INFX 576: Problem Set 3 - Selective Mixing*

Avanti Chande

Due: Thursday, February 2, 2017

Collaborators: Jay Chauhan, Gosuddin Siddiqi

Instructions:

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

1. Download the `problemset3.Rmd` file from Canvas.
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_ps3.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.

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

```
## Warning: package 'statnet' was built under R version 3.2.5
## Warning: package 'tergm' was built under R version 3.2.5
## Warning: package 'statnet.common' was built under R version 3.2.5
## Warning: package 'ergm' was built under R version 3.2.5
## Warning: package 'network' was built under R version 3.2.5
## Warning: package 'networkDynamic' was built under R version 3.2.5
## Warning: package 'ergm.count' was built under R version 3.2.5
## Warning: package 'sna' was built under R version 3.2.5
```

```
library(plyr)
```

```
## Warning: package 'plyr' was built under R version 3.2.5
```

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

Problem 1: Selective Mixing

We will begin with an examination of selective mixing in the Drabek et al. EMON data (which can be accessed in the `network` package via the command `data(emon)`). Recall the `emon` object is a list of seven networks, such that `emon[[1]]` is the *i*th network object. (See `?emon` for details regarding the dataset.)

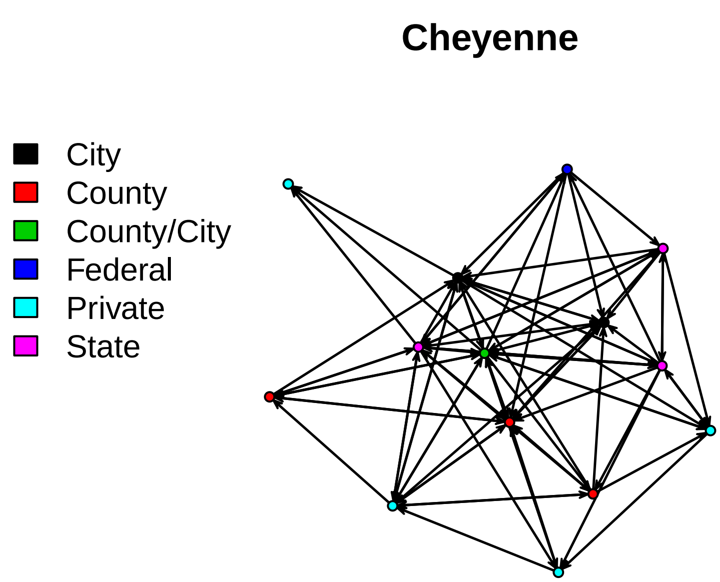
```
data(emon)
```

(a) Visualizing Networks

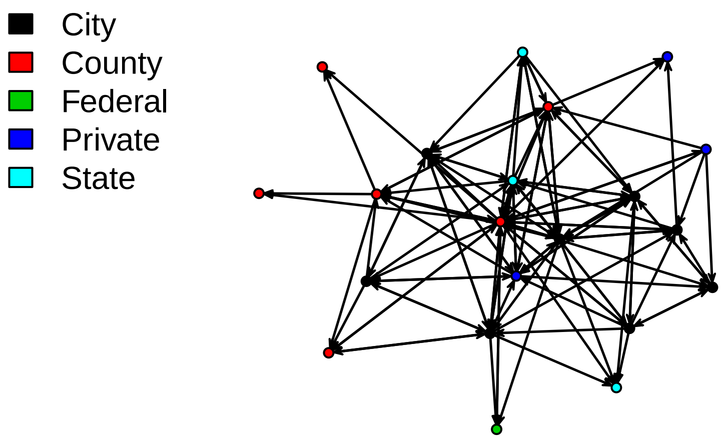
This data set consists of seven individual network data sets of emergent multiorganizational networks (EMONs) in the context of search and rescue activities. These data sets are: the Cheyenne SAR EMON, the Hurricane Frederic SAR EMON, the Lake Pomona SAR EMON, the Mt. Si SAR EMON, the Mt. St. Helens SAR EMON, the Texas Hill Country SAR EMON, and the Wichita Falls SAR EMON. We interpret the relationships in each of these networks as one of salient communication.

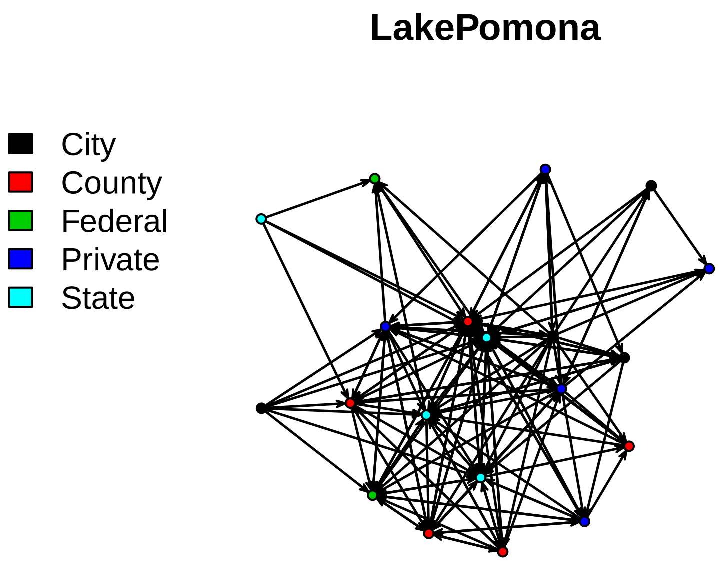
To begin, plot each of the seven networks, coloring vertices by the "Sponsorship" vertex attribute. With each plot, include a legend showing how sponsorship is colored. Comment on what you see.

