INFX 576: Problem Set 3 - Selective Mixing*

Jay Chauhan

Due: Thursday, February 2, 2017

Collaborators: Avanti Chande, Gossudin 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.

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

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 ith network object. (See `?emon` for details regarding the dataset.)

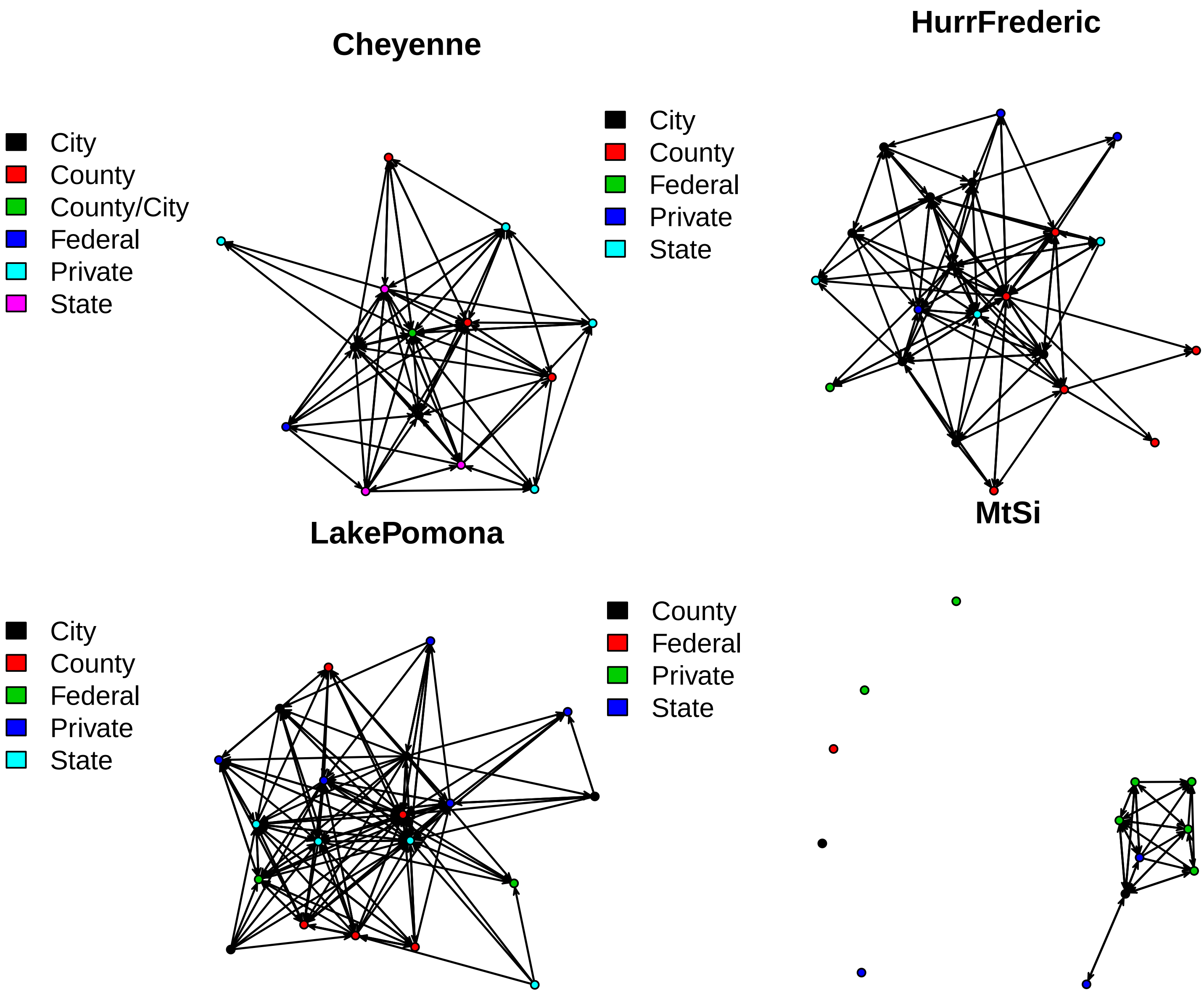
