Main Code

Margaret Perry

November 18, 2018

1. Data

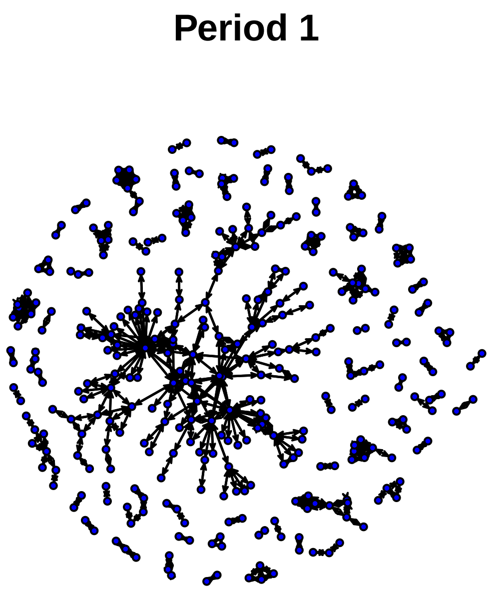
2. creating adjency matrix for the three dataset

3. examing missing data

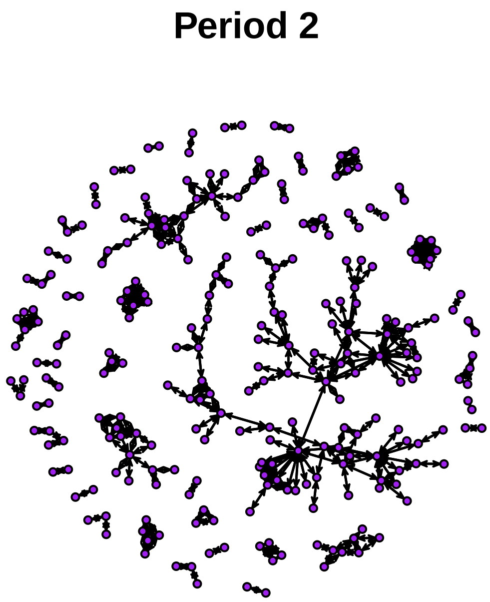
Here we can compare the number of missing data with the number of total data

4. Visualizing

```
gplot(adj.matrix.1,main="Period 1", vertex.col="Blue")
```




```
gplot(adj.matrix.2,main="Period 2", vertex.col="Purple")
```




```
gplot(adj.matrix.3,main="Period 3", vertex.col = "Orange")
```

5.1 Method: Bernoulli Blockmodel

In order to understand how people are connected, we applied the Bernoulli blockmodel to the edge connection among the three time periods.By doing so, we can see the estimate number of groups within each time period has. We will also be able to see the probability each individaul stay in each of those groups.

By assigning people to the group they have more than 50% chance to be in, we created new matrices– "test1"" for period 1 and "test3"" for period 3. Those two matrices can tell us whether or not an individual is in certain group. In order to use the group-membership information alone with other variables, we mutated a new column to our original datasets which include information about node characteristics. This new column, called "group", use number to indicate which group each individual belongs to. We can then use this information to understand what makes an individual in one group but not another.

Note that we tried to run the estimate group function for all three time periods. However, it only works for the first and the third period. We changed the min amd max number of group inside the Bernoulli model and re-examine the adjecency matrxi for period 2, but it still does not work. So in the following part we will just focus on period 1 & 3.

5.2 Method: Add group variable to dataset "daat1", named "data1.2"

Based on the Bernoulli Blockmodel, we know that period one has estimately 5 groups, and period 3 has estimately 7 groups. The above code are used to mutate the group-membership information to the original datasets.