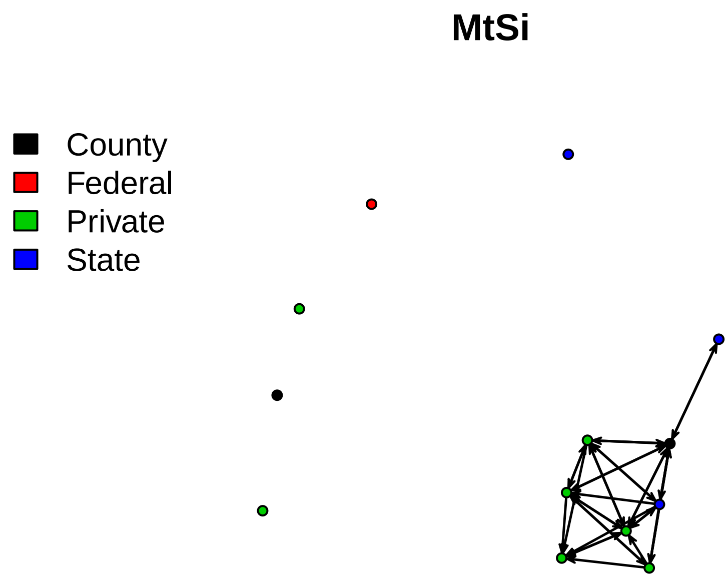
```
names_emon = names(emon)

for (i in 1:7) {
  plot(emon[[i]],vertex.col="Sponsorship")
  title(main = names_emon[i])
  vals<-sort(unique(emon[[i]]%v%"Sponsorship"))    # Let's add a legend....
  legend("topleft",fill=1:length(vals),legend=vals,bty="n")
}
```