(a) Visualizing Networks

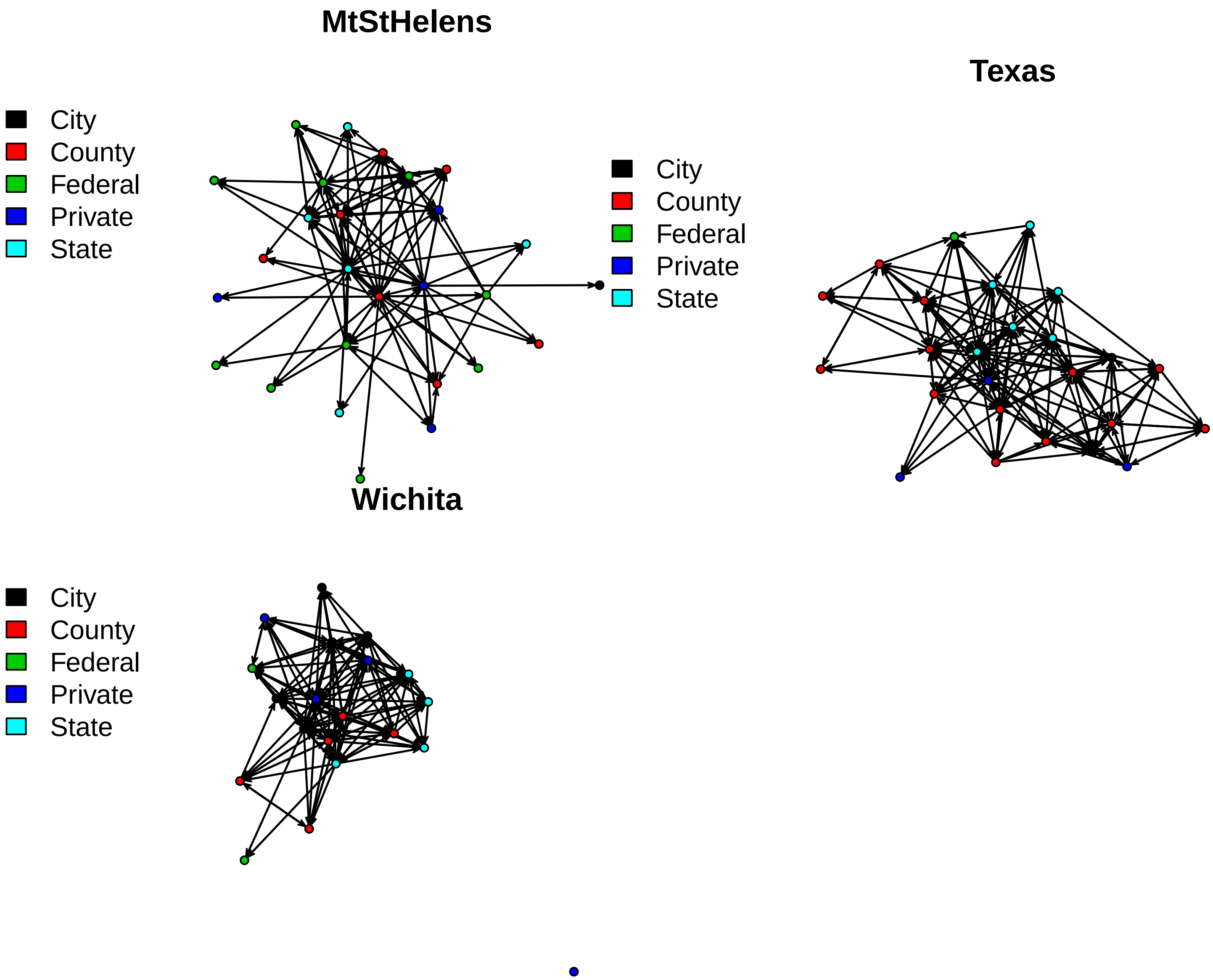
This data set consists of seven individual network data sets of emergent multiorganizational networks (EMONs) in the context of search and rescue activities. These data sets are: the Cheyenne SAR EMON, the Hurricane Frederic SAR EMON, the Lake Pomona SAR EMON, the Mt. Si SAR EMON, the Mt. St. Helens 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.

