Final_Project_Network_Regression

```r
library(rio)
library(igraph)
library(dplyr)
library(readr)

#Loading all the CA House and Senate data

#Each csv file in Edge_Lists represents 1 network. The first two columns are the sender/receiver identified with their "EID"
#        (the identifier provided by NIMP). The third column is used to identify what threshold the edge was

CA_2011_2012_Senate_edges <- read_csv("Edgelist/CA-2011-2012-Senate.csv")
CA_2011_2012_Senate <- read_csv("Metadata/CA-2011-2012-Senate.csv")

CA_2013_2014_Senate_edges <- read_csv("Edgelist/CA-2013-2014-Senate.csv")
CA_2013_2014_Senate <- read_csv("Metadata/CA-2013-2014-Senate.csv")

CA_2015_2016_Senate_edges <- read_csv("Edgelist/CA-2015-2016-Senate.csv")
CA_2015_2016_Senate <- read_csv("Metadata/CA-2015-2016-Senate.csv")

CA_2011_2012_House <- read_csv("Metadata/CA-2011-2012-House.csv")
CA_2011_2012_House_edges <- read_csv("Edgelist/CA-2011-2012-House.csv")

CA_2013_2014_House <- read_csv("Metadata/CA-2013-2014-House.csv")
CA_2013_2014_House_edges <- read_csv("Edgelist/CA-2013-2014-House.csv")

CA_2015_2016_House <- read_csv("Metadata/CA-2015-2016-House.csv")
CA_2015_2016_House_edges <- read_csv("Edgelist/CA-2015-2016-House.csv")

regression_data <- function(nodes, edges) {
  nodes_2 <- nodes %>%
    filter(CatCodeGroup == "Health" | CatCodeGroup == "Labor" | CatCodeGroup == "Agriculture" | CatCodeGroup == "General
        Business" | CatCodeGroup == "Finance, Insurance & Real Estate") %>%
    mutate(party_node_2 = EID) %>%
    mutate(party_group_2 = CatCodeGroup) %>%
    mutate(Total_2 = Total) %>%
    mutate(PerDem_2 = PerDem) %>%
    mutate(PerRep_2 = PerRep) %>%
    select(party_node_2, party_group_2, Total_2, PerDem_2, PerRep_2)

  nodes_1 <- nodes %>%
    filter(CatCodeGroup == "Health" | CatCodeGroup == "Labor" | CatCodeGroup == "Agriculture" | CatCodeGroup == "General
        Business" | CatCodeGroup == "Finance, Insurance & Real Estate") %>%
    mutate(party_node_1 = EID) %>%
    mutate(party_group_1 = CatCodeGroup) %>%
    mutate(Total_1 = Total) %>%
    mutate(PerDem_1 = PerDem) %>%
    mutate(PerRep_1 = PerRep) %>%
    select(EID, party_node_1, party_group_1, Total_1, PerDem_1, PerRep_1) %>%
    inner_join(edges, by = c("EID" = "node_1")) %>%
    inner_join(nodes_2, by = c("node_2" = "party_node_2")) %>%
    select(-EID, -edge) %>%
    mutate(party_node_2 = node_2) %>%
    mutate(party_1_group = party_group_1) %>%
    mutate(party_2_group = party_group_2) %>%
    mutate(party_1_total = Total_1) %>%
    mutate(party_2_total = Total_2) %>%
    mutate(dem_diff = abs(PerDem_1 - PerDem_2)) %>%
    mutate(rep_diff = abs(PerRep_1 - PerRep_2)) %>%
    select(-party_group_1, -party_group_2, -node_2, -Total_1, -Total_2, -PerDem_1, -PerRep_1, -PerDem_2, -PerRep_2) %>%
    #https://www.statology.org/r-convert-true-false-to-1-0/
    mutate(same = as.integer(as.logical(party_1_group == party_2_group))) %>%
    mutate(total_diff = abs(party_1_total - party_2_total))
  return(nodes_1)
}

#First step for ERGM with all House and Senate years prepared

CA_11_12_House <- graph.data.frame(regression_data(CA_2011_2012_House, CA_2011_2012_House_edges))
CA_11_12_House_sim <- get.adjacency(CA_11_12_House, sparse = FALSE, attr = 'same')
CA_11_12_House_total_diff <- get.adjacency(CA_11_12_House, sparse = FALSE, attr = 'total_diff')
CA_11_12_House_per_dem <- get.adjacency(CA_11_12_House, sparse = FALSE, attr = 'dem_diff')
CA_11_12_House_per_rep <- get.adjacency(CA_11_12_House, sparse = FALSE, attr = 'rep_diff')
#our dependent variable is the ties between every donor
CA_11_12_House_ties <- as.matrix(get.adjacency(CA_11_12_House))

CA_13_14_House <- graph.data.frame(regression_data(CA_2013_2014_House, CA_2013_2014_House_edges))
```

```r
CA_13_14_House_sim <- get.adjacency(CA_13_14_House, sparse = FALSE, attr = 'same')
CA_13_14_House_total_diff <- get.adjacency(CA_13_14_House, sparse = FALSE, attr = 'total_diff')
CA_13_14_House_per_dem <- get.adjacency(CA_13_14_House, sparse = FALSE, attr = 'dem_diff')