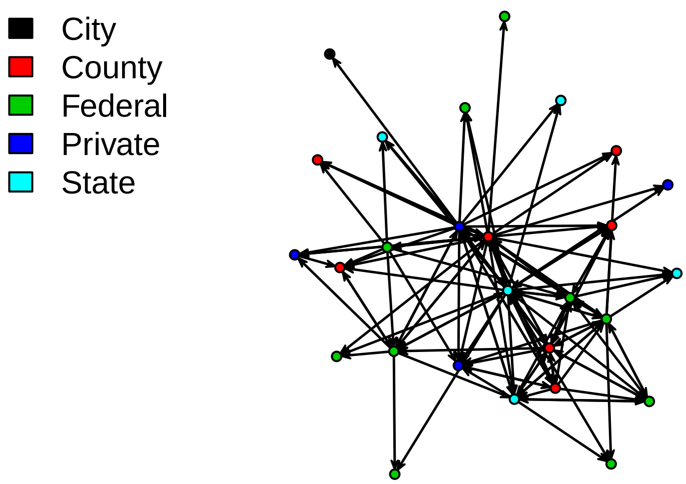
HurrFrederic

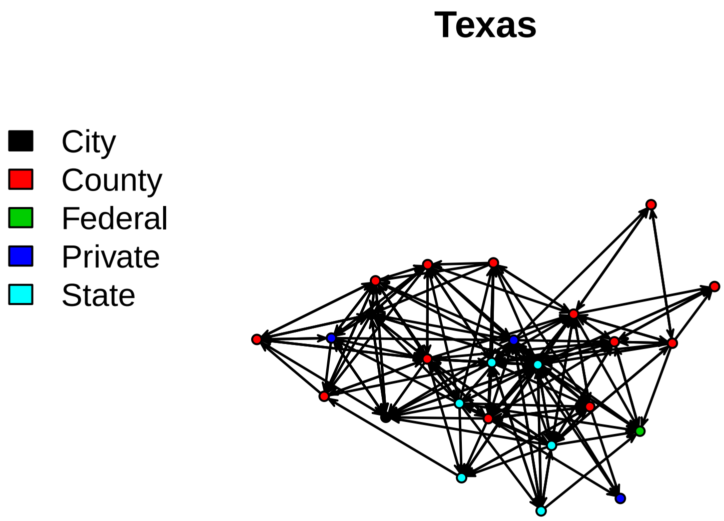


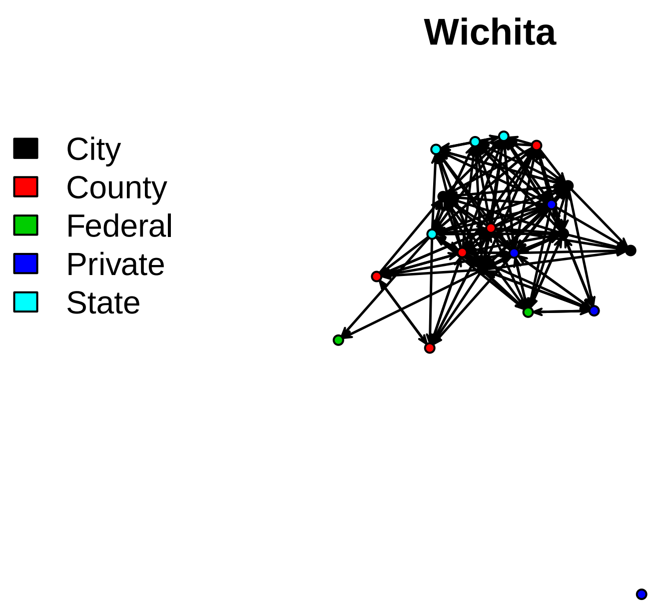

LakePomona
City
County
Federal
Private
State


MtSi
County
Federal
Private
State

MtStHelens




Texas
City
County
Federal
Private
State



1. There is no selective mixing observed in the network of Cheyenne SAR EMON.
2. The HurrFrederic network shows some level of homophily between organizations sponsored at city and county level.
3. LakePamona again doesn't show any selective mixing
4. MtSi network shows some level of homophily among organizations sponsored at private level.
5. No selective mixing observed among organizations at any level in the MtStHelens network.
6. The texas network also does not exhibit any selective mixing
7.

Most organizations report interaction at city level.

Most organization in the Lake Pamola report interaction with organizations at state level.

(b) Dyadic Mixing

Using the `mixingmatrix` command, obtain mixing matrices for all seven EMONs using "Sponsorship" as the relevant vertex attribute. For each network provide:

- The raw mixing matrix.

```
Cheyenne = mixingmatrix(emon[[1]], "Sponsorship")
Hurricane_Frederic = mixingmatrix(emon[[2]], "Sponsorship")
Lake_Pomona = mixingmatrix(emon[[3]], "Sponsorship")
Mt_Si = mixingmatrix(emon[[4]], "Sponsorship")
Mt_Helens = mixingmatrix(emon[[5]], "Sponsorship")
```

```
Texas_Hill = mixingmatrix(emon[[6]], "Sponsorship")
Wichita = mixingmatrix(emon[[7]], "Sponsorship")
```

