INFX 576: Problem Set 4 - Selective Mixing*

Harkar Talwar

Due: Friday, April 27, 2018

Collaborators: Prateek Tripathi, Aakash Agrawal

Instructions:

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

1. Download the `problemset4.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_ps4.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)
```

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.)

```r
# Loading the EMON data
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.

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

```r
# Generic function to plot network data for the EMON networks
plot.emon.network = function(net.obj, color.attr, name, sub.title) {
  # Plot the network, vertices colored by the passed attribute 'color.attr'
  plot(net.obj, main = name, vertex.col = color.attr,
       vertex.cex = 2.0, edge.lwd = 0.5, sub = sub.title)
  # Create a legend based on the vertex attribute specified by 'color.attr'
  values = sort(unique(net.obj %v% color.attr))
  legend(x = "topleft", y = NULL, fill = 1:length(values),
         bty = "n", legend = values, cex = 0.75)
}
```

Network 1

```r
plot.emon.network(emon[[1]], "Sponsorship", names(emon)[1], "Figure 1a.")
```

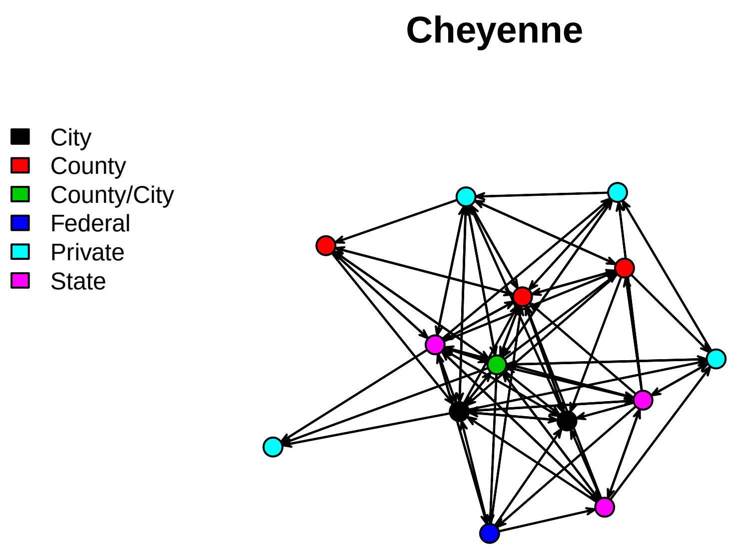

Figure 1a.

Observations

For the Cheyenne network (Figure 1a.), there does not seem to be any visible selective mixing, entities with different sponsorships seem to be collaborating with each other. Another observation is that entities at the Federal and Private level don't seem to be interacting with each other. County and City authorities seem to be the ones coordinating the efforts, with a high number of ties dirercted towards them.

Network 2

```r
plot.emon.network(emon[[2]], "Sponsorship", names(emon)[2], "Figure 1b.")
```

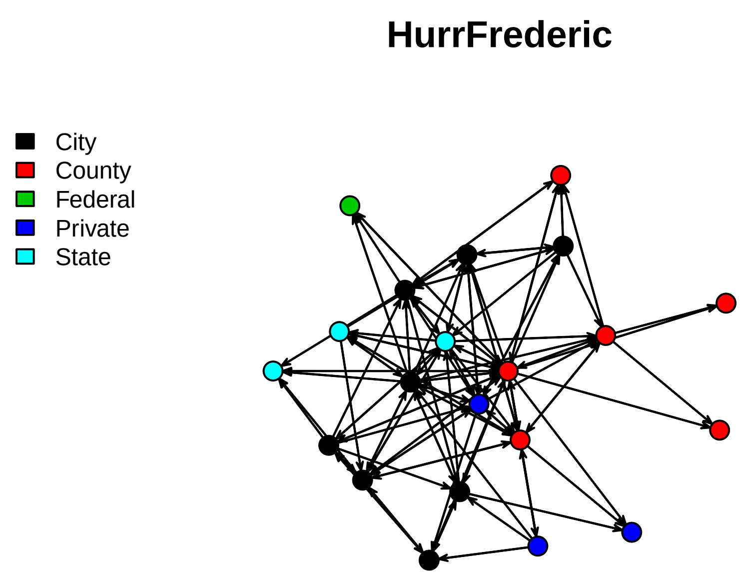

Figure 1b.

Observations

For the HurrFrederic network (Figure 1b.), entities sponsored at the City and County levels seem to be showing homophily, with a high proportion of interactions taking place with entities of the same type. This may also appear to be happening because the bulk of the entities in the network are those at the City and County levels. Further, we see that there is no interaction between entities at the private level, and also from Federal to Private entities and vice-versa.