CA_13_14_House_per_rep <- get.adjacency(CA_13_14_House, sparse = FALSE, attr = 'rep_diff')
CA_13_14_House_ties <- as.matrix(get.adjacency(CA_13_14_House))

CA_15_16_House <- graph.data.frame(regression_data(CA_2015_2016_House, CA_2015_2016_House_edges))
CA_15_16_House_sim <- get.adjacency(CA_15_16_House, sparse = FALSE, attr = 'same')
CA_15_16_House_total_diff <- get.adjacency(CA_15_16_House, sparse = FALSE, attr = 'total_diff')
CA_15_16_House_per_dem <- get.adjacency(CA_15_16_House, sparse = FALSE, attr = 'dem_diff')
CA_15_16_House_per_rep <- get.adjacency(CA_15_16_House, sparse = FALSE, attr = 'rep_diff')
CA_15_16_House_ties <- as.matrix(get.adjacency(CA_15_16_House))

CA_11_12_Senate <- graph.data.frame(regression_data(CA_2011_2012_Senate, CA_2011_2012_Senate_edges))
CA_11_12_Senate_sim <- get.adjacency(CA_11_12_Senate, sparse = FALSE, attr = 'same')
CA_11_12_Senate_total_diff <- get.adjacency(CA_11_12_Senate, sparse = FALSE, attr = 'total_diff')
CA_11_12_Senate_per_dem <- get.adjacency(CA_11_12_Senate, sparse = FALSE, attr = 'dem_diff')
CA_11_12_Senate_per_rep <- get.adjacency(CA_11_12_Senate, sparse = FALSE, attr = 'rep_diff')
CA_11_12_Senate_ties <- as.matrix(get.adjacency(CA_11_12_Senate))

CA_13_14_Senate <- graph.data.frame(regression_data(CA_2013_2014_Senate, CA_2013_2014_Senate_edges))
CA_13_14_Senate_sim <- get.adjacency(CA_13_14_Senate, sparse = FALSE, attr = 'same')
CA_13_14_Senate_total_diff <- get.adjacency(CA_13_14_Senate, sparse = FALSE, attr = 'total_diff')
CA_13_14_Senate_per_dem <- get.adjacency(CA_13_14_Senate, sparse = FALSE, attr = 'dem_diff')
CA_13_14_Senate_per_rep <- get.adjacency(CA_13_14_Senate, sparse = FALSE, attr = 'rep_diff')
CA_13_14_Senate_ties <- as.matrix(get.adjacency(CA_13_14_Senate))

CA_15_16_Senate <- graph.data.frame(regression_data(CA_2015_2016_Senate, CA_2015_2016_Senate_edges))
CA_15_16_Senate_sim <- get.adjacency(CA_15_16_Senate, sparse = FALSE, attr = 'same')
CA_15_16_Senate_total_diff <- get.adjacency(CA_15_16_Senate, sparse = FALSE, attr = 'total_diff')
CA_15_16_Senate_per_dem <- get.adjacency(CA_15_16_Senate, sparse = FALSE, attr = 'dem_diff')
CA_15_16_Senate_per_rep <- get.adjacency(CA_15_16_Senate, sparse = FALSE, attr = 'rep_diff')
CA_15_16_Senate_ties <- as.matrix(get.adjacency(CA_15_16_Senate))


#Detaching igraph to prepare for final step of our regression model
detach(package:igraph)
library(statnet)


#If they are of the same groups that there is a 48% chance higher that a tie will exist between two donors
CA_11_12_House_matrices <- array(NA, c(4, length(CA_11_12_House_sim[1,]), length(CA_11_12_House_sim[1,])))

CA_11_12_House_matrices[1,,] <- CA_11_12_House_sim
CA_11_12_House_matrices[2,,] <- CA_11_12_House_total_diff
CA_11_12_House_matrices[3,,] <- CA_11_12_House_per_dem
CA_11_12_House_matrices[4,,] <- CA_11_12_House_per_rep

CA_11_12_House_lm <- netlm(CA_11_12_House_ties, CA_11_12_House_matrices, reps=100)

CA_11_12_House_model <- list()
CA_11_12_House_model <- summary(CA_11_12_House_lm)
CA_11_12_House_model$names <- c("Intercept", "Same Group", "Total Contributions Difference", "Per Democratic Difference",
        "Per Republican Difference")

CA_11_12_House_model$coefficients = round(CA_11_12_House_model$coefficients, 2)
CA_11_12_House_model
```

```
OLS Network Model

Residuals:
       0%        25%        50%        75%       100%
-2.7673007 -0.0440547 -0.0440547 -0.0440547  0.9559453

Coefficients:
                                Estimate Pr(<=b) Pr(>=b) Pr(>=|b|)
Intercept                       0.04     1       0       0
Same Group                      0.48     1       0       0
Total Contributions Difference  0.00     1       0       0
Per Democratic Difference       0.01     1       0       0
Per Republican Difference       0.01     1       0       0

Residual standard error: 0.1954 on 487897 degrees of freedom
Multiple R-squared: 0.6699  Adjusted R-squared: 0.6699
F-statistic: 2.475e+05 on 4 and 487897 degrees of freedom, p-value:      0


Test Diagnostics:

    Null Hypothesis: qap
    Replications: 100
    Coefficient Distribution Summary:
```

```
        Intercept Same Group Total Contributions Difference
Min     138.30512 -3.13406                        -7.43826
1stQ    141.22124 -0.92960                        -2.62539
Median  141.87535  0.20246                        -1.16406
Mean    141.73239  0.15528                        -0.81417
3rdQ    142.51870  1.08190                         1.00061
Max     143.76674  4.14995                         5.94950
        Per Democratic Difference Per Republican Difference
Min                      -2.08833                  -5.57160
1stQ                     -0.71009                  -0.35748
Median                   -0.16283                   0.32919
Mean                      0.19618                   0.05846
3rdQ                      0.64386                   1.01444
Max                       5.43444                   2.95466
```

```r
CA_13_14_House_matrices <- array(NA, c(4, length(CA_13_14_House_sim[1,]), length(CA_13_14_House_sim[1,])))

CA_13_14_House_matrices[1,,] <- CA_13_14_House_sim
CA_13_14_House_matrices[2,,] <- CA_13_14_House_total_diff
CA_13_14_House_matrices[3,,] <- CA_13_14_House_per_dem
CA_13_14_House_matrices[4,,] <- CA_13_14_House_per_rep

CA_13_14_House_lm <- netlm(CA_13_14_House_ties, CA_13_14_House_matrices, reps=100)

CA_13_14_House_model <- list()
CA_13_14_House_model <- summary(CA_13_14_House_lm)
CA_13_14_House_model$names <- c("Intercept", "Same Group", "Total Contributions Difference", "Per Democratic Difference",
    "Per Republican Difference")

CA_13_14_House_model$coefficients = round(CA_13_14_House_model$coefficients, 2)
CA_13_14_House_model
```

```
OLS Network Model

Residuals:
        0%          25%          50%          75%         100%
-2.95216420 -0.06340898 -0.06340898 -0.06340898  0.93659102

Coefficients:
                               Estimate Pr(<=b) Pr(>=b) Pr(>=|b|)
Intercept                      0.06     1.00    0.00     0
Same Group                     0.50     1.00    0.00     0
Total Contributions Difference 0.00     1.00    0.00     0
Per Democratic Difference      0.02     0.13    0.87     1
Per Republican Difference      NA       NA      NA       NA

Residual standard error: 0.229 on 419252 degrees of freedom
Multiple R-squared: 0.633   Adjusted R-squared: 0.633
F-statistic: 2.411e+05 on 3 and 419252 degrees of freedom, p-value:     0


Test Diagnostics:

    Null Hypothesis: qap
    Replications: 100
    Coefficient Distribution Summary:

        Intercept Same Group Total Contributions Difference
Min     5.329e-12 -5.044e-10                     -7.245e-15
1stQ    5.459e-12 -1.200e-10                     -2.064e-15
Median  5.488e-12 -3.053e-11                     -4.017e-16
Mean    5.490e-12 -1.554e-11                     -5.566e-17
3rdQ    5.524e-12  1.092e-10                      2.064e-15
Max     5.647e-12  5.229e-10                      8.162e-15
        Per Democratic Difference Per Republican Difference
Min                     -4.061e+00                -4.041e+00
1stQ                     2.711e-01                 2.718e-01
Median                   2.762e-01                 2.759e-01
Mean                     4.292e-03                 7.163e-02
3rdQ                     2.796e-01                 2.791e-01
Max                      1.990e+00                 3.483e+00
```

```r
CA_15_16_House_matrices <- array(NA, c(4, length(CA_15_16_House_sim[1,]), length(CA_15_16_House_sim[1,])))

CA_15_16_House_matrices[1,,] <- CA_15_16_House_sim
CA_15_16_House_matrices[2,,] <- CA_15_16_House_total_diff
CA_15_16_House_matrices[3,,] <- CA_15_16_House_per_dem
CA_15_16_House_matrices[4,,] <- CA_15_16_House_per_rep

CA_15_16_House_lm <- netlm(CA_15_16_House_ties, CA_15_16_House_matrices, reps=100)
```

```
CA_15_16_House_model <- list()
CA_15_16_House_model <- summary(CA_15_16_House_lm)
CA_15_16_House_model$names <- c("Intercept", "Same Group", "Total Contributions Difference", "Per Democratic Difference",
        "Per Republican Difference")

CA_15_16_House_model$coefficients = round(CA_15_16_House_model$coefficients, 2)
CA_15_16_House_model
```

