Final Report

Margaret Perry, Stella Li, Yuqing Geng

12/21/2018

Abstract:

The Provisional Irish Republican Army (PIRA) was a terrorist group that was formed after the Irish revolution. We are examining and analyzing the descriptive capabilities of homophily attributes, namely marital status and other mundane node attributes, on grouping within the PIRA. To answer our questions, we used a combination of three different models: exponential random graph model (ERGM), the stochastic block model, and multinomial logistic regression. First we used ERGM to examine the importance of marital status in the formation of the network and found that for all the time periods we examined, save for period 6, that marital status was a significant factor in whether two nodes were connected. Next, we used the combination of the Bernoulli block model and multinomial logistic regression to find the groups within each period then determine what factors help to identify group membership. With those three models, we can explore the formation of social network groups among PIRA members.

Introduction:

The use of social network analysis on terrorist organizations is not particularly new. Networks of this nature are often referred to as darknets due to the risk of inaccurate and possible biases. However, even the newest period of our data was collected 20 years prior to our analysis, and none of the information provides identifying; thus there is very little risk involved in our research. Rather than focusing on activities and other commonly explored reasons of social group forming, we decide to explore the impact of marital status and other human-central elements. In the past years, terrorists are usually depicted dehumanized, and thus not many researches have explored the influence of human-centuraled elements among terrorists' social behavior. WIth our research, such a blank can be filled. Using data about PIRA members from 1970 to 1998, the period which they were most active, we can have a better understanding of terrorists' social behaviors. PIRA is a terrorist organization that formed in Northern Ireland in hope of freeing the region and reunify Ireland as a whole. However the conflict was not confine to this one country, according to the New York Times, "The I.R.A. drew significant support from groups as disparate as Irish-Americans in the United States and the Libyan dictator Muammar el-Qaddafi, who supplied significant amounts of arms and powerful explosives."(Cowell). By examining the mundane attributes, focusing on the marital status of an individual, we will discover how outside factors lead the formation of edges for groups, and eventually find correlation between this and possible other attributes of interest. Additionally, we are interested in mapping out the the ties between members of PIRA, estimate how many social groups exist among those PIRA members, and evaluate what node characteristics contribute to the group-membership. In the first part, we are going to use ERGM node-matching technique to detect common background on PIRA membership; for the second part, we are going to use Bernoulli Block Model to detect potential social groups, and then use Multinomial Regression to explore variables which contribute to group membership. This will give us an overall sense what groups that model believes exist, visualizing this will give us insight to the structure of the group and large and help us contextualize our other findings.

Data:

The data that we are using is from the time period 1970 to 1998, which is often referred to as the Troubles. This was a time of heightened terrorist activity in Northern Ireland in attempts to force its independence.

They see themselves as a successor to the original Irish Republican Army, which was the revolutionary group that lead to the freedom of Ireland in 1920.

The data for this experiment was found in the datasets area of the UCINET Software site. According to the site, the data was collected by the International Center for the Study of Terrorism, Pennsylvania State University. The ties are composed of 4 types of relationships involvement in a PIRA activity together, friends before joining PIRA movement, blood relatives, and related through marriage. The network has "binary and symmetric relations between members". The data covers the members of the Provisional Irish Republican Army from 1970 to 1998, which was the height of the troubles back in Ireland and England. For the members, data was collected on information regarding gender, age, marital status, recruiting age, whether or not they attended university, then role and task-related characteristics, lastly brigade membership which is in reference to the cell that the belonged to. The observations are divided in 6 different time periods across the 28 years. However the period 4 and 5 are combined into one set on the website, and so there is no clear way to distinguish them, but we will not be using this subset in our analysis so it is not our main concern.

