

High School Moreno Data Analysis

Jingxin Huang and Douglas Tan, University of Washington Bothell

Table of Contents

Narrative Summary	1
Setup Code	1
Network Plots	2
Regular Plot & Community Detection	2
Centrality Plot.....	4
Analysis.....	4
Centrality.....	4
Conclusion.....	6

Narrative Summary

[For this project, we decided to analyze the relationships in the forms of communities using the data set created by Jacob Moreno of the friendships between boys within a certain high school. By analyzing the data set using community detection and looking at the degrees, our goal is to see if we are able to analyze this older data set as well as reach a conclusion about the relationships between these boys at this specific high school.] [The data that will used for this project is collected by Jacob L. Moreno. This data set is a network data set that contains the friendships between boys in a certain high school in Illinois, United States. Each boy was asked twice for the creation of this data set, once in the Fall of 1957 and once in the Spring of 1958. This data set combines the results from both dates. A node in this data set represents a boy and an edge in this data set represents friendship. There is a total number of 70 components in this data set.

Link to access: <https://networkrepository.com/soc-highschool-moreno.php>]

Setup Code

```
library(igraph) # Loading igraph package

##
## Attaching package: 'igraph'

## The following objects are masked from 'package:stats':
##
##     decompose, spectrum
```

```

## The following object is masked from 'package:base':
##
##      union

#reading in required data:
highschool.moreno <- read.table("highschool-moreno.txt", header = 1)
highschool <- graph_from_data_frame(highschool.moreno, directed = T)

#betweenness of graph
between <- betweenness(highschool, directed = T, normalized=F)
bet_normalized <- betweenness(highschool, directed=T, normalized=T)
bet.dat <- data.frame(between, bet_normalized)

#Edge betweenness community detection
gv <- cluster_edge_betweenness(highschool)

#Degrees
highschool.degree <- degree(highschool, mode = "total")
highschool.indegree <- degree(highschool, mode = "in")
highschool.outdegree <- degree(highschool, mode = "out")
dat <- data.frame(highschool.degree, highschool.indegree, highschool.outdegree)

```

Network Plots

Regular Plot & Community Detection

```

par(mfrow=c(1,2))
#regular plot of the moreno data
set.seed(2022)
plot(highschool, vertex.label.color = "black", vertex.color = "yellow", edge.
arrow.size=.3,
      main="Moreno Friendship Network")
#betweenness plot of the moreno data
set.seed(2022)
plot(gv,highschool, vertex.label.color="black",vertex.size=13, edge.arrow.size=.3,
      main="Community Detection")

```

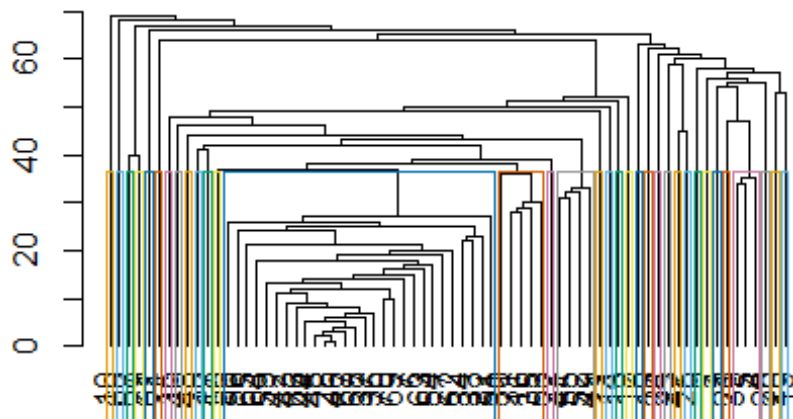
Moreno Friendship Network

Community Detection



Dendrogram

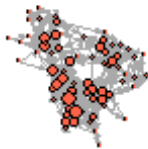
```
dendPlot(gv, mode="hclust")
```



Centrality Plot

```
par(mfrow=c(1,3))
#total degree plot
set.seed(2022)
plot(highschool, edge.arrow.size=.2, vertex.label=NA, vertex.color="tomato",
      vertex.size=highschool.degree, main="Total Degree")
#in-degree plot
set.seed(2022)
plot(highschool, edge.arrow.size=.2, vertex.label=NA, vertex.color="tomato",
      vertex.size=highschool.indegree, main="In-Degree")
#out-degree plot
set.seed(2022)
plot(highschool, edge.arrow.size=.2, vertex.label=NA, vertex.color="tomato",
      vertex.size=highschool.outdegree, main="Out-Degree")
```

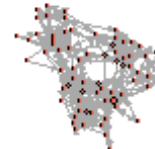
Total Degree



In-Degree



Out-Degree



Analysis

Centrality

highschool.indegree

```
##  1  8 10 11 17 18 24 23  9 29 12 26 19  2  3 13 20 14 15 21  4  5 40 47 48
38
##  2  0  0  3  4  2  1  2  4  1  3  4  6  1  2  5  4  5  6  7 12 13  2  0  1
3
## 25 22 49 57 50 30 51 52 53 55 39 54 41 35 31 16 60 62 58 61 56 63 32 27 33
```

```

34
## 5 5 5 2 2 6 3 4 5 4 7 7 5 6 7 10 1 5 2 4 6 8 7 11 8
8
## 6 7 65 64 69 59 42 70 36 43 66 44 68 67 37 45 28 46
## 9 13 2 1 2 2 4 3 8 8 5 6 8 8 10 8 18 14

```

Edge Betweenness

```

modularity(gv)

## [1] 0.2485645

length(gv)

## [1] 34

membership(gv)

## 1 8 10 11 17 18 24 23 9 29 12 26 19 2 3 13 20 14 15 21 4 5 40 47 48
38
## 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 21 22 23 24
25
## 25 22 49 57 50 30 51 52 53 55 39 54 41 35 31 16 60 62 58 61 56 63 32 27 33
34
## 26 27 28 29 30 31 32 28 31 28 21 32 28 31 31 31 33 32 32 34 32 32 32 32 32
32
## 6 7 65 64 69 59 42 70 36 43 66 44 68 67 37 45 28 46
## 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32

sizes(gv)

## Community sizes
## 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
26
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 3 1 1 1 1
1
## 27 28 29 30 31 32 33 34
## 1 4 1 1 5 28 1 1

```

[Using the in-degree of this network plot, we find that number 28 and 46 were the most interconnected with the other boys at this particular high school. They had the most connections out of all the other 70 boys at this schools. Three of the boys had no connections at this school, meaning they had no one to call as their friend at this high school. From our subgroup detection, we see that there were various groups at this high school being 34 different types of communities. This means that almost everyone had participated in some sort of a community besides one person. This shows that a high percentage of those who went to this school were still interconnected with each other in some sort of way.]

Conclusion

[A major takeaway from this project is how basic network analysis can still be applied to old data such as this data set. Despite newer and more efficient ways to analyze data, network analytic skills can still be applied to older data sets as long as there is some sort of relationship between the data. From our findings, we can see that in this high school, most boys are well connected with each other, with only one being not as connected as the rest. A majority of the boys knew each other or felt like they had at least one friend at their school.]