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

```
#plot the seven emon networks by coloring the vertices by Sponsorship attribure
nam<-names(emon)
for(i in 1:7){
plot(emon[[i]],vertex.col="Sponsorship", main=nam[i])
vals<-sort(unique(emon[[i]]%v%"Sponsorship"))   # Let's add a legend....
legend("topleft",fill=1:length(vals),legend=vals,bty="n")
}
```





- The network of Cheyenne shows no significant slective mixing with respect to the interactions between organizations sponsored at different level.
- We observe some level of homophily in the Hurricane Fredric network between the organizations sponsored at the City and County levels
- The network of Lake Pomona shows no significant slective mixing with respect to the interactions between organizations sponsored at different level.
- We observe some level of homophily in the Mt. Si network between the organizations sponsored at the Private level
- The network of Mt. St. Helens shows no significant slective mixing with respect to the interactions between organizations sponsored at different level.
- The network of Texas shows no significant slective mixing with respect to the interactions between organizations sponsored at different level.
- The network of Wichita shows no significant slective mixing with respect to the interactions between organizations sponsored at different 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.
- The matrix of mixing rates/block densities (this was called r in class).
- The matrix of marginal z-scores, using the Poisson approximation.
- A plot of the reduced form blockmodel, with edge widths set based on mixing rates.
- A disucssion of your findings.

```r
#create mixing matrix for all networks
che_spon<-mixingmatrix(emon[[1]],"Sponsorship")
hurr_spon<-mixingmatrix(emon[[2]],"Sponsorship")
pom_spon<-mixingmatrix(emon[[3]],"Sponsorship")
si_spon<-mixingmatrix(emon[[4]],"Sponsorship")
hel_spon<-mixingmatrix(emon[[5]],"Sponsorship")
tex_spon<-mixingmatrix(emon[[6]],"Sponsorship")
wich_spon<-mixingmatrix(emon[[7]],"Sponsorship")
```

```r
#get count of number of vertex in each network
che_obs_vertex<-as.data.frame(table(get.vertex.attribute(emon[[1]],attrname = "Sponsorship")))

hur_obs_vertex<-as.data.frame(table(get.vertex.attribute(emon[[2]],attrname = "Sponsorship")))

pom_obs_vertex<-as.data.frame(table(get.vertex.attribute(emon[[3]],attrname = "Sponsorship")))

si_obs_vertex<-as.data.frame(table(get.vertex.attribute(emon[[4]],attrname = "Sponsorship")))

hel_obs_vertex<-as.data.frame(table(get.vertex.attribute(emon[[5]],attrname = "Sponsorship")))

tex_obs_vertex<-as.data.frame(table(get.vertex.attribute(emon[[6]],attrname = "Sponsorship")))

wich_obs_vertex<-as.data.frame(table(get.vertex.attribute(emon[[7]],attrname = "Sponsorship")))
```

```r
#calculate all possible ties
possible_ties<-function(network_ver){
  ties_mat<-matrix(0,nrow=nrow(network_ver["Freq"]),ncol=nrow(network_ver["Freq"]))
  for(i in 1:nrow(network_ver["Freq"])){
    for(j in 1:nrow(network_ver["Freq"])){
      #print(i)
      if(i==j)
        ties_mat[i,j]<-network_ver[i,2]*((network_ver[j,2])-1)
      else
        ties_mat[i,j]<-network_ver[i,2]*network_ver[j,2]
    }
  }
  return(ties_mat)
}
```