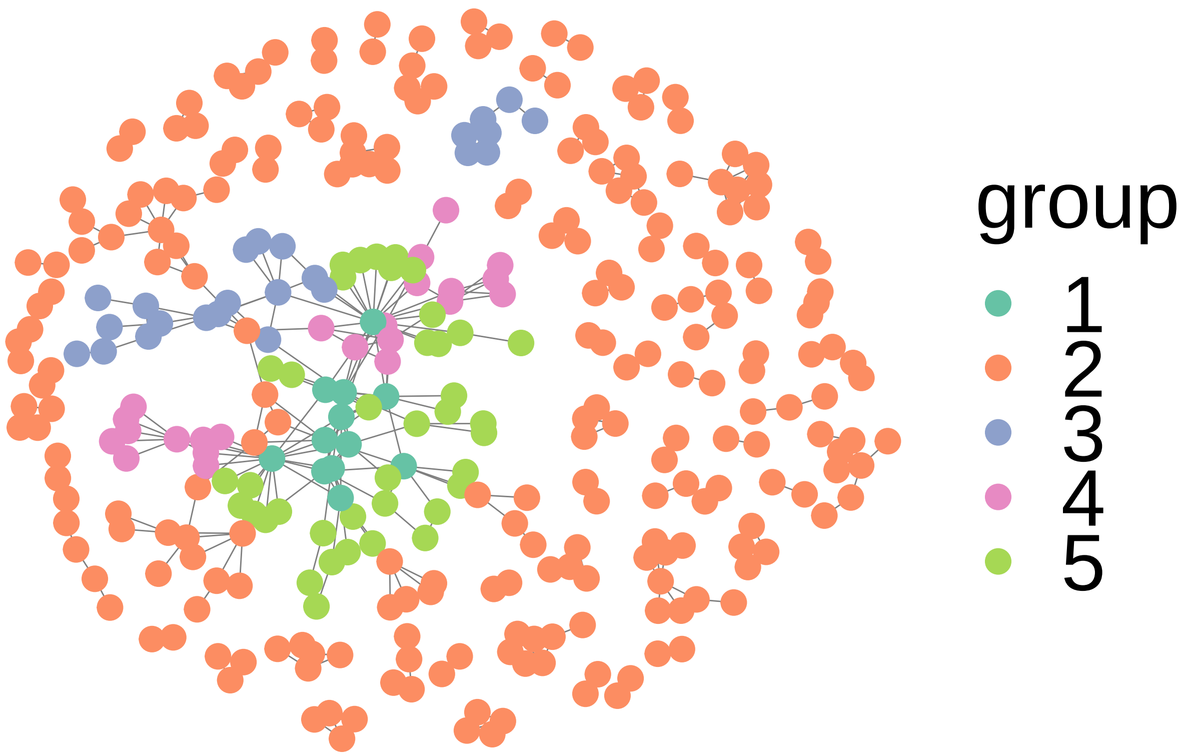
Method

Bernoulli Blockmodel

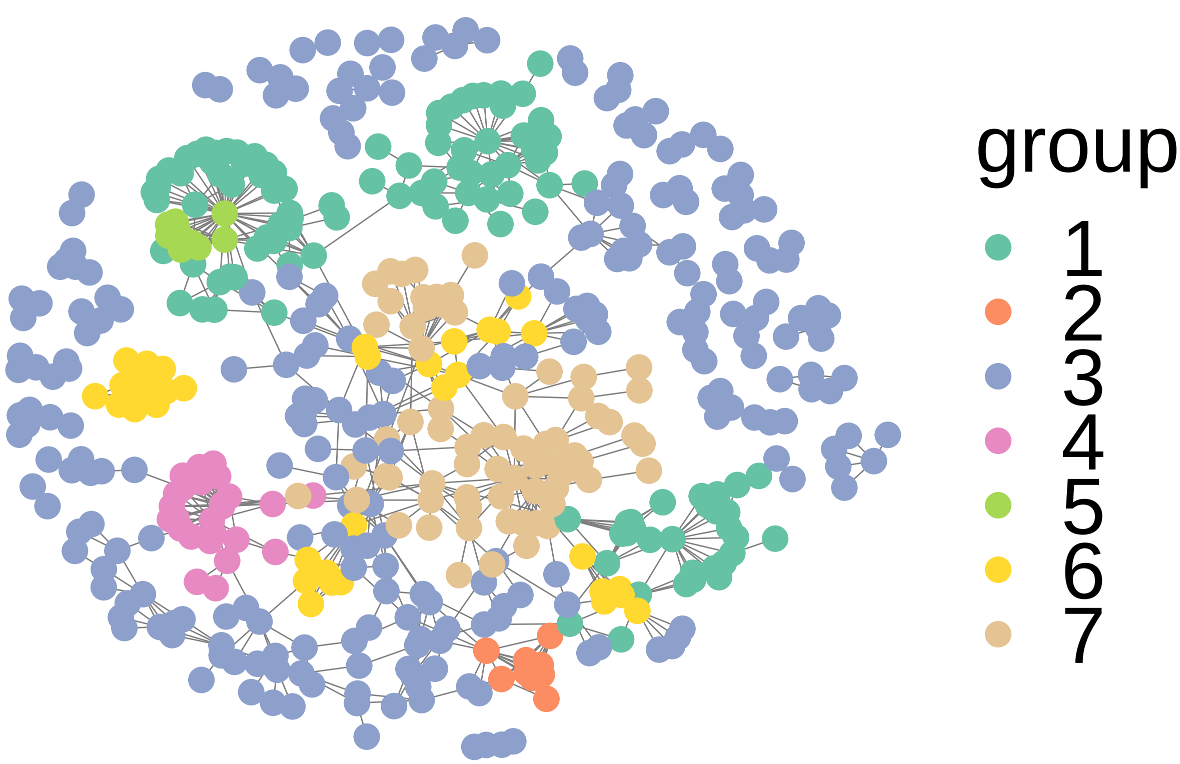
In order to understand how people are connected, we applied the Bernoulli Blockmodel to the edge connection among the three time periods. By doing so, we can see the estimated number of groups within each time period. We will also be able to see the probability each individual belong to each of these 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 (TRUE/FALSE expression). 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", we used number to indicate which group each individual belongs to. We can then use this information to understand what makes an individual in a certain group.The code are used to add the group-membership information onto the original datasets.