Network 3

```r
plot.emon.network(emon[[3]], "Sponsorship", names(emon)[3], "Figure 1c.")
```

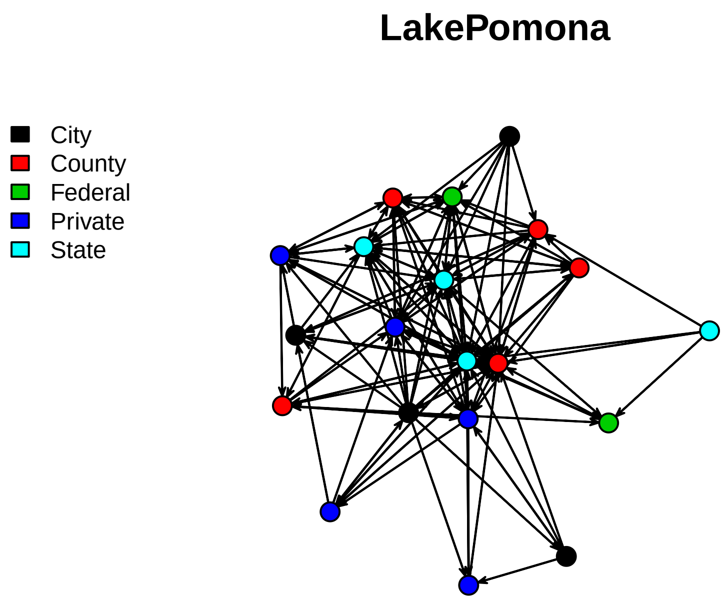

Figure 1c.

Observations

For the Lake Pomona network, the plot does not indicate strong ties within organizations sponsored at the same level, or between organizations at specific levels. Therefore, selective mixing does not appear to be taking place. We also observe that there are no interactions between the two organizations at the Federal level.

Network 4

```
plot.emon.network(emon[[4]], "Sponsorship", names(emon)[4], "Figure 1d.")
```

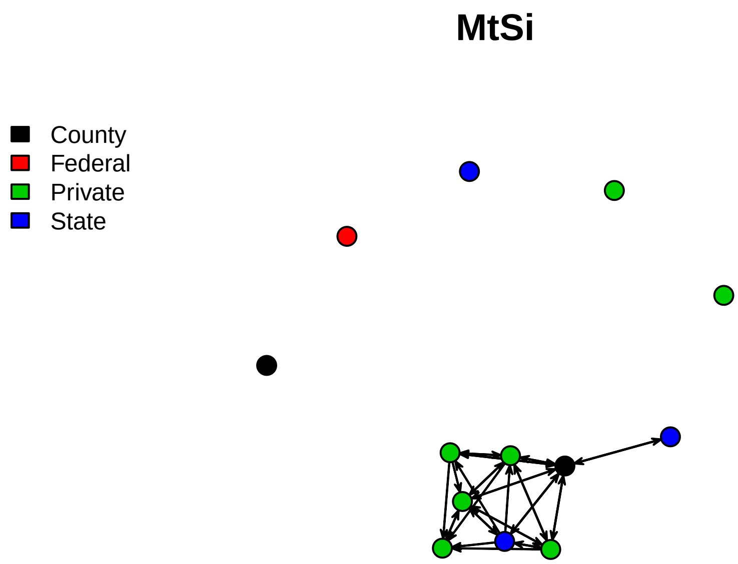

Figure 1d.

Observations

In the MtSi network (figure 1d.), homophily seems to be demonstrated among the private organizations, most of which have strong ties to other private organizations, and very few ties to entities at the County, State or Federal level. Further, there are organizations at all levels which have no interactions with other entities (isolates). Lastly, the number of private organizations taking part in the rescue activities is significantly higher than the number of non-private bodies.

Network 5

```r
plot.emon.network(emon[[5]], "Sponsorship", names(emon)[5], "Figure 1e.")
```

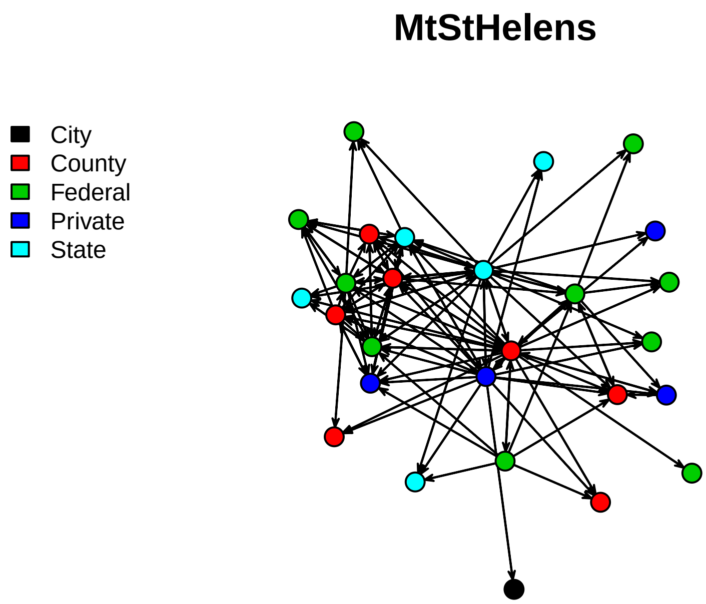

Figure 1e.

