Social Media Analytics

Module 9 Assignment

Paras Ahuja

November 22, 2020

Abstract

In this paper we analyze Bernie Madoff Fraud network data available through networkdata library in R.[1] Bernie Madoff was involved in a massive financial fraud in the USA and was sentenced to 150 years in prison. This network consists of transaction flows between financial institutions and Madoff. Note that networkdata library is not available by default in the newer versions of RStudio, and R, therefore the package must be installed by utilizing the following code:

```r
remotes::install_github("schochastics/networkdata")
```

Data Collection

Data for this network was gathered from various news sources and court documents. David Schoch, creator of the dataset cites The Network Thinkers (TNT) blog as his inspiration.[2] Furthermore, The Wall Street Journal had published a list of names and addresses of people who fell victim to the ponzi scheme. The document published by The Wall Street Journal is 162 pages long.[3]

Data Files and Formats

All datasets in networkdata are an igraph object, including the dataset in question. In order to use statnet on this data, we need to use intergraph package of R so that we may convert igraph to a network object in statnet. The data represents directed network 61 x 61 firm by firm, showing money flows financial institutions and other organizations, leading ultimately to Bernie Madoff's fund. Relations are money flows. Arrows show direction of money flow.

Data Details

Since this is network data we have ample network statistics available here, which is what we utilize here as details. We first load the required libraries and proceed accordingly.

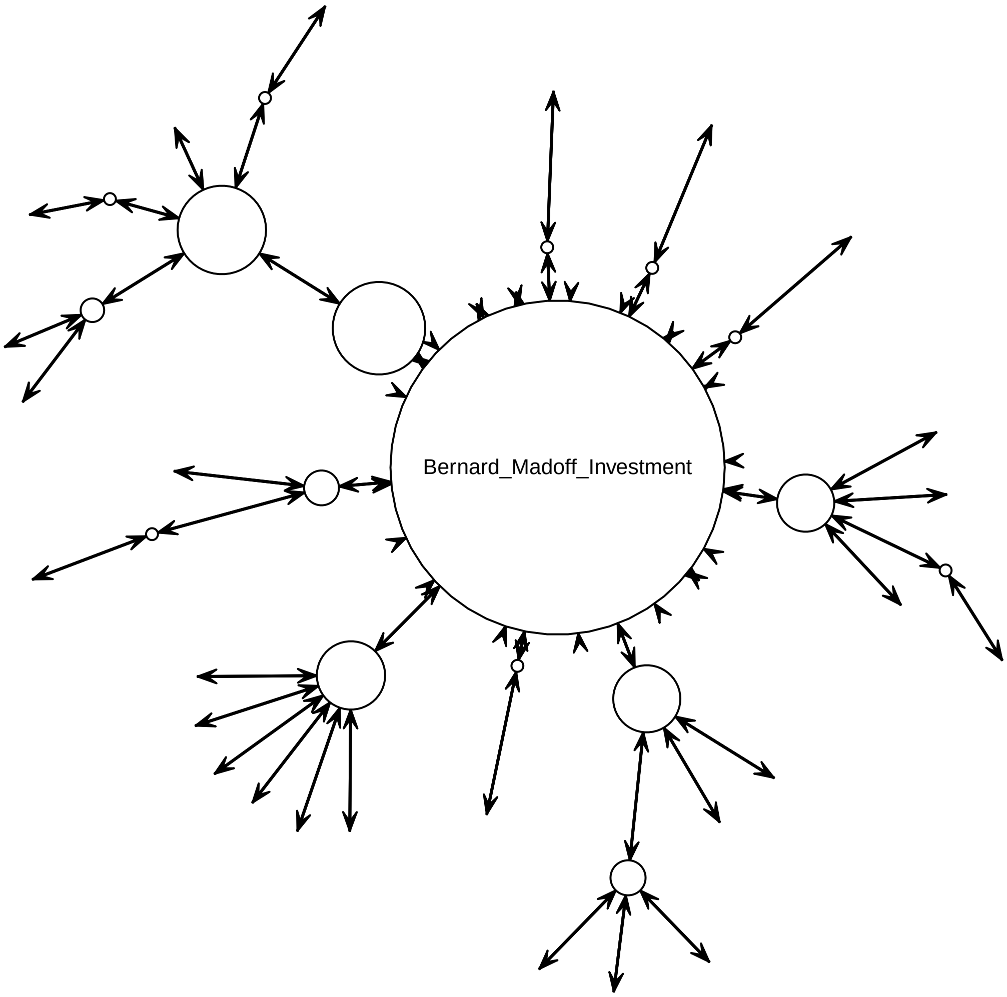
```r
library(statnet)
library(igraph)
library(networkdata)
data("covert_29")

matrix <- as_adjacency_matrix(covert_29, type = c("both", "upper", "lower"),
```