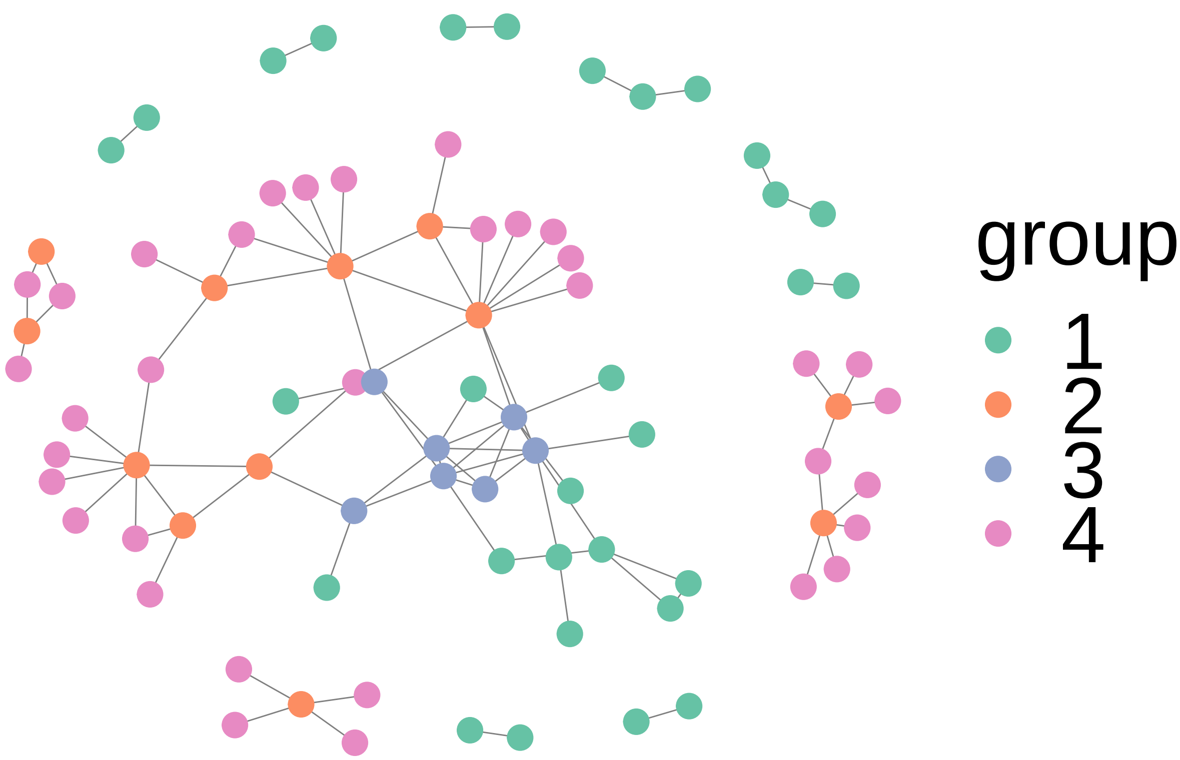
period 1



period 3



period 6



Multinomial Logit Regress

Based on the filtered data, we then ran the multinomial logistic regression model on period 1, period 3, and period 6 data (period 2 data has technical difficulties in Block model as mentioned before, thus we cannot examine the indicator variables for it). The goal is to understand what variable(s) determine which group a certain PIRA member belongs to. We chose four variables: “Gender”, “University” (whether the individual went to university), “Marital Status”, and “Recruitment Age”. The hypothesis is that people are more likely to be with others who share the similar background (married ppl are more likely to befriend with married ppl, etc.).