5.3 Method: Multinomial Logistic Regression

```r
require(nnet)
```

```
## Loading required package: nnet
```

```r
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```r
require(reshape2)
```

```
## Loading required package: reshape2
```

```r
#install.packages("AER")
library(AER)
```

```
## Loading required package: car

## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##     recode

## Loading required package: lmtest

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Loading required package: sandwich

## Loading required package: survival
```

```r
#for period 1:
test1.1 <- multinom(group ~ Gender + Marital.Status+University, data = data1.1)
```

```
## # weights:  25 (16 variable)
## initial  value 537.552263
## iter  10 value 386.636672
```

```
## iter  20 value 313.928768
## iter  30 value 313.697394
## final  value 313.687400
## converged
```

```r
coeftest(test1.1)
```

```
##
## z test of coefficients:
##
##                       Estimate  Std. Error     z value Pr(>|z|)
## 2:(Intercept)     2.6415e+00  5.0969e-11  5.1825e+10  < 2e-16 ***
## 2:Gender          7.4239e+00  7.4941e-13  9.9063e+12  < 2e-16 ***
## 2:Marital.Status  8.8173e-06  5.0968e-06  1.7300e+00  0.08364 .
## 2:University     -1.8937e+00  1.0149e-11 -1.8660e+11  < 2e-16 ***
## 3:(Intercept)     9.9657e-01  6.2894e-11  1.5845e+10  < 2e-16 ***
## 3:Gender         -2.8555e+00  9.5061e-17 -3.0039e+16  < 2e-16 ***
## 3:Marital.Status -3.6227e-06  6.2893e-06 -5.7600e-01  0.56461
## 3:University     -1.8282e+01  9.4416e-20 -1.9363e+20  < 2e-16 ***
## 4:(Intercept)     1.0009e+00  6.7209e-11  1.4893e+10  < 2e-16 ***
## 4:Gender         -2.9543e+00  1.1140e-16 -2.6520e+16  < 2e-16 ***
## 4:Marital.Status -7.1801e-06  6.7207e-06 -1.0684e+00  0.28536
## 4:University     -1.6248e+00  1.4436e-12 -1.1255e+12  < 2e-16 ***
## 5:(Intercept)     1.1370e+00  5.5462e-11  2.0500e+10  < 2e-16 ***
## 5:Gender          6.2659e+00  7.5946e-13  8.2505e+12  < 2e-16 ***
## 5:Marital.Status  4.1546e-06  5.5461e-06  7.4910e-01  0.45379
## 5:University     -2.0087e+00  1.8206e-12 -1.1034e+12  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#for period 3:
test3.1 <- multinom(group ~ Gender + Marital.Status+University, data = data3.1)
```

```
## # weights:  35 (24 variable)
## initial  value 1023.548738
## iter  10 value 798.434969
## iter  20 value 733.685122
## iter  30 value 733.167782
## iter  40 value 733.140054
## final  value 733.139849
## converged
```

```r
coeftest(test3.1)
```

```
##
## z test of coefficients:
##
##                       Estimate  Std. Error     z value  Pr(>|z|)
## 2:(Intercept)    -3.1786e+00  4.7989e-11 -6.6235e+10 < 2.2e-16 ***
## 2:Gender          1.7434e+00  9.1924e-12  1.8965e+11 < 2.2e-16 ***
## 2:Marital.Status  7.1972e-06  4.7989e-06  1.4998e+00  0.133676
## 2:University      2.8847e+00  9.6676e-12  2.9838e+11 < 2.2e-16 ***
## 3:(Intercept)     3.6574e-01  1.7740e-11  2.0617e+10 < 2.2e-16 ***
## 3:Gender         -3.6179e-01  3.9335e-12 -9.1976e+10 < 2.2e-16 ***
## 3:Marital.Status  8.4839e-06  1.7739e-06  4.7825e+00 1.731e-06 ***
## 3:University      2.2845e+00  7.9403e-12  2.8771e+11 < 2.2e-16 ***
```

```
## 4:(Intercept)    -2.6664e+00  2.7276e-11 -9.7754e+10 < 2.2e-16 ***
## 4:Gender          8.6106e-02  7.8024e-13  1.1036e+11 < 2.2e-16 ***
## 4:Marital.Status  1.9567e-05  2.7276e-06  7.1735e+00 7.308e-13 ***
## 4:University     -1.0781e+01  5.9213e-18 -1.8207e+18 < 2.2e-16 ***
## 5:(Intercept)    -2.0675e+00  5.9868e-11 -3.4535e+10 < 2.2e-16 ***
## 5:Gender         -1.7544e+01  6.0062e-20 -2.9210e+20 < 2.2e-16 ***
## 5:Marital.Status -6.1893e-06  5.9867e-06 -1.0338e+00  0.301213
## 5:University      2.5201e+00  6.0193e-12  4.1867e+11 < 2.2e-16 ***
## 6:(Intercept)    -1.6794e+00  2.6082e-11 -6.4390e+10 < 2.2e-16 ***
## 6:Gender          8.1128e-01  1.6650e-12  4.8726e+11 < 2.2e-16 ***
## 6:Marital.Status  1.0507e-05  2.6082e-06  4.0285e+00 5.613e-05 ***
## 6:University      8.9928e-01  7.7036e-13  1.1674e+12 < 2.2e-16 ***
## 7:(Intercept)    -8.5987e-01  2.2800e-11 -3.7715e+10 < 2.2e-16 ***
## 7:Gender          2.9115e-01  1.9757e-12  1.4736e+11 < 2.2e-16 ***
## 7:Marital.Status  6.4292e-06  2.2799e-06  2.8199e+00  0.004803 **
## 7:University      2.4213e+00  2.7526e-12  8.7967e+11 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the filtered data, we then ran the multinomial regression model on period 1 and period 3 data. The goal is to understand what variable(s) determain which group a certain belongs to. We choose three major variable: whether or not went to university, marrital status, and gedner. The hypothesis is that people are more likely to be with people with similar background (married ppl are more likely to befriend with married ppl, etc.).

In period one, we saw that marital status among the five groups does not have statistically significant influence (p-value >0.05 for group 2-4, indicating that marrital status variable does not have significant effect on distinguishing them from group 1). The other two variable, "University" and "Gender", however, has great influence. We can conclude that in period one, gender and education level contribute significantly on who PIRA members befriend with.

In Period two, we saw a different trend. Marriage status has significant influence on the chance of being assigned into certain for most of the time; only in group 2 this variable does not has a significant p-value. The other two variables work efficiently for all seven groups. Thus we can conclude that in most cases at period 2, PIRA members' friend choices is heavily influenced by their marriage status, gender, and education background.

5.4 Method: ERGM

6. Discussion

We can then run the regression with missing data…