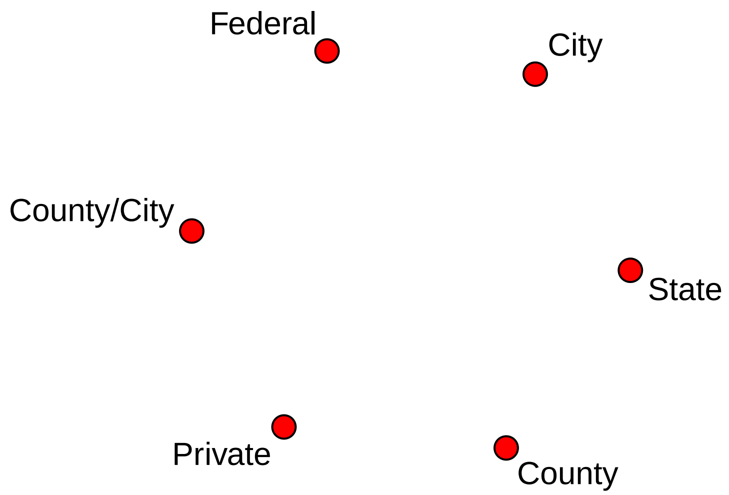
```r
#calculate block densities
block_den<-function(ties_mat,obs_mat){
  block_mat<-matrix(0,nrow=nrow(ties_mat),ncol=nrow(ties_mat))
        block_mat<-(obs_mat$matrix)/ties_mat
  return(block_mat)
}
```

```r
#calculate possible ties for all networks
che_ties_mat<-possible_ties(che_obs_vertex)
hur_ties_mat<-possible_ties(hur_obs_vertex)
pom_ties_mat<-possible_ties(pom_obs_vertex)
si_ties_mat<-possible_ties(si_obs_vertex)
hel_ties_mat<-possible_ties(hel_obs_vertex)
tex_ties_mat<-possible_ties(tex_obs_vertex)
wich_ties_mat<-possible_ties(wich_obs_vertex)

#calculate block denisities for all network
che_block_den<-block_den(che_ties_mat,che_spon)
hur_block_den<-block_den(hur_ties_mat,hurr_spon)
pom_block_den<-block_den(pom_ties_mat,pom_spon)
si_block_den<-block_den(si_ties_mat,si_spon)
hel_block_den<-block_den(hel_ties_mat,hel_spon)
tex_block_den<-block_den(tex_ties_mat,tex_spon)
wich_block_den<-block_den(wich_ties_mat,wich_spon)
```

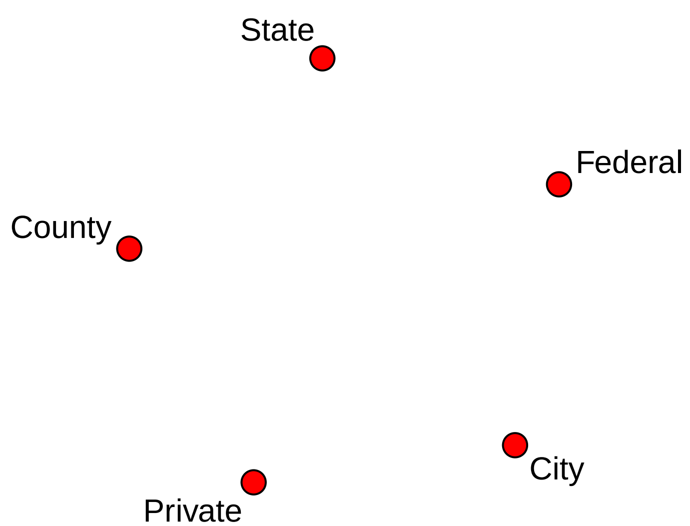
```r
#calculate expected ties
expected_ties<-function(obs_mat){
  exp_mat<-matrix(0,nrow=nrow(obs_mat),ncol=nrow(obs_mat))
  row_sum<-rowSums(obs_mat)
  col_sum<-colSums(obs_mat)
  gr_sum<-sum(obs_mat)
  for(i in 1:length(row_sum)){
    for(j in 1:length(col_sum)){
      exp_mat[i,j]<-row_sum[i]*col_sum[j]/gr_sum
    }
  }
  return(exp_mat)
}
```

```r
#calculate expected ties for all networks
che_exp_ties<-expected_ties(che_spon$matrix)
hur_exp_ties<-expected_ties(hurr_spon$matrix)
pom_exp_ties<-expected_ties(pom_spon$matrix)
si_exp_ties<-expected_ties(si_spon$matrix)
hel_exp_ties<-expected_ties(hel_spon$matrix)
tex_exp_ties<-expected_ties(tex_spon$matrix)
wich_exp_ties<-expected_ties(wich_spon$matrix)

#calculate mixing z-score for all netowrks
che_z<-(che_spon$matrix-che_exp_ties)/sqrt(che_exp_ties)
hur_z<-(hurr_spon$matrix-hur_exp_ties)/sqrt(hur_exp_ties)
pom_z<-(pom_spon$matrix-pom_exp_ties)/sqrt(pom_exp_ties)
si_z<-(si_spon$matrix-si_exp_ties)/sqrt(si_exp_ties)
hel_z<-(hel_spon$matrix-hel_exp_ties)/sqrt(hel_exp_ties)
tex_z<-(tex_spon$matrix-tex_exp_ties)/sqrt(tex_exp_ties)
wich<-(wich_spon$matrix-wich_exp_ties)/sqrt(wich_exp_ties)
```