[1] https://www.rdocumentation.org/packages/networkdata/versions/0.1.4

[2] https://www.rdocumentation.org/packages/networkdata/versions/0.1.4/topics/covert_29

[3] http://www.thenetworkthinkers.com/2009/02/madoff-feeder-funds.html.

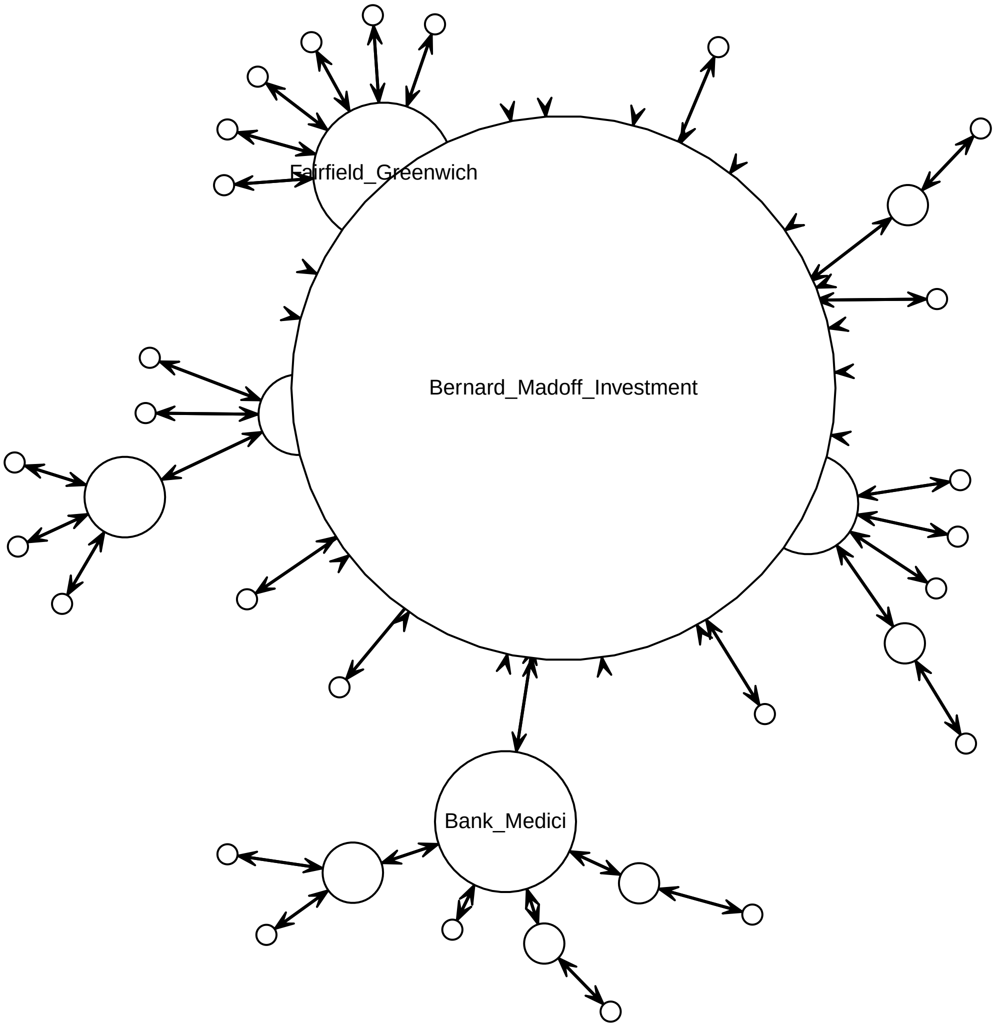
```
                              attr = NULL, edges = FALSE, names = TRUE,
                              sparse = igraph_opt("sparsematrices"))

network <- network(matrix, directed=TRUE, loops=FALSE)

plot.network(network,
             vertex.col = "White",
             displaylabels = T,
             label.cex = 0.7,
             vertex.cex= betweenness(covert_29)/100,
             label = ifelse(betweenness(covert_29)/100 > 5,
                            V(covert_29)$name, NA),
             label.pos = 5)
```