OLS Network Model

Residuals:
```
        0%          25%          50%          75%         100%
-2.46411463  -0.06208445  -0.06208445  -0.06208445   0.93791555
```

Coefficients:

	Estimate	Pr(<=b)	Pr(>=b)	Pr(>=|b|)
Intercept	0.06	1.00	0.00	0
Same Group	0.46	1.00	0.00	0
Total Contributions Difference	0.00	1.00	0.00	0
Per Democratic Difference	0.02	0.15	0.85	1
Per Republican Difference	NA	NA	NA	NA

Residual standard error: 0.2249 on 273002 degrees of freedom
Multiple R-squared: 0.6648 Adjusted R-squared: 0.6648
F-statistic: 1.805e+05 on 3 and 273002 degrees of freedom, p-value: 0

Test Diagnostics:

    Null Hypothesis: qap
    Replications: 100
    Coefficient Distribution Summary:

	Intercept	Same Group	Total Contributions Difference
Min	1.436e-11	-1.172e-10	-2.464e-15
1stQ	1.477e-11	-4.041e-11	-8.243e-16
Median	1.491e-11	-7.362e-12	-1.848e-16
Mean	1.489e-11	-5.338e-12	-7.126e-17
3rdQ	1.499e-11	2.791e-11	6.872e-16
Max	1.523e-11	1.032e-10	2.998e-15