Observations

For MtStHelens, we observe some amount of selective mixing among organizations at the Federal level. Overall, however, the organizations at different levels seem to be interacting well with each other. One exception is the only City level organization, which is not initiating interactions with any other bodies, and receives communication from just one private organization.

Network 6

```r
plot.emon.network(emon[[6]], "Sponsorship", names(emon)[6], "Figure 1f.")
```

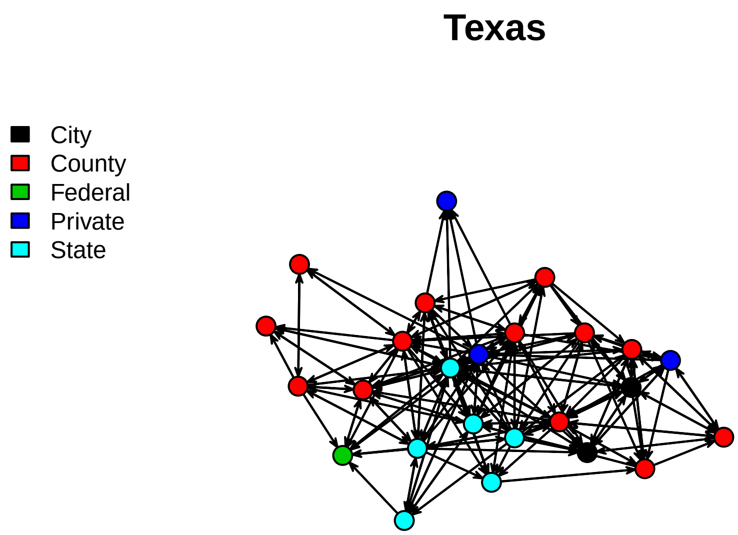

Figure 1f.

Observations

In the Texas network, there seem to be a lot of ties within organizations at the County level, and within those at the State level, indicating some sort of selective mixing. This could also appear to be happening in the plot since the number of organizations at the County and State level are much more than those at the City, Private or Federal level. Lastly, there doesn't seem to be any direct interaction between the Federal and State level bodies.

Network 7

```
plot.emon.network(emon[[7]], "Sponsorship", names(emon)[7], "Figure 1g.")
```

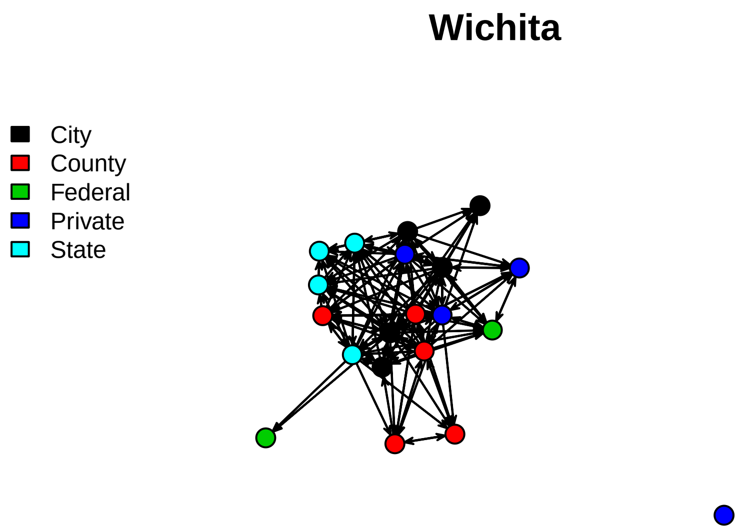

Figure 1g.

Observations

The Wichita network does not show any signs of homophily between organizations at the same level. There seem to be evenly distributed ties between organizations at various levels. One of the Private organizations, however, does not seem to be interacting with any of the other bodies. Out of the two Federal level entities, one seems to be having significantly more interactions with the rest of the network, than the other. Lastly, the two Federal bodies don't seem to have any direct interactions with each other.

(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
# Function to perform an analysis of dyadic mixing for a given EMON.
analyze.dyadic.mixing = function(network, net.name, sub.title) {
  mms = mixingmatrix(network, "Sponsorship")
  cat("Mixing Matrix:\n")
  print(mms)
  # Extract dimensions of mixing matrix
  mms.dim = dim(mms$matrix)
```