- The matrix of mixing rates/block densities (this was called r in class).

```
#Frequency Counts of the Sponsorship attribute in each of the EMON networks
Cheyenne_f = count(get.vertex.attribute(emon[[1]], "Sponsorship"))
Hurricane_Frederic_f = count(get.vertex.attribute(emon[[2]], "Sponsorship"))
Lake_Pomona_f = count(get.vertex.attribute(emon[[3]], "Sponsorship"))
Mt_Si_f = count(get.vertex.attribute(emon[[4]], "Sponsorship"))
Mt_Helens_f = count(get.vertex.attribute(emon[[5]], "Sponsorship"))
Texas_Hill_f = count(get.vertex.attribute(emon[[6]], "Sponsorship"))
Wichita_f = count(get.vertex.attribute(emon[[7]], "Sponsorship"))
```

```
#Function for Possible Ties
possible_ties <- function(x) {
  mat1 = matrix(0, nrow = NROW(x), ncol = NROW(x))

  for (i in 1:nrow(mat1)){
    for(j in 1:ncol(mat1)){
      if (i == j) {
        mat1[i,j] = x[i,2] * (x[i,2] - 1)
      }
      else{
        mat1[i,j] = x[i,2] * x[j,2]
      }
    }
  }
 return (mat1)
}
```

```
#Calling the functions for possible ties
Cheyenne_pt = possible_ties(Cheyenne_f)
Hurricane_Frederic_pt = possible_ties(Hurricane_Frederic_f)
Lake_Pomona_pt = possible_ties(Lake_Pomona_f)
Mt_Si_pt = possible_ties(Mt_Si_f)
Mt_Helens_pt = possible_ties(Mt_Helens_f)
Texas_Hill_pt = possible_ties(Texas_Hill_f)
Wichita_pt = possible_ties(Wichita_f)
```

```
#Block Density Function
block_densities <- function(x,y){
  bd = x/y
  return (bd)
}
```

```
#Calling Block Densities
Cheyenne_bd = block_densities(Cheyenne$matrix, Cheyenne_pt)
Cheyenne_bd
```

```
##                To
## From                City    County County/City   Federal   Private
##   City        1.0000000 0.3333333   0.5000000 0.0000000 0.3750000
##   County      0.6666667 0.6666667   0.6666667 0.0000000 0.2500000
##   County/City 1.0000000 1.0000000             1.0000000 1.0000000
##   Federal     1.0000000 0.3333333   0.0000000           0.0000000
```

```
##   Private     0.2500000 0.3333333   0.7500000 0.0000000 0.1666667
##   State       1.0000000 0.6666667   1.0000000 0.6666667 0.5000000
##              To
## From              State
##   City        0.1666667
##   County      0.2222222
##   County/City 0.6666667
##   Federal     0.6666667
##   Private     0.1666667
##   State       1.0000000
```

```
Hurricane_Frederic_bd = block_densities(Hurricane_Frederic$matrix, Hurricane_Frederic_pt)
Hurricane_Frederic_bd
```

```
##          To
## From           City    County   Federal   Private     State
##   City    0.3928571 0.2500000 0.2500000 0.2916667 0.5000000
##   County  0.2500000 0.4333333 0.1666667 0.2222222 0.2222222
##   Federal 0.0000000 0.1666667           0.0000000 0.0000000
##   Private 0.3333333 0.2222222 0.0000000 0.0000000 0.1111111
##   State   0.3333333 0.2777778 0.0000000 0.1111111 0.1666667
```

```
Lake_Pomona_bd = block_densities(Lake_Pomona$matrix, Lake_Pomona_pt)
Lake_Pomona_bd
```

```
##          To
## From           City    County   Federal   Private     State
##   City    0.1666667 0.4500000 0.3750000 0.4500000 0.6250000
##   County  0.1500000 0.5000000 0.4000000 0.3600000 0.6500000
##   Federal 0.0000000 0.3000000 0.0000000 0.2000000 0.5000000
##   Private 0.1500000 0.4400000 0.4000000 0.2500000 0.5500000
##   State   0.1250000 0.6500000 0.5000000 0.3500000 0.5833333
```

```
Mt_Helens_bd = block_densities(Mt_Helens$matrix, Mt_Helens_pt)
Mt_Helens_bd
```

```
##          To
## From           City    County   Federal   Private     State
##   City              0.0000000 0.0000000 0.0000000 0.0000000
##   County  0.0000000 0.2619048 0.2142857 0.1785714 0.2285714
##   Federal 0.0000000 0.1857143 0.1000000 0.1000000 0.1800000
##   Private 0.2500000 0.3928571 0.1000000 0.1666667 0.2000000
##   State   0.0000000 0.1714286 0.2400000 0.2500000 0.2000000
```

```
Texas_Hill_bd = block_densities(Texas_Hill$matrix, Texas_Hill_pt)
Texas_Hill_bd
```

```
##          To
## From            City     County    Federal    Private      State
##   City    1.00000000 0.23076923 0.00000000 0.33333333 0.16666667
##   County  0.34615385 0.30128205 0.23076923 0.20512821 0.34615385
##   Federal 0.00000000 0.07692308            0.00000000 0.16666667
##   Private 0.66666667 0.23076923 0.33333333 0.33333333 0.22222222
##   State   0.50000000 0.33333333 0.66666667 0.27777778 0.56666667
```

```
Wichita_bd = block_densities(Wichita$matrix, Wichita_pt)
Wichita_bd
```

```
##          To
## From           City    County   Federal   Private     State
##   City    0.6500000 0.2400000 0.5000000 0.4000000 0.6000000
##   County  0.4800000 0.7500000 0.2000000 0.3000000 0.5000000
##   Federal 0.1000000 0.0000000 0.0000000 0.2500000 0.0000000
##   Private 0.5500000 0.3500000 0.3750000 0.4166667 0.3750000
##   State   0.3500000 0.2500000 0.2500000 0.1875000 0.6666667
```

```
Mt_Si_bd = block_densities(Mt_Si$matrix, Mt_Si_pt)
Mt_Si_bd
```

```
##          To
## From         County   Federal   Private     State
##   County  0.0000000 0.0000000 0.2857143 0.3333333
##   Federal 0.0000000           0.0000000 0.0000000
##   Private 0.2857143 0.0000000 0.3333333 0.0952381
##   State   0.3333333 0.0000000 0.2380952 0.0000000
```

