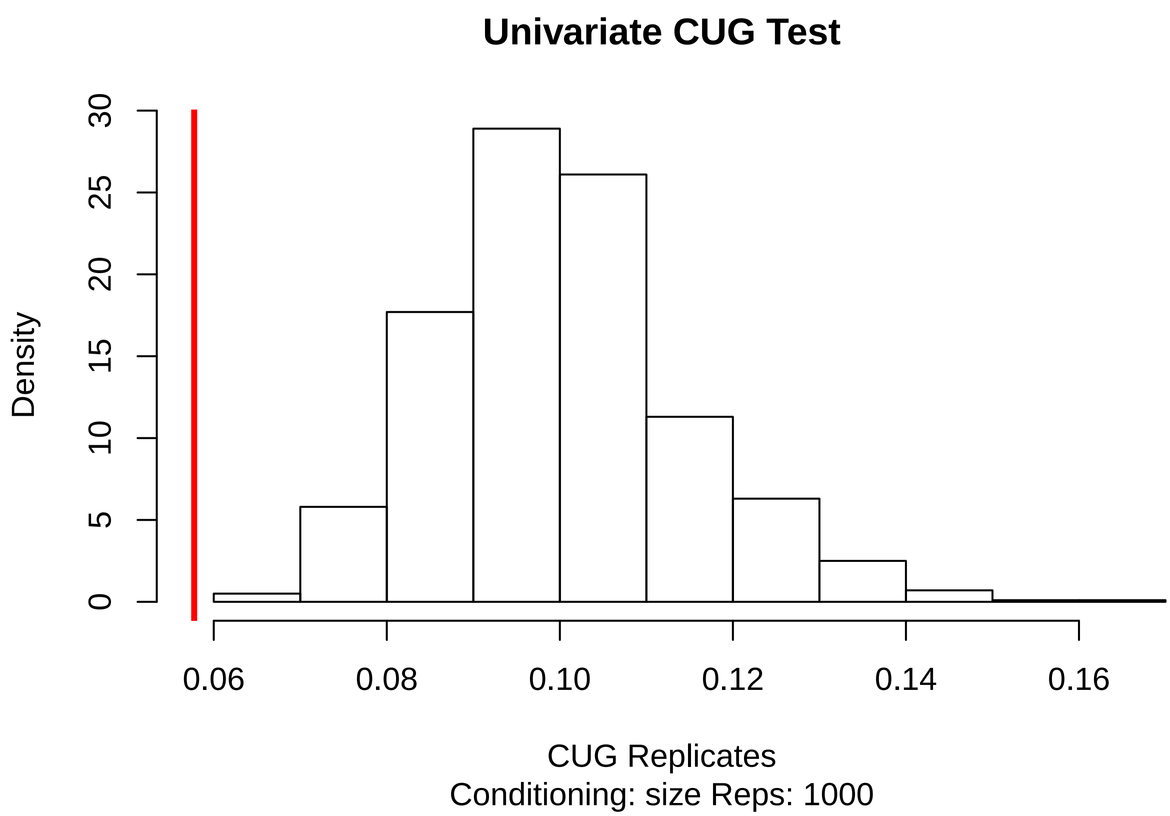
```
##          Length Class  Mode
## obs.stat    1   -none- numeric
## rep.stat 1000   -none- numeric
```

```
## mode        1    -none- character
## diag        1    -none- logical
## cmode       1    -none- character
## plteobs     1    -none- numeric
## pgteobs     1    -none- numeric
## reps        1    -none- numeric
```

```
#perform cugtetst of indegree centralization using size as baseline distribution
cug_3_size<-cug.test(mids_1993,centralization,cmode="size",FUN.args=c(FUN="degree",cmode="indegree"))
```

```
## Warning in FUN.args$mode <- mode: Coercing LHS to a list
```

```
plot(cug_3_size)
```




```
cug_3_size
```

```
##
## Univariate Conditional Uniform Graph Test
##
## Conditioning Method: size
## Graph Type: digraph
## Diagonal Used: FALSE
## Replications: 1000
##
## Observed Value: 0.05773557
## Pr(X>=Obs): 1
## Pr(X<=Obs): 0
```

```
summary(cug_3_size)
```

```
##           Length Class  Mode
```

```
## obs.stat    1   -none- numeric
## rep.stat 1000   -none- numeric
## mode        1   -none- character
## diag        1   -none- logical
## cmode       1   -none- character
## plteobs     1   -none- numeric
## pgteobs     1   -none- numeric
## reps        1   -none- numeric
```

- Since the caluclated higer tail p-value is less than the critical 0.05, and there is a significant deviation from the basleine model in the plots, we reject the null hypothesis and conclude that given the number of disputes at any given time, as small number of nations will receive a disproportionate share of aggressive acts