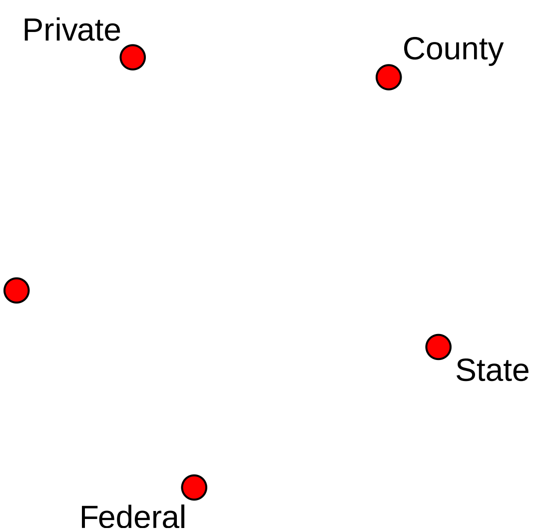
```r
# plot z-score for all the networks
gplot(abs(che_z)>2,edge.col=sign(che_z)+3,
      label=rownames(che_z),boxed.lab=FALSE,diag=TRUE)
```



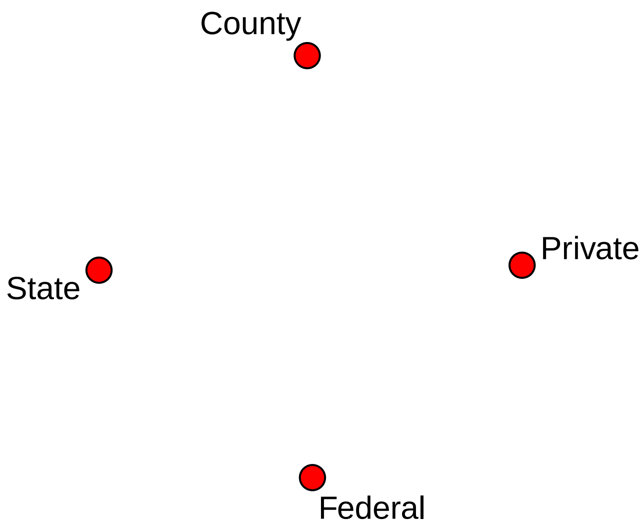

```
gplot(abs(hur_z)>2,edge.col=sign(hur_z)+3,
      label=rownames(hur_z),boxed.lab=FALSE,diag=TRUE)
```



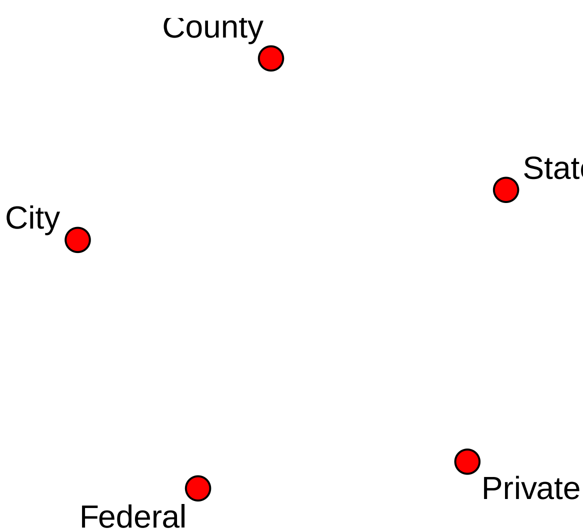

```
gplot(abs(pom_z)>2,edge.col=sign(pom_z)+3,
      label=rownames(pom_z),boxed.lab=FALSE,diag=TRUE)
```