```r
num.rows = mms.dim[1]
num.cols = mms.dim[2]
# Compute grand total of all entries
matrix.sum = sum(mms$matrix)
# Matrix to store expected tie values
block.dens = array(dim = c(num.rows,num.cols))
# Matrix to store z-scores for mixing
zmm_attribute = array(dim = c(num.rows,num.cols))
# Set row and column names as in the mixing matrix
row.names(zmm_attribute) = row.names(block.dens) = row.names(mms$matrix)
colnames(zmm_attribute) = colnames(block.dens) = colnames(mms$matrix)
for (i in 1:num.rows) {
  for (j in 1: num.cols) {
    # Compute possible ties
    poss.ties = 0
    # Count number of vertices with Sponsorship same as current row name
    vertex.count = sum(network %v% "Sponsorship" == row.names(mms$matrix)[i])
    if (i == j) {
      # Possible ties within organizations at the same level
      poss.ties = vertex.count * (vertex.count - 1)
    } else {
      # Count number of vertices with Sponsorship same as current column name
      vertex.count2 =  sum(network %v% "Sponsorship" == colnames(mms$matrix)[j])
      # Possible ties between organizations at different levels
      poss.ties = vertex.count * vertex.count2
    }
    # Compute block densities
    if (poss.ties == 0) {
      block.dens[i,j] = NA
    } else {
      block.dens[i,j] = mms$matrix[i,j]/poss.ties
    }
    # Compute expected ties
    exp.tie = (sum(mms$matrix[i,]) * sum(mms$matrix[,j]))/ matrix.sum

    # Compute sz-scores
    if (exp.tie == 0) {
      zmm_attribute[i,j] = NA
    } else {
      zmm_attribute[i,j] = (mms$matrix[i,j] - exp.tie)/sqrt(exp.tie)
    }
  }
}
cat("\nBlock Densities:\n")
print(block.dens)
cat("\nMarginal Z-Scores:\n")
print(zmm_attribute)
zmm_attribute[is.na(zmm_attribute)] = 0
# Plot the reduced form blockmodel
# Edge widths reflect the ratio of current |z| to max |z|
# Edge colors reflect the polarity of z-scores
gplot(abs(zmm_attribute),edge.col=sign(zmm_attribute)+3,
      edge.lwd = abs(zmm_attribute)/max(abs(zmm_attribute)) * 5,
```

```r
        label=rownames(zmm_attribute),boxed.lab=FALSE,diag=TRUE,
        main = net.name, sub=sub.title)
}
```

Note:

In each of the below network plots, edge colors indicate the type of mixing between two blocks (blue => strong mixing between two blocks/within one block, red => weak mixing between two blocks/within one block). Further, the edge widths indicate the extent/strength of mixing in the strong/weak direction. The edge widths are always relative to the max interaction observed in either direction, for a given network. Therefore they should be analyzed in combination with the matrix of z-scores.

Network 1

```r
analyze.dyadic.mixing(emon[[1]], names(emon)[1], "Figure 2a.")
```

```
## Mixing Matrix:
##              To
## From           City County County/City Federal Private State Total
##   City            2      2           1       0       3     1     9
##   County          4      4           2       0       3     2    15
##   County/City     2      3           0       1       4     2    12
##   Federal         2      1           0       0       0     2     5
##   Private         2      4           3       0       2     2    13
##   State           6      6           3       2       6     6    29
##   Total          18     20           9       3      18    15    83
##
## Block Densities:
##                  City    County County/City   Federal   Private     State
## City        1.0000000 0.3333333   0.5000000 0.0000000 0.3750000 0.1666667
## County      0.6666667 0.6666667   0.6666667 0.0000000 0.2500000 0.2222222
## County/City 1.0000000 1.0000000          NA 1.0000000 1.0000000 0.6666667
## Federal     1.0000000 0.3333333   0.0000000        NA 0.0000000 0.6666667
## Private     0.2500000 0.3333333   0.7500000 0.0000000 0.1666667 0.1666667
## State       1.0000000 0.6666667   1.0000000 0.6666667 0.5000000 1.0000000
##
## Marginal Z-Scores:
##                    City      County County/City    Federal    Private
## City         0.03449558 -0.11453883  0.02439206 -0.5703518  0.7502788
## County       0.41416249  0.20279176  0.29285710 -0.7363210 -0.1402808
## County/City -0.37342560  0.06376727 -1.14070365  0.8598200  0.8663474
## Federal      0.87933284 -0.18659924 -0.73632104 -0.4251152 -1.0413152
## Private     -0.48793496  0.49012491  1.33949764 -0.6854776 -0.4879350
## State       -0.11530203 -0.37373245 -0.08153085  0.9296691 -0.1153020
##                  State
## City        -0.4912442
## County      -0.4317394
## County/City -0.1145388
## Federal      1.1533781
## Private     -0.2279511
## State        0.3315560
```

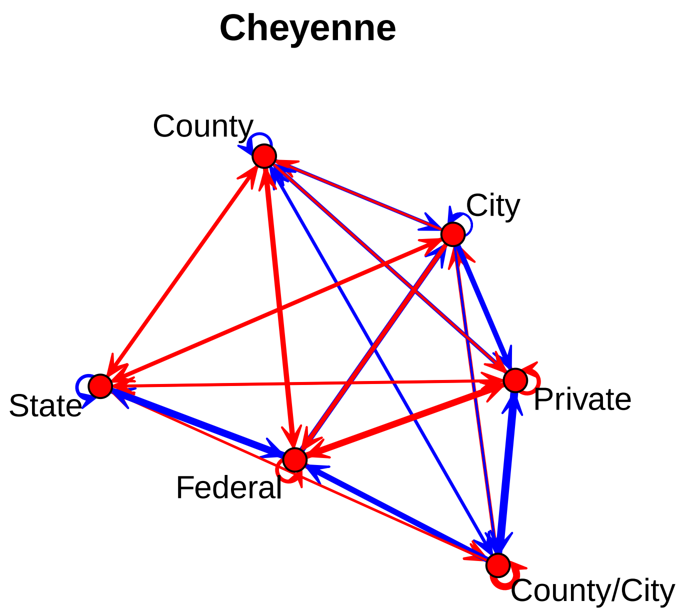

Figure 2a.

Observations

For the Cheyenne network (Figure 2a.), the Z-scores for mixing rates indicate that there does not appear any extreme observation (no z-score above our significance level of |z| >= 2) with regards to selective mixing between organizations at different sponsorship levels. The plot further shows that there are relatively higher number of interactions from 1. Federal to State level, 2. Private to County/City level entities, than one would expect in a random network with similar configuration. Lastly, the interactions from Federal to Private entities are fewer than expected.

Network 2