The Logistic Regression Model shows how likely each variable can distinguish other social groups from group 1 and predict the probabilities of the different possible outcomes of “Group” (social group membership) given “Gender”, “University”, “Marital Status”, and “Recruitment Age”. A small p-value ($p<0.05$) indicates a higher likelihood a certain indicator variable can work as an indicator variable.

Here is the result for period 1,period3, and period 6:

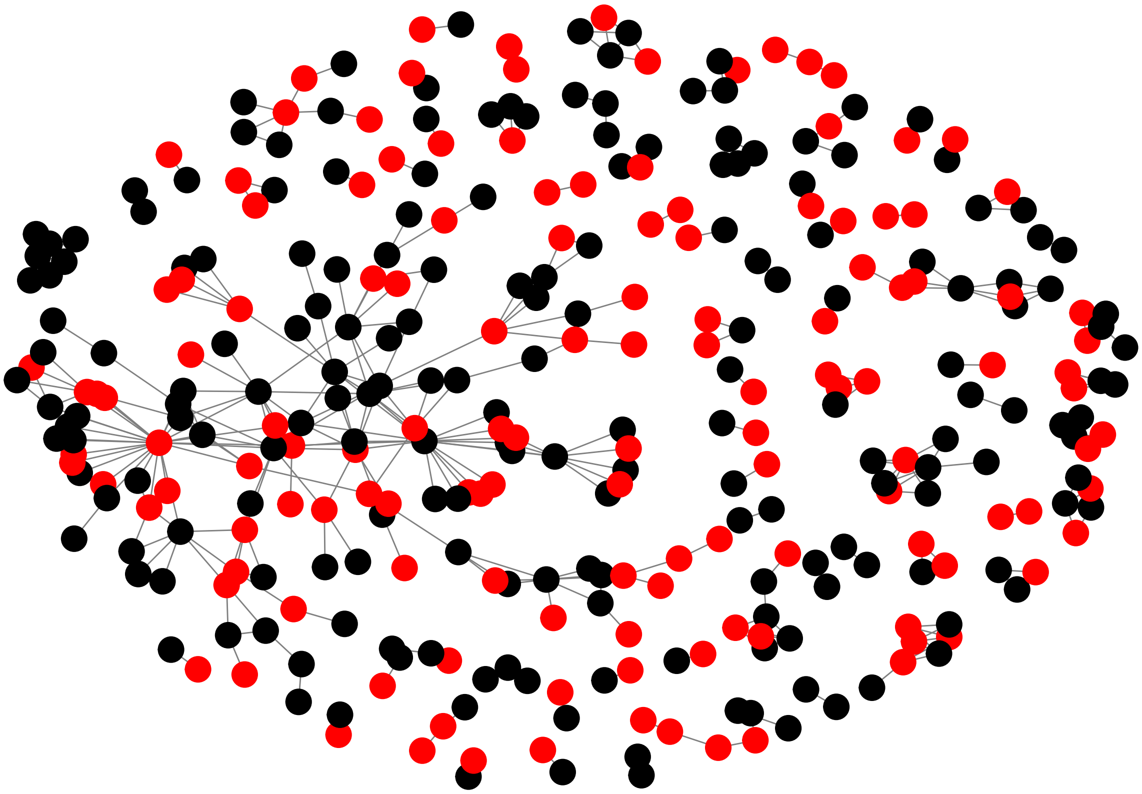
```
##
## z test of coefficients:
##
##                          Estimate  Std. Error     z value  Pr(>|z|)
## 2:(Intercept)          4.0518e+00  1.4235e+00  2.8463e+00  0.004423 **
## 2:Gender               1.3748e+01  5.5163e-01  2.4922e+01 < 2.2e-16 ***
## 2:Marital.Status      -2.5562e+00  1.1098e+00 -2.3032e+00  0.021265 *
## 2:University          -4.3806e-01  1.2242e+00 -3.5780e-01  0.720462
## 2:Age.at.Recruitment -2.5152e-04  4.7341e-02 -5.3000e-03  0.995761
## 3:(Intercept)          1.0240e+00  1.5433e+00  6.6350e-01  0.506995
## 3:Gender              -5.7087e+00  1.7969e-09 -3.1770e+09 < 2.2e-16 ***
```

```
## 3:Marital.Status     -1.6095e+00  1.2188e+00 -1.3205e+00  0.186655
## 3:University         -1.6727e+01  5.0871e-07 -3.2881e+07 < 2.2e-16 ***
## 3:Age.at.Recruitment  3.7652e-02  5.1605e-02  7.2960e-01  0.465619
## 4:(Intercept)         1.3777e+00  1.5589e+00  8.8370e-01  0.376849
## 4:Gender             -5.7831e+00  1.5160e-09 -3.8148e+09 < 2.2e-16 ***
## 4:Marital.Status     -2.7026e+00  1.2403e+00 -2.1789e+00  0.029338 *
## 4:University         -2.2619e-01  1.5647e+00 -1.4460e-01  0.885057
## 4:Age.at.Recruitment  3.8163e-02  5.3622e-02  7.1170e-01  0.476642
## 5:(Intercept)         1.4398e+00  1.5185e+00  9.4820e-01  0.343039
## 5:Gender              1.3025e+01  5.5163e-01  2.3612e+01 < 2.2e-16 ***
## 5:Marital.Status     -2.4788e+00  1.2111e+00 -2.0468e+00  0.040678 *
## 5:University         -5.2395e-01  1.5450e+00 -3.3910e-01  0.734509
## 5:Age.at.Recruitment  3.8382e-02  5.1451e-02  7.4600e-01  0.455679
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## z test of coefficients:
##
##                        Estimate  Std. Error     z value  Pr(>|z|)
## 2:(Intercept)        -3.1886e+00  2.3329e+00 -1.3668e+00 0.1716800
## 2:Gender              1.2029e+00  1.1493e+00  1.0466e+00 0.2952781
## 2:Marital.Status      2.3731e+00  1.1826e+00  2.0067e+00 0.0447824 *
## 2:University          2.3072e+00  1.6735e+00  1.3786e+00 0.1680029
## 2:Age.at.Recruitment -4.4700e-02  1.0250e-01 -4.3610e-01 0.6627636
## 3:(Intercept)        -1.6790e+00  7.5380e-01 -2.2273e+00 0.0259241 *
## 3:Gender             -7.5868e-01  6.2152e-01 -1.2207e+00 0.2222099
## 3:Marital.Status      8.9215e-01  3.7273e-01  2.3936e+00 0.0166856 *
## 3:University          8.6672e-01  1.1544e+00  7.5080e-01 0.4527875
## 3:Age.at.Recruitment  8.4160e-02  3.3179e-02  2.5365e+00 0.0111956 *
## 4:(Intercept)        -7.5936e+00  1.6990e+00 -4.4693e+00 7.846e-06 ***
## 4:Gender             -9.5543e-01  1.5730e+00 -6.0740e-01 0.5435895
## 4:Marital.Status      2.1491e+00  1.2139e+00  1.7703e+00 0.0766693 .
## 4:University         -1.3123e+01  3.7870e-06 -3.4654e+06 < 2.2e-16 ***
## 4:Age.at.Recruitment  1.6044e-01  4.9121e-02  3.2663e+00 0.0010896 **
## 5:(Intercept)        -4.9536e+00  1.2532e+00 -3.9528e+00 7.724e-05 ***
## 5:Gender             -1.4484e+01  6.2485e-07 -2.3181e+07 < 2.2e-16 ***
## 5:Marital.Status     -9.4855e-01  1.1255e+00 -8.4280e-01 0.3993649
## 5:University         -1.2052e+01  9.8946e-06 -1.2180e+06 < 2.2e-16 ***
## 5:Age.at.Recruitment  1.3826e-01  4.7873e-02  2.8881e+00 0.0038760 **
## 6:(Intercept)        -5.5393e+00  1.4421e+00 -3.8411e+00 0.0001225 ***
## 6:Gender             -5.4974e-01  9.6610e-01 -5.6900e-01 0.5693338
## 6:Marital.Status      3.6066e+00  1.0844e+00  3.3259e+00 0.0008815 ***
## 6:University          1.0264e+00  1.5688e+00  6.5430e-01 0.5129344
## 6:Age.at.Recruitment  7.9469e-02  4.5768e-02  1.7363e+00 0.0825052 .
## 7:(Intercept)        -1.3714e+00  9.8575e-01 -1.3912e+00 0.1641531
## 7:Gender             -1.1529e+00  9.2215e-01 -1.2502e+00 0.2112083
## 7:Marital.Status      1.4188e+00  4.7427e-01  2.9915e+00 0.0027760 **
## 7:University          2.2905e+00  1.2162e+00  1.8834e+00 0.0596453 .
## 7:Age.at.Recruitment  8.6460e-03  4.3942e-02  1.9680e-01 0.8440163
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## z test of coefficients:
```