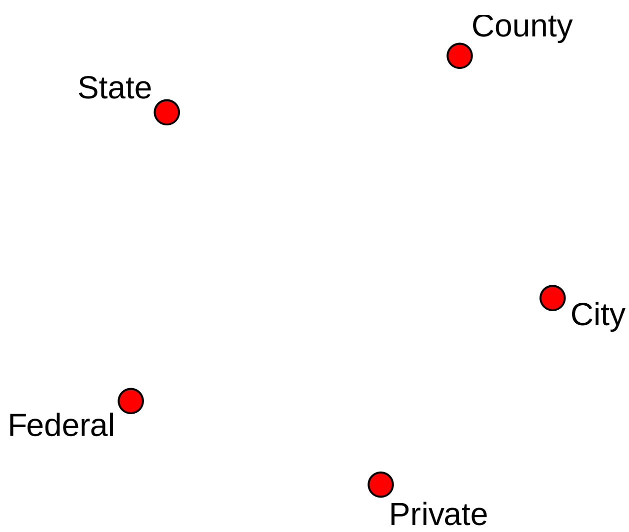
```
gplot(abs(si_z)>2,edge.col=sign(si_z)+3,
      label=rownames(si_z),boxed.lab=FALSE,diag=TRUE)
```



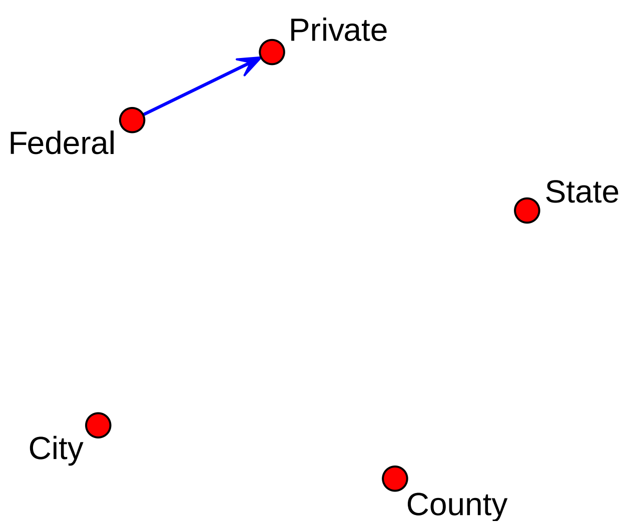

```
gplot(abs(hel_z)>2,edge.col=sign(hel_z)+3,
      label=rownames(hel_z),boxed.lab=FALSE,diag=TRUE)
```



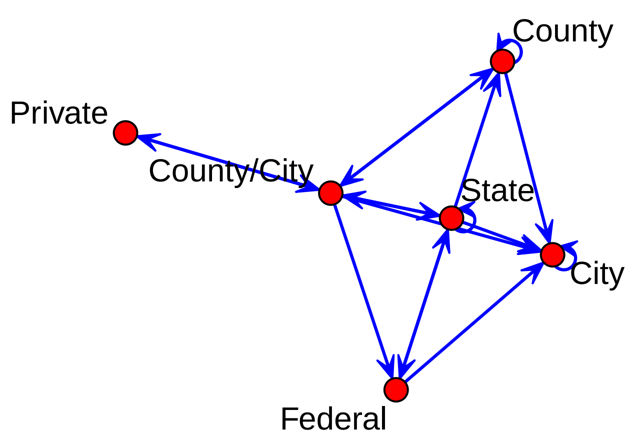

```
gplot(abs(tex_z)>2,edge.col=sign(tex_z)+3,
      label=rownames(tex_z),boxed.lab=FALSE,diag=TRUE)
```