```r
analyze.dyadic.mixing(emon[[2]], names(emon)[2], "Figure 2b.")
```

```
## Mixing Matrix:
##          To
## From      City County Federal Private State Total
##   City      22     12       2       7    12    55
##   County    12     13       1       4     4    34
##   Federal    0      1       0       0     0     1
##   Private    8      4       0       0     1    13
##   State      8      5       0       1     1    15
##   Total     50     35       3      12    18   118
##
## Block Densities:
##              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        NA 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
##
## Marginal Z-Scores:
##               City      County    Federal    Private      State
## City    -0.2703419 -1.06797580  0.5088331  0.5948331  1.2463815
## County  -0.6340929  0.91800270  0.1458408  0.2916815 -0.5209683
## Federal -0.6509446  1.29152574 -0.1594482 -0.3188964 -0.3905667
## Private  1.0615725  0.07336725 -0.5748987 -1.1497973 -0.6980862
## State    0.6521239  0.26115161 -0.6175402 -0.4254166 -0.8515707
```

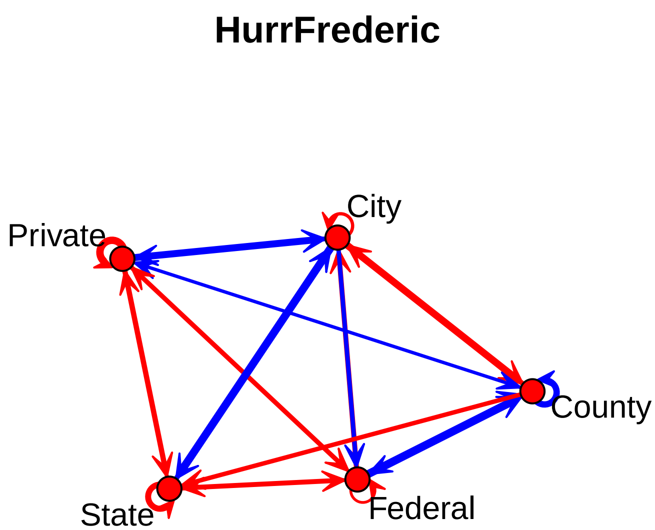

Figure 2b.

Observations

From the z-scores of the HurrFrederic network as well, we do not observe any extreme cases of mixing or it's absence, between organizations at different levels. There are a few blocks, where the interactions are relatively higher/lower than the others (Figure 2b). For instance, there are a relatively high number of interactions from Federal to County level organizations. Moreover, private organizations seem to be interacting less with each other. This latter observation could be due to the fact that private organizations are more likely to be reporting/coordinating with non-private organizations rather than among themselves.

Network 3