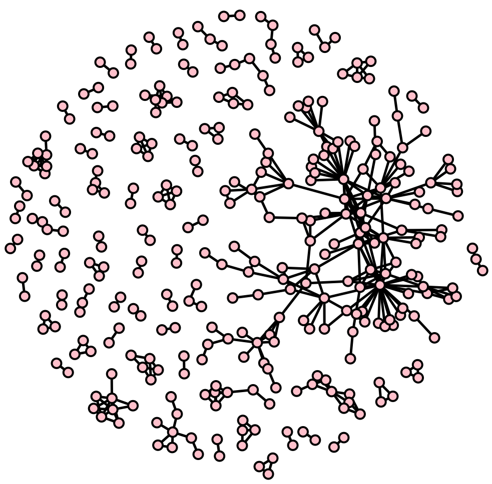
```
##
##                        Estimate  Std. Error     z value Pr(>|z|)
## 2:(Intercept)       -2.6333e-01  2.4343e+00 -1.0820e-01   0.9139
## 2:Gender            -1.6325e+01  1.8794e-07 -8.6865e+07   <2e-16 ***
## 2:Marital.Status     2.6047e-01  1.6020e+00  1.6260e-01   0.8708
## 2:University         1.1434e+01  8.9418e-01  1.2787e+01   <2e-16 ***
## 2:Age.at.Recruitment  8.7370e-03  7.7924e-02  1.1210e-01   0.9107
## 3:(Intercept)        1.0117e+00  2.3723e+00  4.2650e-01   0.6698
## 3:Gender            -1.6366e+01  1.8777e-07 -8.7160e+07   <2e-16 ***
## 3:Marital.Status    -6.3919e-01  1.4375e+00 -4.4470e-01   0.6566
## 3:University         1.1311e+01  9.0012e-01  1.2566e+01   <2e-16 ***
## 3:Age.at.Recruitment -1.2910e-02  8.1662e-02 -1.5810e-01   0.8744
## 4:(Intercept)        1.3702e+00  1.9072e+00  7.1840e-01   0.4725
## 4:Gender             2.2478e-01  1.3089e+00  1.7170e-01   0.8636
## 4:Marital.Status     4.9640e-02  1.2925e+00  3.8400e-02   0.9694
## 4:University         1.0665e+01  8.6799e-01  1.2287e+01   <2e-16 ***
## 4:Age.at.Recruitment -1.8535e-02  6.5795e-02 -2.8170e-01   0.7782
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

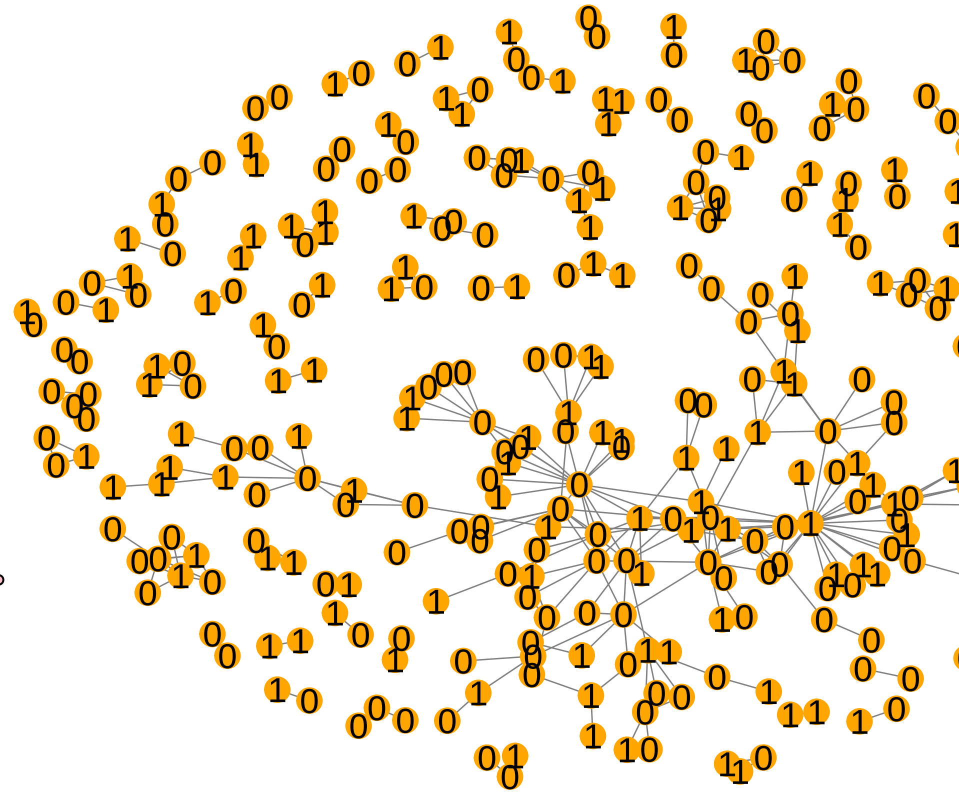
ERGM

For our exponential random graph model (ERGM) we ran periods 1, 2, 3, and 6. For each of these time periods we ran on edges, which attempts to build model similar to the original model based on the edges that exist, and nodes match, which is an argument in the code that attempts to build a similar network based on the status of the node attribute being called, for marital status. We after this we look at the node attributes for the education of an individual and the age of recruitment, but only some of the periods yield interesting results with these attributes. The main issue we ran into with many of the possible arguments for this model was that there was a noticeable lack in interptiblity for complex models of this type. We also ran in to an issue when we were first attempting to run the model because of how we formed our networks that the column names did not line up with the correct attribute which gave us wildly incorrect results the first time we ran through these models and so we had to run them a second time with the corrected networks to obtain usable results. In addition to this, for our ERGM networks we drop all observations with missing values for marital status as to avoid any underlying issues that may come of that.

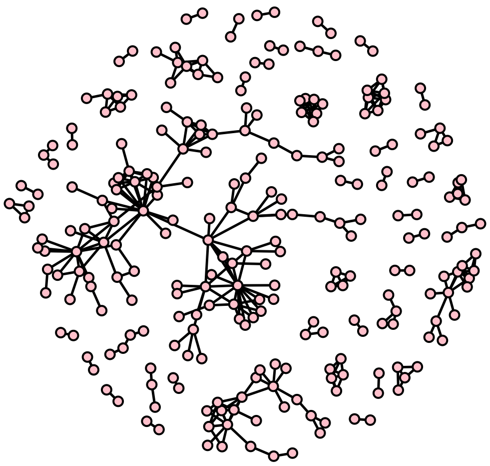


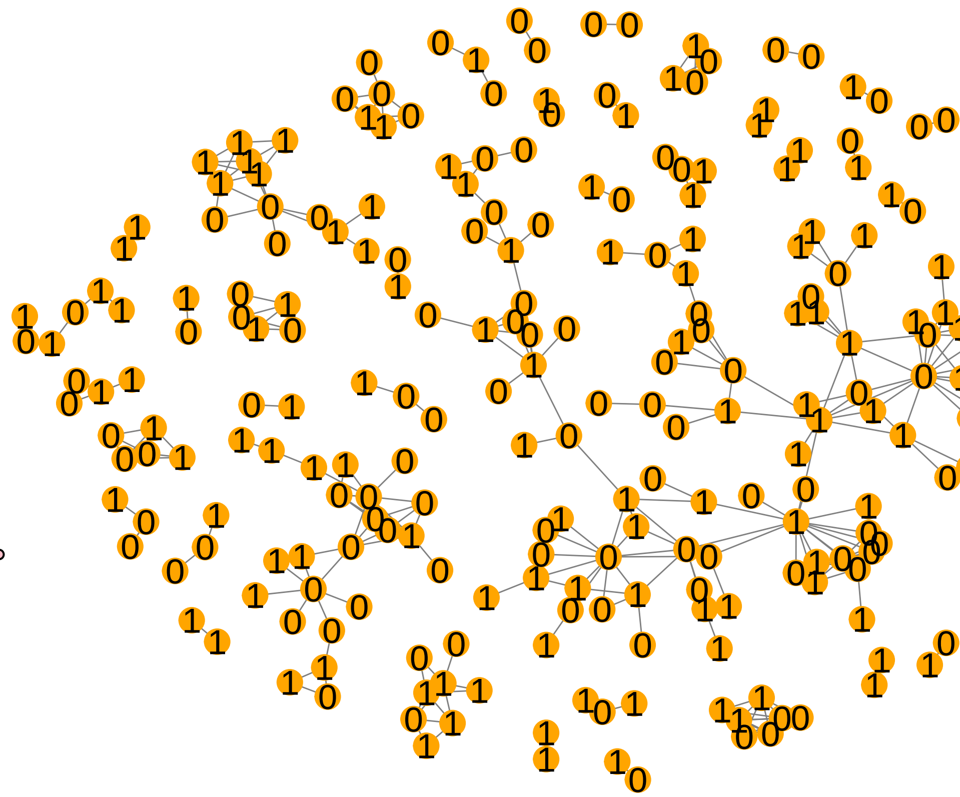
```