We notice here that Bernie Madoff Investment has a betweenness centrality of greater than 5. Betweenness centrality is a measure of number of time a node acts as a bridge between two other nodes. We also notice that there are a few large financial institutions that perhaps facilitated Bernie Madoff's rise. By simply lowering the betweenness centrality requirement to 2, we find out that chief among Madoff's clients and ultimately victims, were Cohmad Securities, Fairfield Greenwich, Bank Medici, Tremont Group Holdings, and Gabriel Capital. Another interesting measure to dwell upon is degree centrality. Degree centrality is defined as the number of links incident upon a node. Just by looking at the graph we notice that Madoff certainly had a lot of connections.

Below we notice that Bernie Madoff, for obvious reasons, has the highest degree centrality. What's also interesting is that two of his clients seemed to have acted as a conduit.

```
plot.network(network,
             vertex.col = "White",
             displaylabels = T,
             label.cex = 0.7,
             vertex.cex= degree(covert_29),
             label = ifelse(degree(covert_29) > 5,
                            V(covert_29)$name, " "),
             label.pos = 5)
```



Existing Publications

Nearly all existing publications are either from financial institutions, news media, SEC, or court documents. There are not many research papers that have been written about Madoff, and his fall. One document that served as division exhibit 8 in SEC filings against Madoff is Barron's report published on May 07, 2001. It's title: **Don't Ask, Don't Tell: Bernie Madoff is so secretive, he even asks his investors to keep mum.**[4]

[4] https://www.sec.gov.edgekey-staging.net/news/studies/2009/oig-509/exhibit-0156.pdf

ERGM

Note that the network has no attributes, and simply consists of adjacency matrix, and directed graph indicating direction of the transaction. Furthermore, attributes can be collected from publicly available information. One attribute that can be collected to expand the dataset is money lost by the investors, and their country. However, not all investors chose to reveal the amount they lost so attribute information will be incomplete. Country of investor is available in documents revealed by the news media.

```r
e2 <- ergm(network~edges)
summary(e2)
```

```
## Call:
## ergm(formula = network ~ edges)
##
## Iterations:  6 out of 20
##
## Monte Carlo MLE Results:
##       Estimate Std. Error MCMC % z value Pr(>|z|)
## edges -3.38439    0.09282      0  -36.46   <1e-04 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      Null Deviance: 5074  on 3660  degrees of freedom
##  Residual Deviance: 1056  on 3659  degrees of freedom
##
## AIC: 1058    BIC: 1064    (Smaller is better.)
```

```r
e1 <- ergm(network~edges+triangle)
summary(e1)
```

```
## Call:
## ergm(formula = network ~ edges + triangle)
##
## Iterations:  2 out of 20
##
## Monte Carlo MLE Results:
##          Estimate Std. Error MCMC % z value Pr(>|z|)
## edges     -3.1441     0.1105      0  -28.46   <1e-04 ***
## triangle     -Inf     0.0000      0    -Inf   <1e-04 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      Null Deviance: 5074  on 3660  degrees of freedom
##  Residual Deviance:  NaN  on 3658  degrees of freedom
##
## AIC: NaN    BIC: NaN    (Smaller is better.)
##
##  Warning: The following terms have infinite coefficient estimates:
##   triangle
```