	Per Democratic Difference	Per Republican Difference
Min	-4.084e+00	-3.964e+00
1stQ	2.721e-01	2.722e-01
Median	2.752e-01	2.747e-01
Mean	-1.639e-02	-1.028e-01
3rdQ	2.773e-01	2.775e-01
Max	2.135e+00	2.870e-01

```
CA_11_12_Senate_matrices <- array(NA, c(4, length(CA_11_12_Senate_sim[1,]), length(CA_11_12_Senate_sim[1,])))

CA_11_12_Senate_matrices[1,,] <- CA_11_12_Senate_sim
CA_11_12_Senate_matrices[2,,] <- CA_11_12_Senate_total_diff
CA_11_12_Senate_matrices[3,,] <- CA_11_12_Senate_per_dem
CA_11_12_Senate_matrices[4,,] <- CA_11_12_Senate_per_rep

CA_11_12_Senate_lm <- netlm(CA_11_12_Senate_ties, CA_11_12_Senate_matrices, reps=100)

CA_11_12_Senate_model <- list()
CA_11_12_Senate_model <- summary(CA_11_12_Senate_lm)
CA_11_12_Senate_model$names <- c("Intercept", "Same Group", "Total Contributions Difference", "Per Democratic Difference",
        "Per Republican Difference")

CA_11_12_Senate_model$coefficients = round(CA_11_12_Senate_model$coefficients, 2)
CA_11_12_Senate_model
```

OLS Network Model

Residuals:
```
        0%          25%          50%          75%         100%
-1.56256609  -0.07662802  -0.07662802  -0.07662802   0.92337198
```

Coefficients:

	Estimate	Pr(<=b)	Pr(>=b)	Pr(>=|b|)
Intercept	0.08	1.0	0.0	0
Same Group	0.45	1.0	0.0	0
Total Contributions Difference	0.00	1.0	0.0	0
Per Democratic Difference	0.01	0.8	0.2	1
Per Republican Difference	NA	NA	NA	NA