## Starting maximum pseudolikelihood estimation (MPLE):
## Evaluating the predictor and response matrix.
## Maximizing the pseudolikelihood.
## Finished MPLE.
## Stopping at the initial estimate.
## Evaluating log-likelihood at the estimate.
```



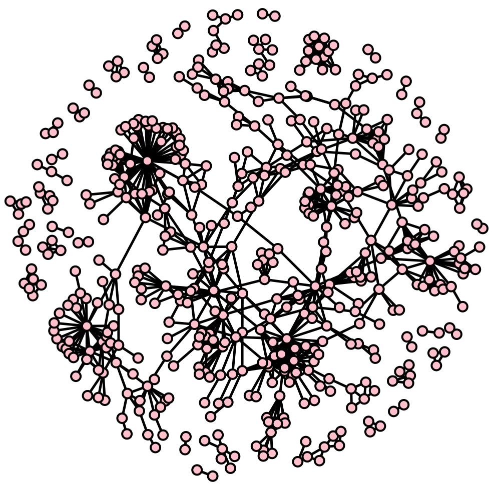


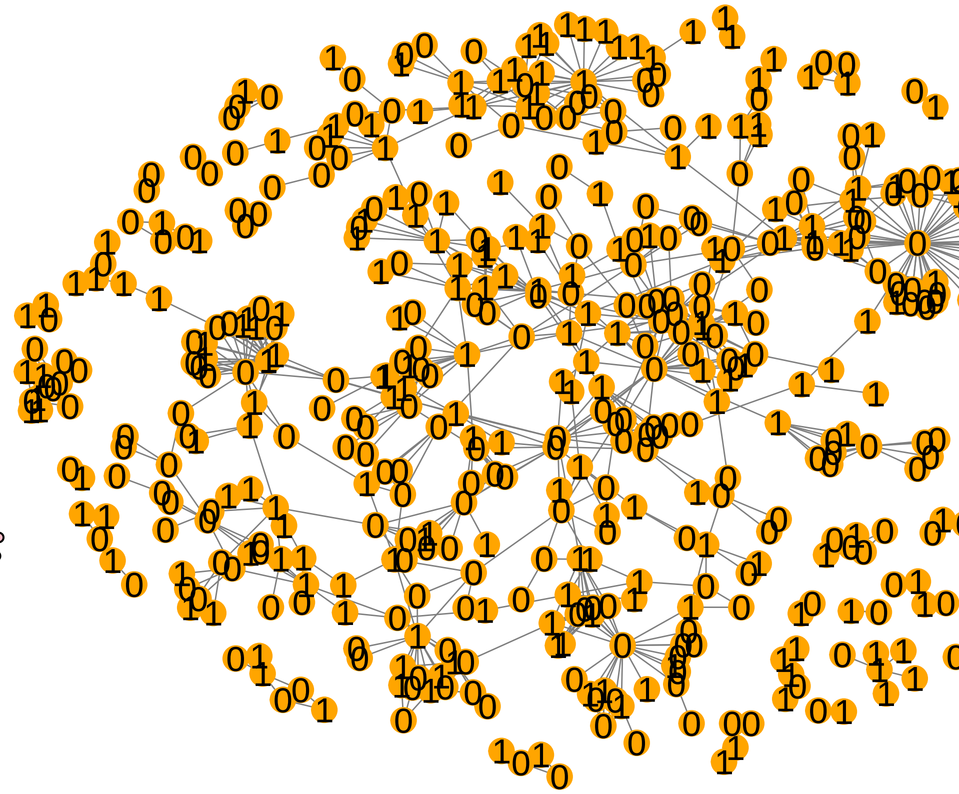
```
## Starting maximum pseudolikelihood estimation (MPLE):
## Evaluating the predictor and response matrix.
## Maximizing the pseudolikelihood.
## Finished MPLE.
## Stopping at the initial estimate.
## Evaluating log-likelihood at the estimate.
```



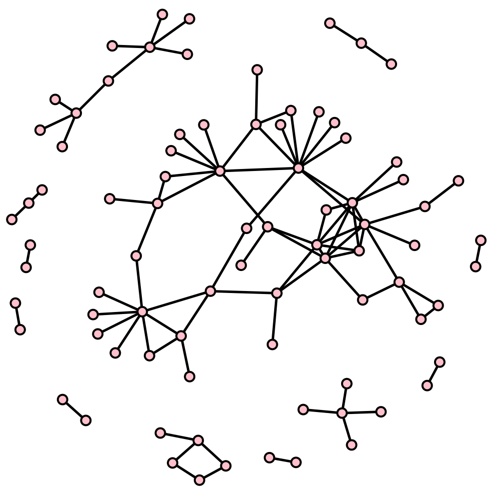


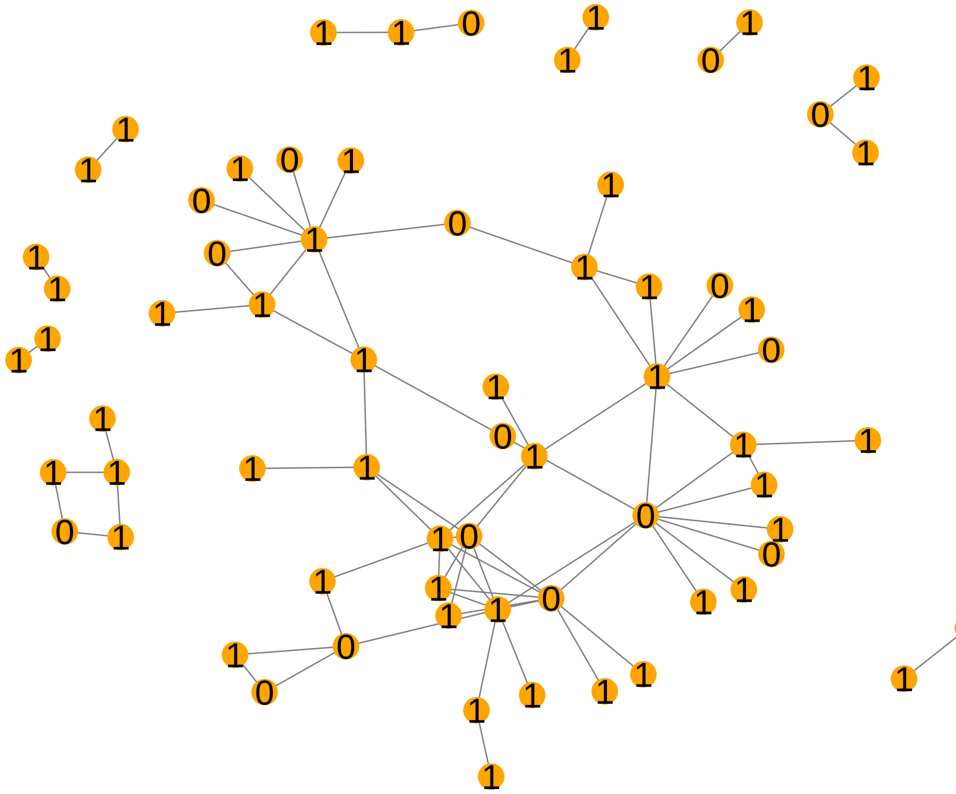
```
## Starting maximum pseudolikelihood estimation (MPLE):
## Evaluating the predictor and response matrix.
## Maximizing the pseudolikelihood.
## Finished MPLE.
## Stopping at the initial estimate.
## Evaluating log-likelihood at the estimate.
```



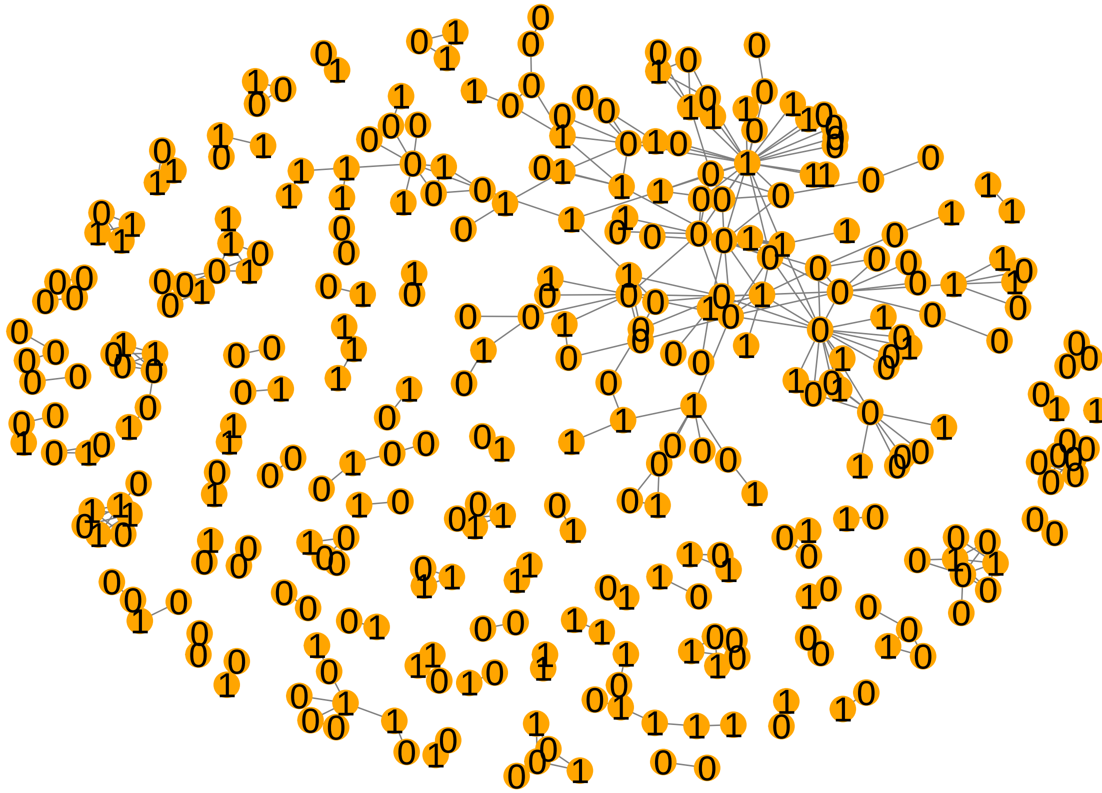


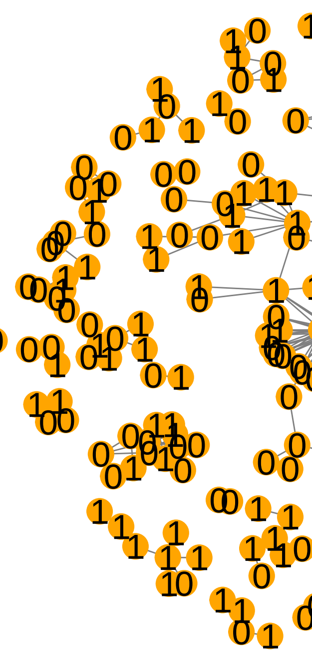
```
## Starting maximum pseudolikelihood estimation (MPLE):
## Evaluating the predictor and response matrix.
## Maximizing the pseudolikelihood.
## Finished MPLE.
## Stopping at the initial estimate.
## Evaluating log-likelihood at the estimate.
```

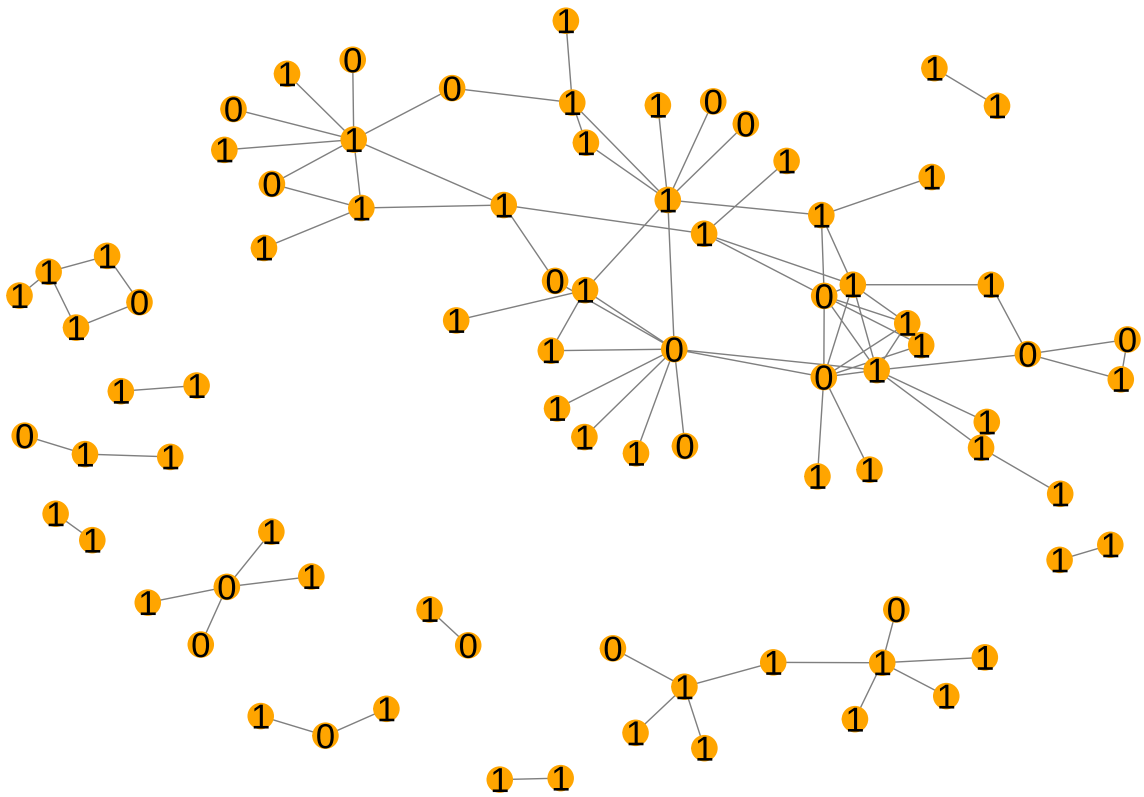




Here are the graphs for ERGM in Period 1, 2, 3, 6:







Result

Based on the Bernoulli Blockmodel, we found 5 estimated social groups in period 1, 7 in period 3, and 4 in period 6. Our definition of "being in one specific group" means the probability an individual belongs to a group is greater than 50%, but based on the Bernoulli Blockmodel's calculation, the likelihood for an individual being in the most likely group is usually greater than 90%, and there are only less than 5% probability that it belongs to other groups. So we are pretty confident that our group membership, based on calculation, is accurate.

Note that we tried to run the Bernoulli Blockmodel on period 2 too, but for unknown reasons it kept report error. We changed the minimum and maximum number of group inside the Bernoulli Blockmodel and re-examine the adjacency matrix for period 2, but it still was not work. So in the following part we will just focus on period 1, 3, and 6.

The Logistic Regression Model shows what are major indicator variables for group memberships among those three time periods.

In period 1, we saw that "Marital Status" can significantly distinguish the probability of being in groups other than group 1 (p-value <0.05 for all but group3, indicating this indicator variable works well for distinguishing most groups). "Gender" is also significant for all groups. The other two variable, "University" and "Recruitment Age", however, has less influence. They have p-value >0.05 for almost all groups, indicating that they cannot welly distinguish each group as indicator variables. We can then conclude that in period one, "Gender" and "Marital Status" contribute significantly on who PIRA members befriend with.