```
analyze.dyadic.mixing(emon[[3]], names(emon)[3], "Figure 2c.")
```

```
## Mixing Matrix:
##          To
## From      City County Federal Private State Total
##   City       2      9       3       9    10    33
##   County     3     10       4       9    13    39
##   Federal    0      3       0       2     4     9
##   Private    3     11       4       5    11    34
##   State      2     13       4       7     7    33
##   Total     10     46      15      32    45   148
##
## Block Densities:
##              City County Federal Private     State
## City    0.1666667   0.45   0.375    0.45 0.6250000
## County  0.1500000   0.50   0.400    0.36 0.6500000
## Federal 0.0000000   0.30   0.000    0.20 0.5000000
## Private 0.1500000   0.44   0.400    0.25 0.5500000
## State   0.1250000   0.65   0.500    0.35 0.5833333
##
## Marginal Z-Scores:
##               City     County     Federal     Private       State
## City    -0.1538477 -0.3924155 -0.18842422  0.69814602 -0.01066537
## County   0.2247660 -0.6093791  0.02378972  0.19545246  0.33160211
## Federal -0.7798129  0.1211965 -0.95507181  0.03874921  0.76380630
## Private  0.4636209  0.1330242  0.29846809 -0.86722929  0.20594417
## State   -0.1538477  0.8565630  0.35837547 -0.05059029 -0.95775021
```

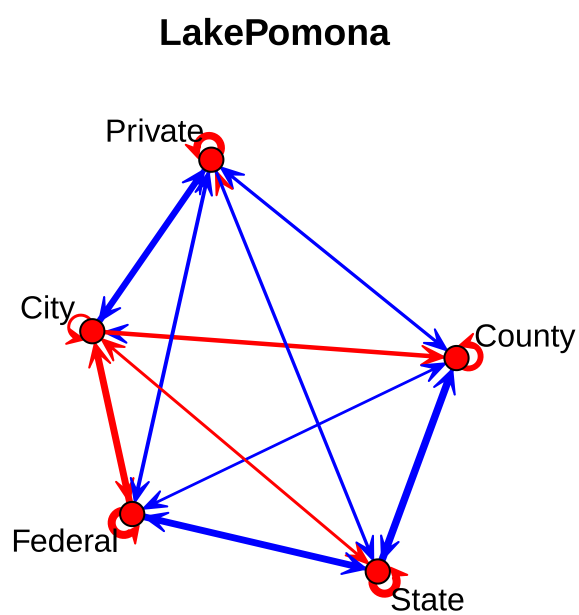

Figure 2c.

Observations

The z-scores for the LakePomona rescue network indicate that interactions between organizations at different levels are well distributed, and there is no z-score higher than 1. Relative comparisons can be done based on Figure 2c. There are a higher number of interaction ties from State to County and Federal to State level entities. There seem to be lower level of interactions from Federal to City level bodies. Lastly, we observe that Federal level entities do not interact with each other.

Network 4

```
analyze.dyadic.mixing(emon[[4]], names(emon)[4], "Figure 2d.")
```

```
## Mixing Matrix:
##          To
## From      County Federal Private State Total
##   County       0       0       4     2     6
##   Federal      0       0       0     0     0
##   Private      4       0      14     2    20
##   State        2       0       5     0     7
##   Total        6       0      23     4    33
##
## Block Densities:
##            County Federal   Private     State
## County  0.0000000       0 0.2857143 0.3333333
## Federal 0.0000000      NA 0.0000000 0.0000000
## Private 0.2857143       0 0.3333333 0.0952381
## State   0.3333333       0 0.2380952 0.0000000
##
## Marginal Z-Scores:
##             County Federal     Private      State
## County  -1.0444659      NA -0.08891084  1.4924050
## Federal         NA      NA          NA         NA
## Private  0.1906925      NA  0.01623283 -0.2724746
## State    0.6446584      NA  0.05487696 -0.9211324
```

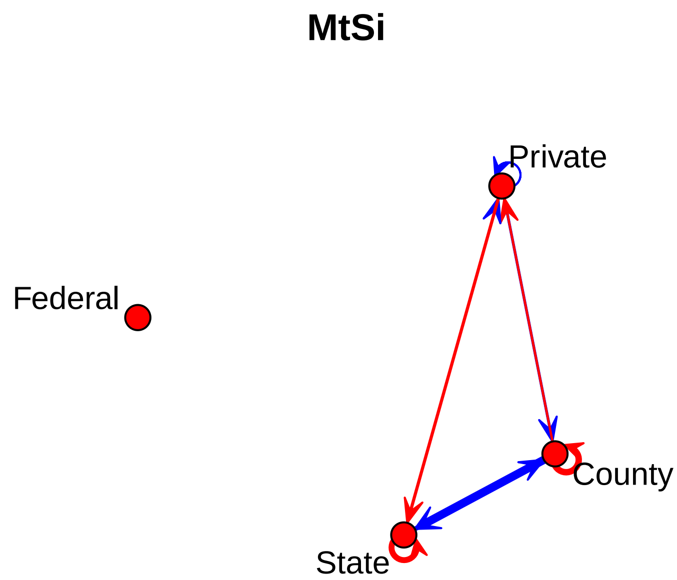

Figure 2d.

Observations

In the MtSi network, we observe that there are no interactions taking place between many of the entities. The z-scores and the plot in figure 2d. indicate that there isn't any strong selective mixing taking place as well. While looking at Figure 1d from part a) of our problem, we felt there could be homophily between private organizations. It seems however that this was only because the number of private organizations were more. From the plot and z-scores, we also see that there are relatively higher number of interactions from County to State level bodies, and no interactions reported from/to Federal level entities.

Network 5

```
analyze.dyadic.mixing(emon[[5]], names(emon)[5], "Figure 2e.")
```

```
## Mixing Matrix:
##          To
## From      City County Federal Private State Total
##   City       0      0       0       0     0     0
##   County     0     11      15       5     8    39
##   Federal    0     13       9       4     9    35
##   Private    1     11       4       2     4    22
##   State      0      6      12       5     4    27
##   Total      1     41      40      16    25   123
##
## Block Densities:
##         City    County   Federal   Private     State
## City      NA 0.0000000 0.0000000 0.0000000 0.0000000
```