- The matrix of marginal z-scores, using the Poisson approximation.

```
calculate_zscore <- function(observed_matrix) {
  row_sums = rowSums(observed_matrix)
  col_sums = colSums(observed_matrix)
  gr_sum = sum(observed_matrix)

  expected_matrix = matrix(0, nrow = NROW(observed_matrix), ncol = NROW(observed_matrix))
  diff = matrix(0, nrow = NROW(observed_matrix), ncol = NROW(observed_matrix))
  zscore = matrix(0, nrow = NROW(observed_matrix), ncol = NROW(observed_matrix))

  for (i in 1:length(row_sums)){
    for (j in 1:length(col_sums)){

      expected_matrix[i,j] = row_sums[i] * col_sums[j] /gr_sum
    }
  }

  diff = observed_matrix - expected_matrix
  zscore = diff/sqrt(expected_matrix)
  return (zscore)
}
```

```
#Calling the Expected Matrix

Cheyenne_z = calculate_zscore(Cheyenne$matrix)
Cheyenne_z
```

```
##              To
## From                 City      County County/City     Federal     Private
##   City         0.03449558 -0.11453883  0.02439206 -0.57035183  0.75027884
##   County       0.41416249  0.20279176  0.29285710 -0.73632104 -0.14028084
##   County/City -0.37342560  0.06376727 -1.14070365  0.85982004  0.86634738
##   Federal      0.87933284 -0.18659924 -0.73632104 -0.42511515 -1.04131520
##   Private     -0.48793496  0.49012491  1.33949764 -0.68547758 -0.48793496
##   State       -0.11530203 -0.37373245 -0.08153085  0.92966915 -0.11530203
##              To
## From                State
```

```
##   City        -0.49124417
##   County      -0.43173942
##   County/City -0.11453883
##   Federal      1.15337814
##   Private     -0.22795108
##   State        0.33155599
```

```
Hurricane_Frederic_z = calculate_zscore(Hurricane_Frederic$matrix)
Hurricane_Frederic_z
```

```
##          To
## From             City      County     Federal     Private       State
##   City    -0.27034191 -1.06797580  0.50883312  0.59483309  1.24638151
##   County  -0.63409288  0.91800270  0.14584075  0.29168150 -0.52096833
##   Federal -0.65094455  1.29152574 -0.15944820 -0.31889640 -0.39056673
##   Private  1.06157247  0.07336725 -0.57489866 -1.14979733 -0.69808621
##   State    0.65212387  0.26115161 -0.61754023 -0.42541660 -0.85157068
```

```
Lake_Pomona_z = calculate_zscore(Lake_Pomona$matrix)
Lake_Pomona_z
```

```
##          To
## From             City      County     Federal     Private       State
##   City    -0.15384773 -0.39241554 -0.18842422  0.69814602 -0.01066537
##   County   0.22476599 -0.60937912  0.02378972  0.19545246  0.33160211
##   Federal -0.77981287  0.12119654 -0.95507181  0.03874921  0.76380630
##   Private  0.46362092  0.13302423  0.29846809 -0.86722929  0.20594417
##   State   -0.15384773  0.85656296  0.35837547 -0.05059029 -0.95775021
```

```
Mt_Si_z = calculate_zscore(Mt_Si$matrix)
Mt_Si_z
```

```
##          To
## From           County Federal     Private       State
##   County  -1.04446594         -0.08891084  1.49240501
##   Federal
##   Private  0.19069252          0.01623283 -0.27247463
##   State    0.64465837          0.05487696 -0.92113237
```

```
Mt_Helens_z = calculate_zscore(Mt_Helens$matrix)
Mt_Helens_z
```

```
##          To
## From             City      County     Federal     Private       State
##   City
##   County  -0.56309251 -0.55470020  0.65062389 -0.03248611  0.02598888
##   Federal -0.53343495  0.39036003 -0.70607534 -0.25909698  0.70718233
##   Private  1.94158978  1.35400640 -1.17933675 -0.50942702 -0.22299447
##   State   -0.46852129 -1.00000000  1.08650256  0.79388329 -0.63510663
```

```
Texas_Hill_z = calculate_zscore(Texas_Hill$matrix)
Texas_Hill_z
```

```
##          To
## From             City      County     Federal     Private       State
##   City     0.55427400  0.10769589 -0.71842121  0.86245755 -0.71134299
##   County  -0.49509804  0.30142026 -0.51872399 -0.20176593  0.24145120
##   Federal -0.47519096  0.04396666 -0.29329423 -0.42754614  0.60984958
```

```
##   Private  1.15921474 -0.18422111  0.15071514  0.12724890 -0.63365450
##   State   -0.21429896 -0.33269948  0.95310900 -0.13076548  0.27502654
```

```r
Wichita_z = calculate_zscore(Wichita$matrix)
Wichita_z
```

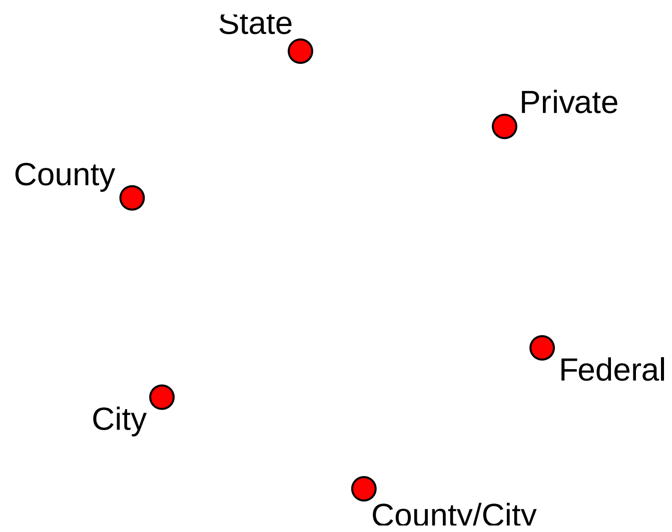
```
##          To
## From              City       County      Federal      Private        State
##   City     0.001861891 -1.199658834  0.773659100  0.342857627  0.419913120
##   County  -0.353488671  1.594427963 -0.853149915 -0.463668863 -0.264601749
##   Federal  0.121218477 -0.815125469 -0.491539152  2.181971415 -0.851370786
##   Private  0.504333153 -0.032773156  0.263379399 -0.067991417 -0.622730780
##   State   -0.140794399 -0.228175956 -0.009459675 -0.511708547  0.797385736
```

```r
class(Wichita_z)
```

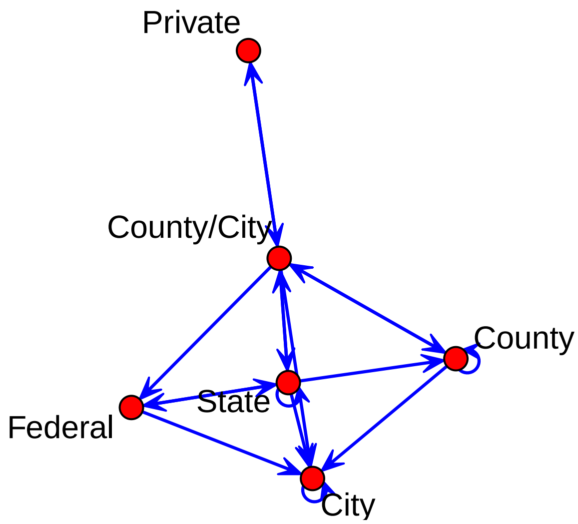
```
## [1] "table"
```

- A plot of the reduced form blockmodel, with edge widths set based on mixing rates.