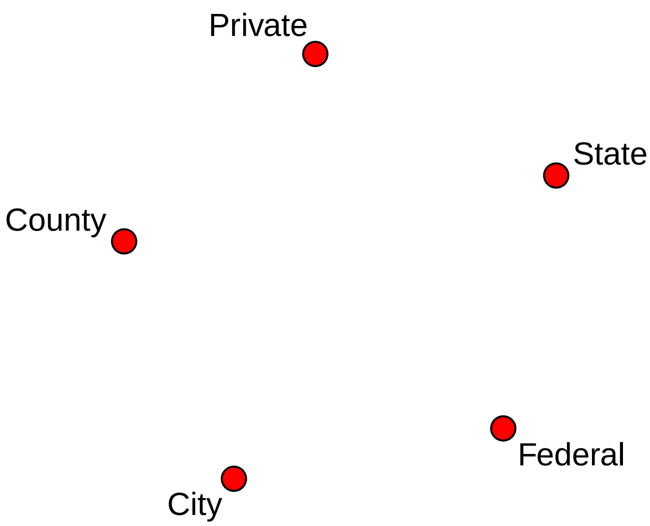

```r
gplot(abs(wich)>2,edge.col=sign(wich)+3,
      label=rownames(wich),boxed.lab=FALSE,diag=TRUE)
```



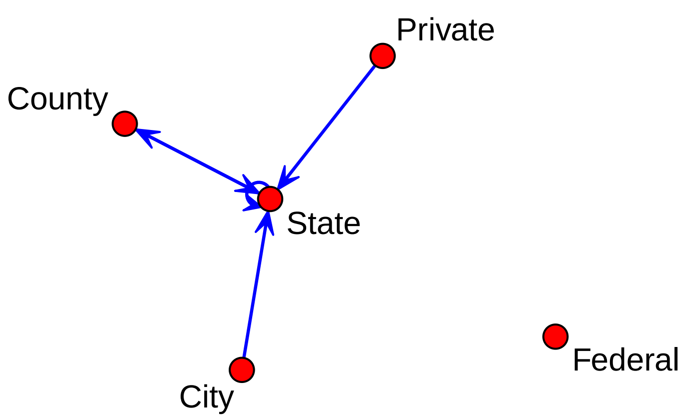

```r
# plot block densities for all the netoworks
gplot(abs(che_block_den)>0.5,edge.col=sign(che_block_den)+3,
      label=rownames(che_block_den),boxed.lab=FALSE,diag=TRUE)
```



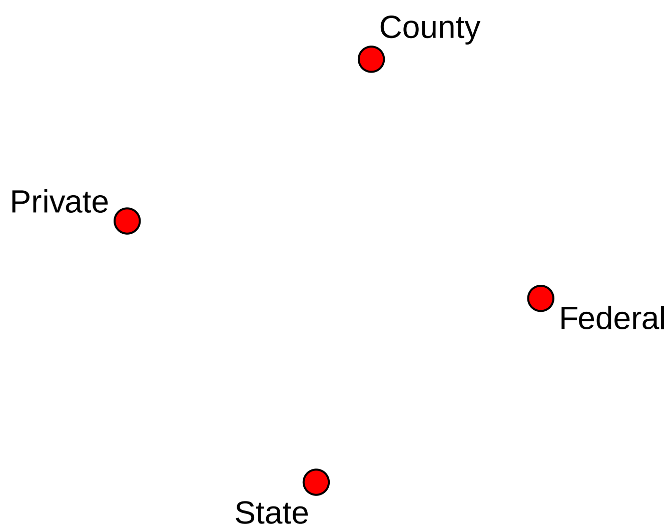
```
gplot(abs(hur_block_den)>0.5,edge.col=sign(hur_block_den)+3,
      label=rownames(hur_block_den),boxed.lab=FALSE,diag=TRUE)
```