```
## County  0.00 0.2619048 0.2142857 0.1785714 0.2285714
## Federal 0.00 0.1857143 0.1000000 0.1000000 0.1800000
## Private 0.25 0.3928571 0.1000000 0.1666667 0.2000000
## State   0.00 0.1714286 0.2400000 0.2500000 0.2000000
##
## Marginal Z-Scores:
##               City     County    Federal     Private       State
## City            NA         NA         NA          NA          NA
## County  -0.5630925 -0.5547002  0.6506239 -0.03248611  0.02598888
## Federal -0.5334349  0.3903600 -0.7060753 -0.25909698  0.70718233
## Private  1.9415898  1.3540064 -1.1793368 -0.50942702 -0.22299447
## State   -0.4685213 -1.0000000  1.0865026  0.79388329 -0.63510663
```

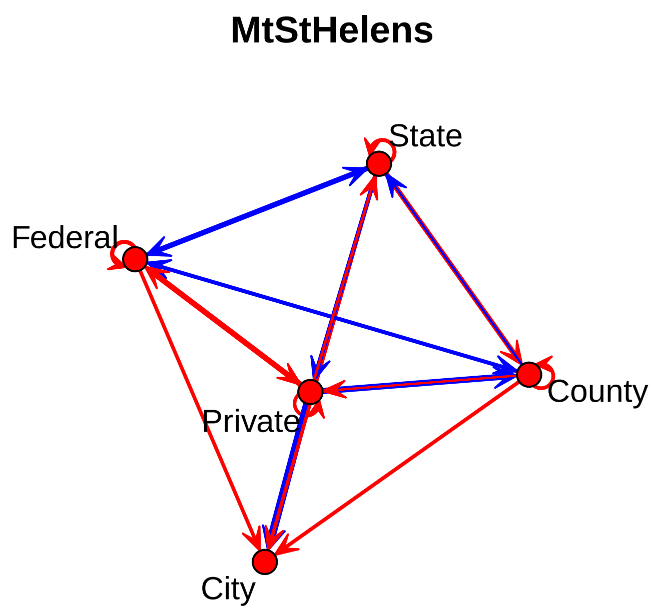

Figure 2e.

Observations

In the MtStHelens network (Figure 2e.), we see that overall there are no extremely high z-score values, with the exception of interactions from Private to City level organizations. Recalling from part a) of our problem (Figure 1e.), we realize that this exception occurs since the only city level body in the network receives communications from just one organization, which happens to be a private body. This suggests some sort of selecting mixing for the pair. Other than this, we do not observe any cases of high mixing/isolation for any of the blocks, suggesting good collaboration between the organizations.

Network 6

```
analyze.dyadic.mixing(emon[[6]], names(emon)[6], "Figure 2f.")
```

```
## Mixing Matrix:
##          To
## From      City County Federal Private State Total
##   City       2      6       0       2     2    12
##   County     9     47       3       8    27    94
##   Federal    0      1       0       0     1     2
##   Private    4      9       1       2     4    20
##   State      6     26       4       5    17    58
##   Total     21     89       8      17    51   186
##
## Block Densities:
##              City     County   Federal   Private     State
## City    1.0000000 0.23076923 0.0000000 0.3333333 0.1666667
## County  0.3461538 0.30128205 0.2307692 0.2051282 0.3461538
## Federal 0.0000000 0.07692308        NA 0.0000000 0.1666667
## Private 0.6666667 0.23076923 0.3333333 0.3333333 0.2222222
## State   0.5000000 0.33333333 0.6666667 0.2777778 0.5666667
##
## Marginal Z-Scores:
##              City      County    Federal    Private      State
## City     0.554274  0.10769589 -0.7184212  0.8624575 -0.7113430
## County  -0.495098  0.30142026 -0.5187240 -0.2017659  0.2414512
## Federal -0.475191  0.04396666 -0.2932942 -0.4275461  0.6098496
## Private  1.159215 -0.18422111  0.1507151  0.1272489 -0.6336545
## State   -0.214299 -0.33269948  0.9531090 -0.1307655  0.2750265
```

Texas

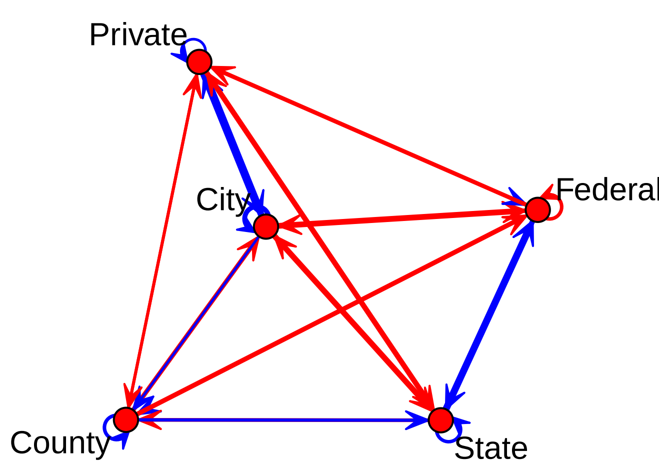

Figure 2f.

Observations

Looking at the z-scores and the plot in Figure 2f., we see that our deduction from part a) Figure 1f, wasn't entirely accurate for the Texas network. The z-scores indicate that there is not signficantly extreme amount of interaction within County or State level organizations. Also, the block densities matrix confirms our deduction that there are no direct interactions taking place between the only Federal level body and the City level organizations. Lastly, the proportion of interactions between Private and City level entities seems to be higher than that for other pairs. Overall, however, there does not appear to be a strong case for selective mixing in the network.

Network 7