Period 2 cannot be accessed due to unknown technical difficulties as mentioned before. In Period 3, we saw a different trend. "Gender" is no longer significant. Instead, "Marital Status" has significant influence group membership as indicated by the p-value for each group; "University" becomes significant in two out of seven groups, and "Recruitment Age" is also more significant comparing with period 1. Thus we can conclude that in most cases at period 2, PIRA members' friend choices is heavily influenced by their "Marital Status", education background, and "Recruitment Age".

In Period 6, we saw a new trend. "Gender" and "University" become new key indicator variables, while the other two, "Recruitment Age" and "Marital Status" become insignificant. As we can see from those three time periods, with the passing of time, the key determinants of the formation of social group is changing too. But our hypothesis is partially correct: people do like to form relationship with others who share some kind of similar background.

In our ERGM models, we see that for first three of the four periods we have obtain an extremely small p-value, which indicates that marital status is a statistically significant attribute in formation of the network. This indicates to us that within this network of members of the PIRA that there is pattern for members with similar martial statuses to linked by one of the four edge types described earlier in the paper. However it should be noted that in period 6 while the p-value was close it was not low enough to be considered statistically significant.

Discussion:

In this study, we examined the general characteristic, marital status, of PIRA members. However, we are also interested in learning whether or not the marriage link between PIRA members has any contribution to the formation of social group.If we had access to the information it could be avantages to look at the types of relationship best described by marital status save for related for marriage, which would give us a valuable perspective of the inner workings of the network.

Since we are only including marital status in this model, the information revealed from the output might be limited and biased. For example, we are not sure whether two nodes connected solely with each other exclusively mean that they are married to each other or just simply mean they are connected because they share ties with married people. In the future, we would like to add in the gender characteristic, so we can have a