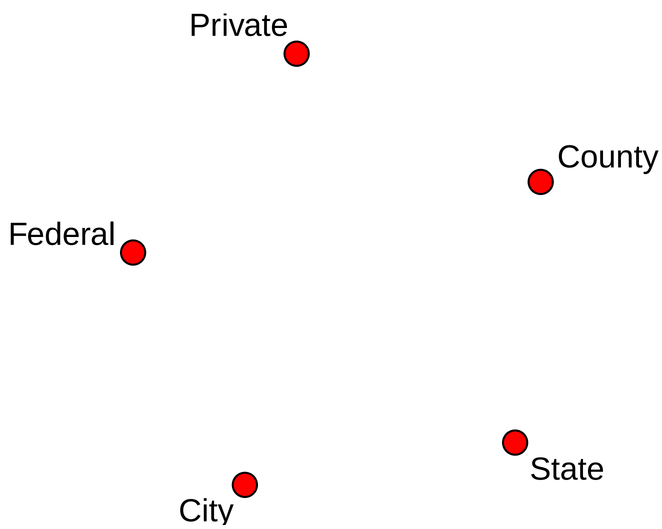
```r
gplot(abs(Cheyenne_z)>2,edge.col=sign(Cheyenne_z)+3,label=rownames(Cheyenne_z),boxed.lab=FALSE,diag=TRU
```



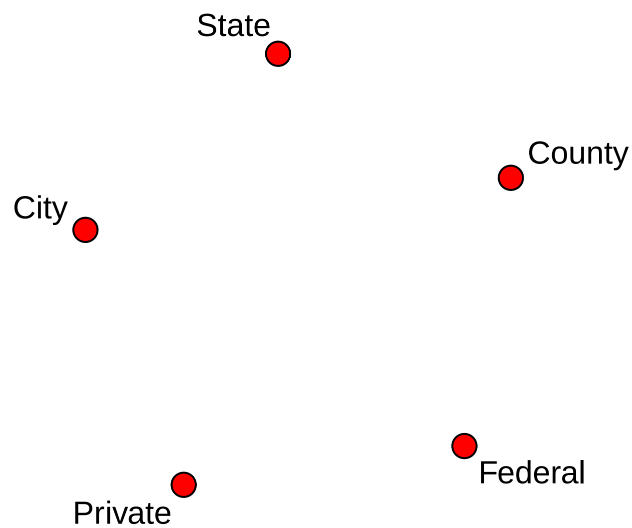

```r
gplot(abs(Cheyenne_bd)>0.5,edge.col=sign(Cheyenne_bd)+3,label=rownames(Cheyenne_bd),boxed.lab=FALSE,dia
```



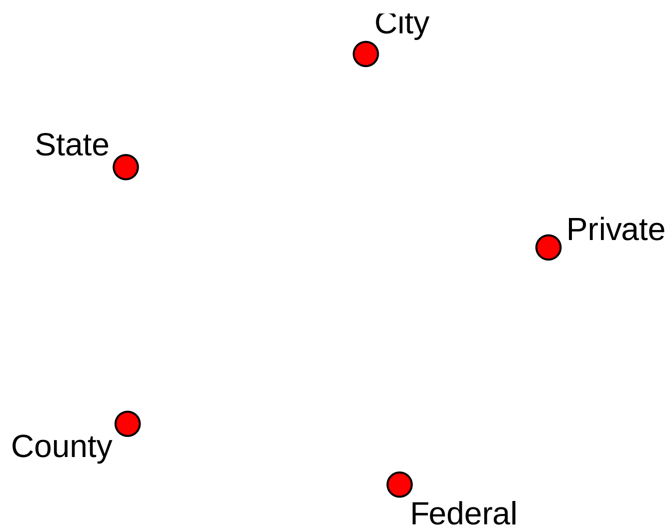

```
gplot(abs(Hurricane_Frederic_z)>2,edge.col=sign(Hurricane_Frederic_z)+3,label=rownames(Hurricane_Freder
```