RSiena s50

```r
library(RSiena)
```

Step 1

```
s501d <- read.table("s50-network1.dat", header = FALSE, sep = " ", quote = "\"'")
s502d <- read.table("s50-network2.dat", header = FALSE, sep = " ", quote = "\"'")
s503d <- read.table("s50-network3.dat", header = FALSE, sep = " ", quote = "\"'")

s501 <- as.matrix(s501d[, 2:51])
s502 <- as.matrix(s502d[, 2:51])
s503 <- as.matrix(s503d[, 2:51])

s50a <- read.table("s50-alcohol.dat", header = FALSE, sep = " ", quote = "\"'")
s50s <- read.table("s50-smoke.dat", header = FALSE, sep = " ", quote = "\"'")
```

Step 2

```
friend.data.w1 <- s501
friend.data.w2 <- s502
friend.data.w3 <- s503
drink <- as.matrix(s50a[, 2:4])
smoke <- as.matrix(s50s[, 2:4])

friendship <- sienaDependent(array(c(friend.data.w1, friend.data.w2, friend.data.w3),
    dim = c(50, 50, 3)))

smoke1 <- coCovar(smoke[, 1])
alcohol <- varCovar(drink)
mydata <- sienaDataCreate(friendship, smoke1, alcohol)
```

Step 3

```
myeff <- getEffects(mydata)
myalgorithm <- sienaAlgorithmCreate(projname = "s50_3")
myeff <- includeEffects(myeff, transTrip, cycle3)
myeff <- includeEffects(myeff, egoX, altX, egoXaltX, interaction1 = "alcohol")
myeff <- includeEffects(myeff, simX, interaction1 = "smoke1")
```

```
ans <- siena07(myalgorithm, data = mydata, effects = myeff)
ans
```

```
myeff[myeff$effectName == "transitive triplets" & myeff$type == "eval", "include"] <- TRUE
myeff[myeff$effectName == "3-cycles" & myeff$type == "eval", "include"] <- FALSE
myeff[myeff$effectName == "smoke1 similarity" & myeff$type == "eval", "include"] <- FALSE
myeff[myeff$effectName == "alcohol alter" & myeff$type == "eval", "include"] <- FALSE
myeff[myeff$effectName == "alcohol ego" & myeff$type == "eval", "include"] <- FALSE
myeff[myeff$effectName == "alcohol ego x alcohol alter" & myeff$type == "eval", "include"] <- TRUE

ans <- siena07(myalgorithm, data = mydata, effects = myeff)
ans
```

```
## Estimates, standard errors and convergence t-ratios
##
##                                         Estimate   Standard   Convergence
##                                                      Error      t-ratio
##
```

```
## Rate parameters:
##   0.1      Rate parameter period 1      6.6823  ( 1.1849   )
##   0.2      Rate parameter period 2      5.2200  ( 0.8741   )
##
## Other parameters:
##   1.  eval outdegree (density)         -2.7118  ( 0.1147   )   0.0352
##   2.  eval reciprocity                  2.4296  ( 0.1950   )   0.0594
##   3.  eval transitive triplets          0.6160  ( 0.0726   )   0.0859
##   4.  eval alcohol ego x alcohol alter  0.1290  ( 0.0479   )   0.0139
##
## Overall maximum convergence ratio:    0.1115
##
##
## Total of 1795 iteration steps.
```