Residual standard error: 0.2405 on 198466 degrees of freedom

```
Multiple R-squared: 0.6878  Adjusted R-squared: 0.6878
F-statistic: 1.458e+05 on 3 and 198466 degrees of freedom, p-value:      0


Test Diagnostics:

    Null Hypothesis: qap
    Replications: 100
    Coefficient Distribution Summary:

       Intercept Same Group Total Contributions Difference
Min    2.531e-11 -2.473e-10                     -1.438e-15
1stQ   2.571e-11 -8.042e-11                     -3.853e-16
Median 2.589e-11 -9.441e-12                     -5.608e-17
Mean   2.594e-11 -9.711e-12                     -5.587e-17
3rdQ   2.615e-11  6.319e-11                      3.140e-16
Max    2.677e-11  2.359e-10                      1.104e-15
       Per Democratic Difference Per Republican Difference
Min                   -4.174e+00                -1.859e+00
1stQ                  -3.207e-01                -3.202e-01
Median                -3.179e-01                -3.165e-01
Mean                  -1.682e-02                 6.839e-03
3rdQ                  -3.143e-01                -3.131e-01
Max                    3.811e+00                 3.705e+00


CA_13_14_Senate_matrices <- array(NA, c(4, length(CA_13_14_Senate_sim[1,]), length(CA_13_14_Senate_sim[1,])))

CA_13_14_Senate_matrices[1,,] <- CA_13_14_Senate_sim
CA_13_14_Senate_matrices[2,,] <- CA_13_14_Senate_total_diff
CA_13_14_Senate_matrices[3,,] <- CA_13_14_Senate_per_dem
CA_13_14_Senate_matrices[4,,] <- CA_13_14_Senate_per_rep

CA_13_14_Senate_lm <- netlm(CA_13_14_Senate_ties, CA_13_14_Senate_matrices, reps=100)

CA_13_14_Senate_model <- list()
CA_13_14_Senate_model <- summary(CA_13_14_Senate_lm)
CA_13_14_Senate_model$names <- c("Intercept", "Same Group", "Total Contributions Difference", "Per Democratic Difference",
        "Per Republican Difference")

CA_13_14_Senate_model$coefficients = round(CA_13_14_Senate_model$coefficients, 2)
CA_13_14_Senate_model


OLS Network Model

Residuals:
         0%         25%         50%         75%        100%
-1.82849596 -0.09960791 -0.09960791 -0.09960791  0.90039209

Coefficients:
                              Estimate Pr(<=b) Pr(>=b) Pr(>=|b|)
Intercept                     0.10     1.00    0.00     0
Same Group                    0.45     1.00    0.00     0
Total Contributions Difference 0.00    1.00    0.00     0
Per Democratic Difference     0.01     0.23    0.77     1
Per Republican Difference     NA       NA      NA      NA

Residual standard error: 0.2657 on 212978 degrees of freedom
Multiple R-squared: 0.6578  Adjusted R-squared: 0.6578
F-statistic: 1.365e+05 on 3 and 212978 degrees of freedom, p-value:      0


Test Diagnostics:

    Null Hypothesis: qap
    Replications: 100
    Coefficient Distribution Summary:

       Intercept Same Group Total Contributions Difference
Min    5.786e-12 -1.948e-10                     -1.645e-15
1stQ   6.007e-12 -7.174e-11                     -5.113e-16
Median 6.093e-12  7.485e-12                      9.604e-17
Mean   6.083e-12  1.070e-11                      6.264e-17
3rdQ   6.152e-12  9.239e-11                      4.928e-16
Max    6.381e-12  2.266e-10                      2.177e-15
       Per Democratic Difference Per Republican Difference
Min                   -2.856e+00                -2.766e+00
1stQ                   3.230e-01                 3.679e-01
Median                 3.722e-01                 3.732e-01
Mean                   4.696e-02                 1.088e-01
3rdQ                   3.785e-01                 3.763e-01
Max                    2.736e+00                 3.199e+00
```

```r
CA_15_16_Senate_matrices <- array(NA, c(4, length(CA_15_16_Senate_sim[1,]), length(CA_15_16_Senate_sim[1,])))

CA_15_16_Senate_matrices[1,,] <- CA_15_16_Senate_sim
CA_15_16_Senate_matrices[2,,] <- CA_15_16_Senate_total_diff
CA_15_16_Senate_matrices[3,,] <- CA_15_16_Senate_per_dem
CA_15_16_Senate_matrices[4,,] <- CA_15_16_Senate_per_rep

CA_15_16_Senate_lm <- netlm(CA_15_16_Senate_ties, CA_15_16_Senate_matrices, reps=100)

CA_15_16_Senate_model <- list()
CA_15_16_Senate_model <- summary(CA_15_16_Senate_lm)
CA_15_16_Senate_model$names <- c("Intercept", "Same Group", "Total Contributions Difference", "Per Democratic Difference",
        "Per Republican Difference")

CA_15_16_Senate_model$coefficients = round(CA_15_16_Senate_model$coefficients, 2)
CA_15_16_Senate_model
```

```
OLS Network Model

Residuals:
        0%          25%          50%          75%         100%
-2.24370822  -0.13638222  -0.13638222  -0.05329781   0.86361778

Coefficients:
                                Estimate Pr(<=b) Pr(>=b) Pr(>=|b|)
Intercept                       0.14     1.00    0.00     0
Same Group                      0.46     1.00    0.00     0
Total Contributions Difference  0.00     1.00    0.00     0
Per Democratic Difference       0.02     0.73    0.27     1
Per Republican Difference       NA       NA      NA       NA

Residual standard error: 0.3091 on 176816 degrees of freedom
Multiple R-squared: 0.5446  Adjusted R-squared: 0.5446
F-statistic: 7.047e+04 on 3 and 176816 degrees of freedom, p-value:     0


Test Diagnostics:

    Null Hypothesis: qap
    Replications: 100
    Coefficient Distribution Summary:

       Intercept Same Group Total Contributions Difference
Min    9.987e-12 -2.162e-10                     -2.615e-15
1stQ   1.024e-11 -6.724e-11                     -6.149e-16
Median 1.037e-11 -9.996e-12                      3.534e-17
Mean   1.037e-11 -1.369e-11                      1.027e-16
3rdQ   1.048e-11  3.983e-11                      8.242e-16
Max    1.076e-11  1.531e-10                      2.751e-15
       Per Democratic Difference Per Republican Difference
Min                   -1.817e+00                -1.996e+00
1stQ                  -4.426e-01                -4.448e-01
Median                -4.368e-01                -4.404e-01
Mean                   7.902e-02                -3.817e-02
3rdQ                   2.791e-01                -4.314e-01
Max                    2.777e+00                 2.677e+00
```

References Consulted

Reuning, Kevin. *Election Donation Networks*. Harvard Dataverse, 4 June 2020, https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi%3A10.7910%2FDVN%2FYMDFPW.

Zach. "How to Convert TRUE and FALSE to 1 and 0 in R." *Statology*, 9 Sept. 2022, https://www.statology.org/r-convert-true-false-to-1-0/.