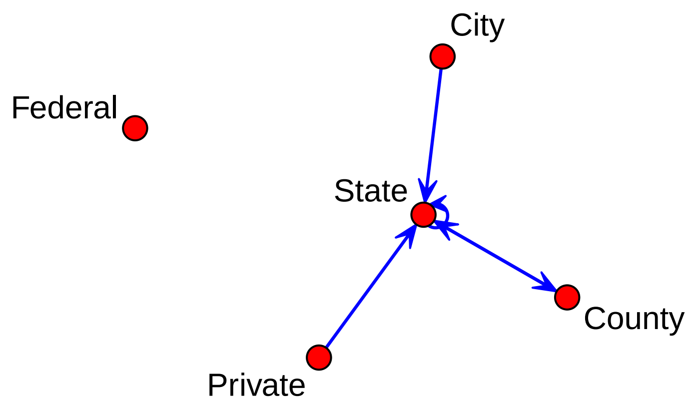

```
gplot(abs(Hurricane_Frederic_bd)>0.5,edge.col=sign(Hurricane_Frederic_bd)+3,label=rownames(Hurricane_Fre
```



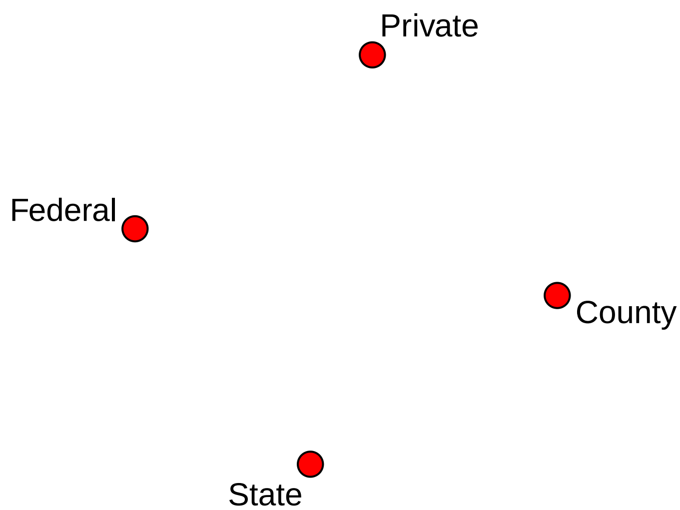

```
gplot(abs(Lake_Pomona_z)>2,edge.col=sign(Lake_Pomona_z)+3,label=rownames(Lake_Pomona_z),boxed.lab=FALSE
```



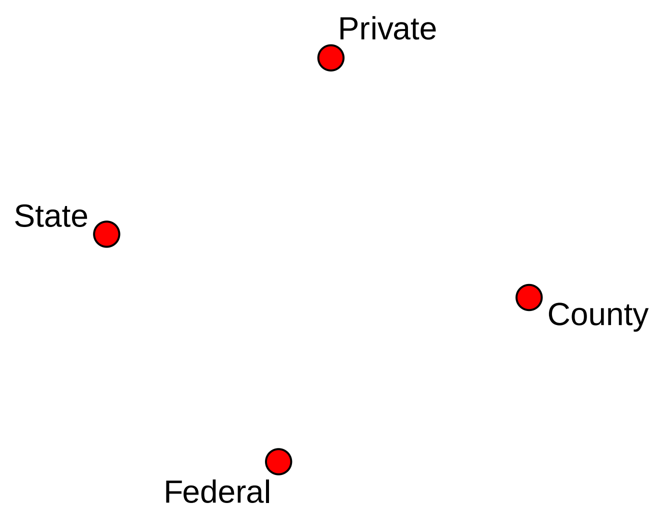

```
gplot(abs(Lake_Pomona_bd)>0.5,edge.col=sign(Lake_Pomona_bd)+3,label=rownames(Lake_Pomona_bd),boxed.lab=
```



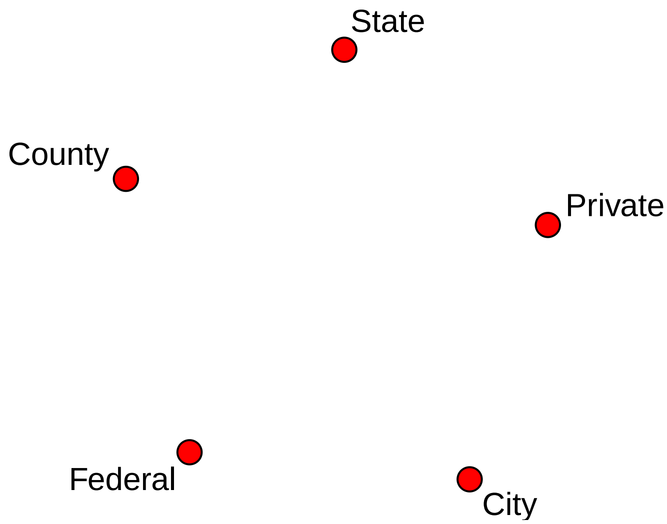

```
gplot(abs(Mt_Si_z)>2,edge.col=sign(Mt_Si_z)+3,label=rownames(Mt_Si_z),boxed.lab=FALSE,diag=TRUE)
```




```
gplot(abs(Mt_Si_bd)>0.5,edge.col=sign(Mt_Si_bd)+3,label=rownames(Mt_Si_bd),boxed.lab=FALSE,diag=TRUE)
```