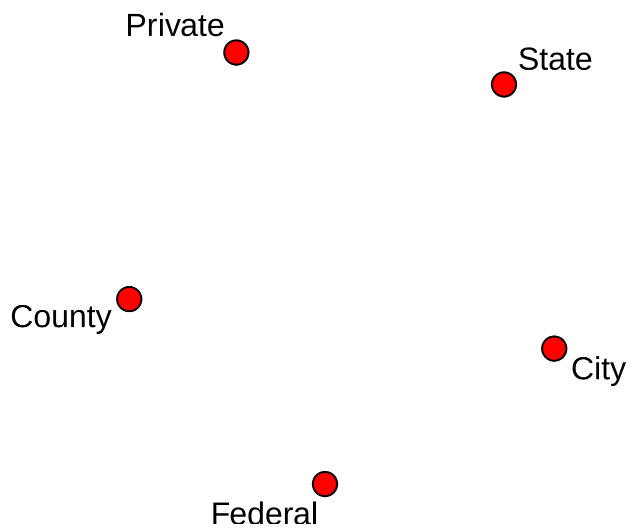

```
gplot(abs(pom_block_den)>0.5,edge.col=sign(pom_block_den)+3,
      label=rownames(pom_block_den),boxed.lab=FALSE,diag=TRUE)
```



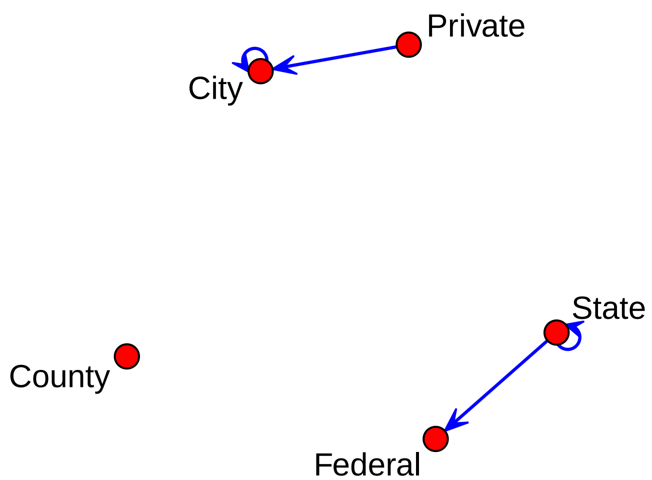

```
gplot(abs(si_block_den)>0.5,edge.col=sign(si_block_den)+3,
      label=rownames(si_block_den),boxed.lab=FALSE,diag=TRUE)
```



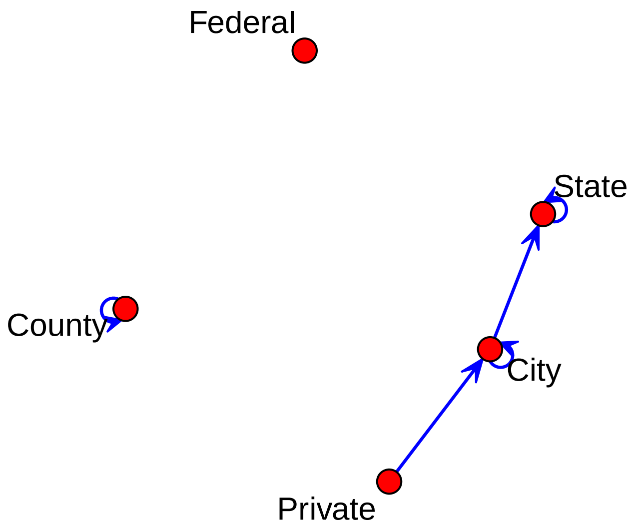

```
gplot(abs(hel_block_den)>0.5,edge.col=sign(hel_block_den)+3,
      label=rownames(hel_block_den),boxed.lab=FALSE,diag=TRUE)
```




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



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



- Observing the block densities of all the 7 networks, we see that there are a few block densities having a significant value which is greater than 0.5 with respect to the interactions between organizations at various levels. From the plots above we see that there is a noticable deviation from the mean of interaction of organization from the Federal level to the Private level in the Wichita 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?

From the above results of the 2 parts, we see that there were only a few cases of selective mising. This was validated from the mixing rates that we calculated in part (b), which showed us there there were indeed a few. When the block densities of the networks are studied coupled with the graphs of block densities of each block, the networks wherever we see an observed value of 0.5 or more for the block densities, we see an interatction between the organizations funded at the state and private levels