more accurate interpretation of the model. We did not originally consider this factor because how the data describes the edges none are form through relationships formed during membership in the PIRA. However, the edge type related through marriage may encapsulate this. There is a concerned since we are unable to determine the type of edge that connects to nodes, and therefore the actual structure within the network. In the future we hope to find the original data set so that we may have access to this information.

Due to the data having already been collected and published, we assume the ethical concerns of the data have hypothetically been addressed. However, little is available on how this data was collected, therefore we have concerns we regarding the timeline of collection and the status of the individual at that point. There is a lot of concern with the validity and the ethics of collecting data from prisoners, this may have been the case with our dataset. It could have been collected retroactively from events and individuals, but this is likely not the case. When data is collected from inmates it is unlikely that the information is voluntary, and there are concerns with confidentiality. This may lead to individuals refusing to disclose more sensitive information about relatives or significant others in hopes of keeping them safe.

Besides, we have some missing data in all of our data sets, due to lack of information, we do not know whether our data is missing completely at random or not. The research can only be carried out under the missing completely at random assumption, but in the future, we would like to analyze the individuals who have certain background information missing (Eg: some have missing marital status, some have missing education background), and see whether there exists certain kind of relationship between them.

Furthermore, since we have data from multiple time period, we are also interested in seeing how new relationships are formed or old relationships are changed overtime. Since this is beyond the scope of answering our initial question, we didn't involve it into our study.

Citation

Provisional Irish Republican Army - UCINET Software." Google Sites, sites.google.com/site/ucinetsoftware/datasets/covert-networks/provisionalirishrepublicanarmy. Slavens, Kate. "Violence Returns to the Streets of Northern Ireland." Penn State University, 2017, news.psu.edu/story/268437/2013/03/13/research/violence-returns-streets-northern-ireland. Cowell, Alan. "50 Years Later, Troubles Still Cast 'Huge Shadow' Over Northern Ireland." The New York Times, The New York Times, 5 Oct. 2018, www.nytimes.com/2018/10/04/world/europe/northern-ireland-troubles.html.