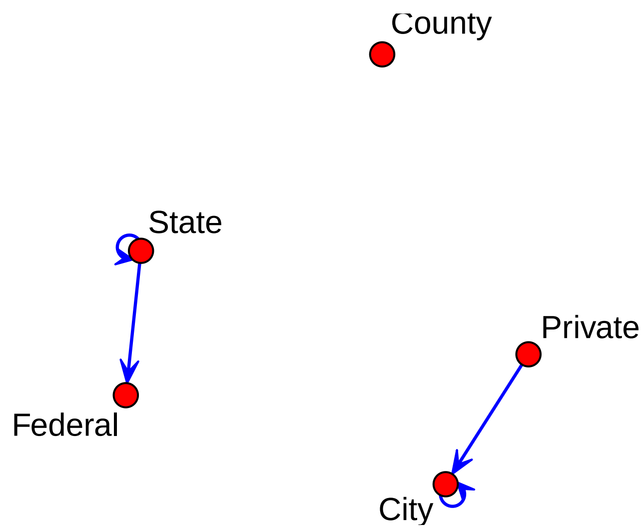

```
gplot(abs(Mt_Helens_z)>2,edge.col=sign(Mt_Helens_z)+3,label=rownames(Mt_Helens_z),boxed.lab=FALSE,diag='
```



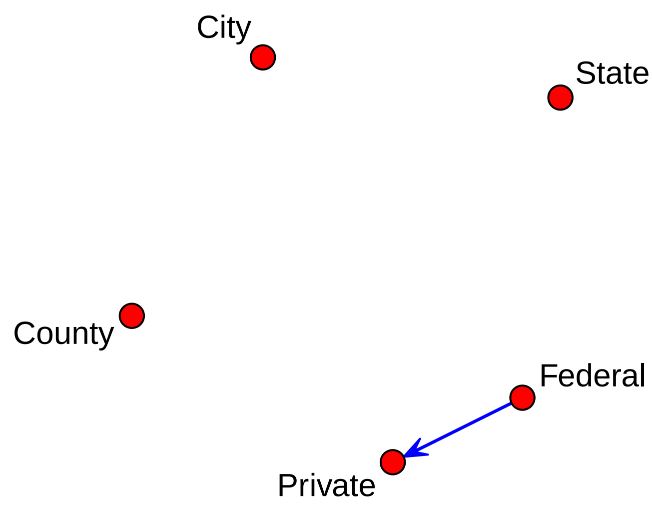

```
gplot(abs(Mt_Helens_bd)>0.5,edge.col=sign(Mt_Helens_bd)+3,label=rownames(Mt_Helens_bd),boxed.lab=FALSE,c
```


```
gplot(abs(Texas_Hill_z)>2,edge.col=sign(Texas_Hill_z)+3,label=rownames(Texas_Hill_z),boxed.lab=FALSE,di
```

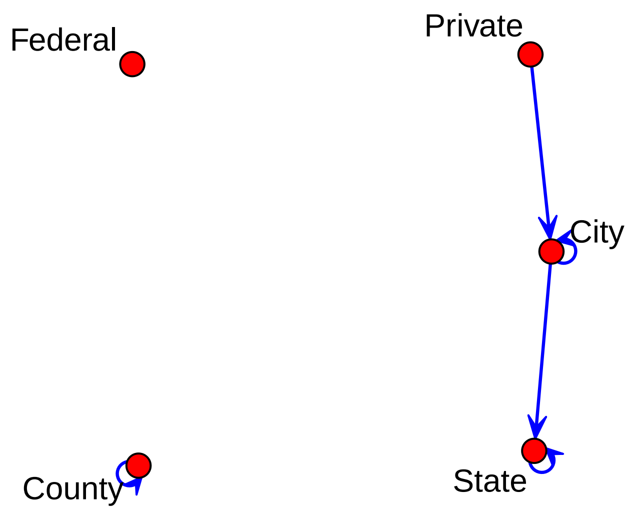

```
gplot(abs(Texas_Hill_bd)>0.5,edge.col=sign(Texas_Hill_bd)+3,label=rownames(Texas_Hill_bd),boxed.lab=FALS
```




```
gplot(abs(Wichita_z)>2,edge.col=sign(Wichita_z)+3,label=rownames(Wichita_z),boxed.lab=FALSE,diag=TRUE)
```




```
gplot(abs(Wichita_bd)>0.5,edge.col=sign(Wichita_bd)+3,label=rownames(Wichita_bd),boxed.lab=FALSE,diag=T
```



- A disucssion of your findings.

1. Keeping a threshold of 0.5 for the block densities, we see that some network interactions show a value above 0.5 which means that there is a good tie probability between them. For example, in the Cheyenne network, the block density of the organizations sponsored at state at County/City level and Private level is 0.75. This means that the probability of ties between these two organizations are strong.
2. The z-score values show us deviations from mixing rate distribution of the expected network matrices (standard). All the networks except the Wichita network show normal ranges of z-score values. The Wichita network shows a significant departure and hence is an unusual network.

(c) Discussion

Based on your analysis in parts (a)-(b) how would you describe the overall pattern of communication mixing in the Drabek et al. SAR EMONs?

1. From the plots of the network structures, we see that that the networks in the Drabek et al. SAR EMONS have few cases of selective mixing. Rest of the networks are heterogenous with interactions among organizations at different levels.

2. Looking at the block densities, we see that the commmunication mixing between organizations with high block densities(>0.5) tend to include interactions among state and privately sponsored organizations.