```
analyze.dyadic.mixing(emon[[7]], names(emon)[7], "Figure 2g.")
```

```
## Mixing Matrix:
##          To
## From      City County Federal Private State Total
##   City      13      6       5       8    12    44
##   County    12     15       2       6    10    45
##   Federal    1      0       0       2     0     3
##   Private   11      7       3       5     6    32
##   State      7      5       2       3     8    25
##   Total     44     33      12      24    36   149
##
## Block Densities:
##         City County Federal   Private     State
## City    0.65   0.24   0.500 0.4000000 0.6000000
## County  0.48   0.75   0.200 0.3000000 0.5000000
## Federal 0.10   0.00   0.000 0.2500000 0.0000000
## Private 0.55   0.35   0.375 0.4166667 0.3750000
## State   0.35   0.25   0.250 0.1875000 0.6666667
##
## Marginal Z-Scores:
##                 City      County      Federal     Private      State
## City     0.001861891 -1.19965883  0.773659100  0.34285763  0.4199131
## County  -0.353488671  1.59442796 -0.853149915 -0.46366886 -0.2646017
## Federal  0.121218477 -0.81512547 -0.491539152  2.18197141 -0.8513708
## Private  0.504333153 -0.03277316  0.263379399 -0.06799142 -0.6227308
## State   -0.140794399 -0.22817596 -0.009459675 -0.51170855  0.7973857
```

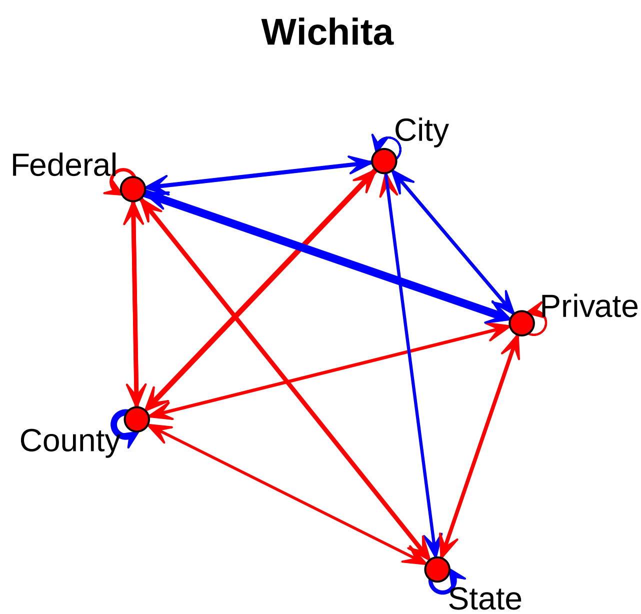

Figure 2g.

Observations

Looking at the z-scores and the plot in Figure 2g, the interactions from Federal to Private bodies stand out, with a z-score value greater than our significance level of 2, and consequently a thick edge between the two in the blockmodel plot. This is suggestive of selective mixing between the two types of bodies. We also observe that there is no interaction between the two Federal bodies (as noted in part a., Figure 1g). Lastly, there are strong interactions among County level organizations (z-score 1.59). Overall, this network does not show a pattern of selective mixing, apart from the two instances highlighted above.

(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?

The overall pattern of communication in the SAR EMONs can be described as follows:

- There appears to be only a few instances of selective mixing in the networks, and mostly, the interactions between organizations at different levels or within organizations at the same level of sponsorship do not show any extreme trends.
- This also suggests that there isn't an overall hierarchy in terms of the interactions/reporting structures between organizations at different levels of sponsorship.
- We also observe that whereever there are multiple organizations at the Federal level, there is no or very little direct interaction between them.

- In terms of interaction between Federal and Private entities, there are contrasting observations. The Cheyenne network shows no interaction between these two types of entities, whereas there is strong interaction between the two for the Wichita network (suggestive of selective mixing).
- Moreover, there is almost always a good amount of interaction between bodies at the City and Private level of sponsorship, across the seven networks.
- Lastly, Federal bodies show stronger interaction with only one out of City, County or State level bodies in all the networks.