George Washington University
6289,2018 fall

Report of Homework 4

Jiaying Liu

October 7, 2018

1 Question

The original data set is about hollywood actors and their movie. This is an affiliation network. In this homework, I will go through the analysis between slides 31 to 81 and create the similar graph like slide 55 and slide 81.

1.1 Theory

This time, the network we use is affiliation network. The definition of Affiliation networks is that it is a network where members are affiliated with one another based on

- co-membership in a group, or
- co-participation in some type of even

In our data set, it is about hollywood actors and their movie. The actors interact with each other when they work together. one movie can have several actors and one actors can have several movie.
In affiliation networks, there are always two type of nodes:

- actors
- groups to which the actors belong.

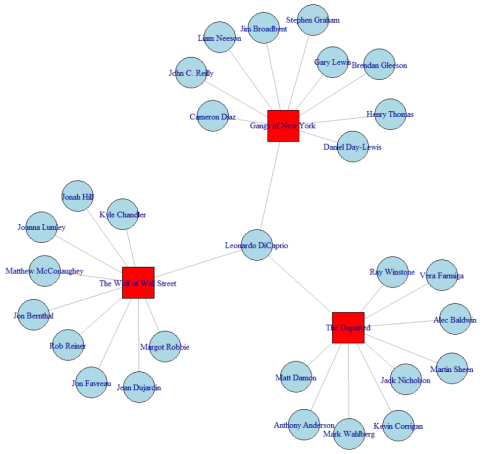
So in conclusion, if we use the projections if examine the connection among actors only or movies only.

1.2 Result

1.2.1 Repeat Result

Basic information about network:we can see the network is about movie and actors where the type of movies is TRUE and the type of actors is False. So by separate the node type and color with TRUE and FALSE, from the figure 1, we was able to find out the connection of movie and actors.

Figure 1: Degree Centrality



We curious about the degree of each node. About the degree is the the total number of edge of a node. So by using the degree of node, we was able to find out those popular actors who have a lot of work. In the following step, first I define the degree for network and find out the actors who have more then 4 movies. The result shows in the table 1.

Table 1: Popular movie actors

Name	**Name**
Leonardo DiCaprio	Emma Watson
Richard Griffiths	Harry Melling
Daniel Radcliffe	Rupert Grint
James Franco	Ian McKellen
Martin Freeman	Bradley Cooper
Christian Bale	Samuel L.Jackson
Natalie Portman	Brad Pitt
Liam Neeson	

From the table 2, with the degree of node, we was able to find out the busy actors.
There are another variable in the data set such as IMDBrating. This is a variable to assess the popularity of the movies they appeared in.From the table 3, we can see the actors and their average socre of BMDBrating.

Table 2: Busy actors

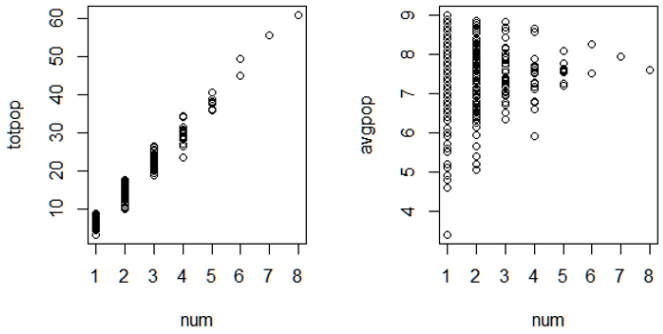
index	Name	number of movie
5	Daniel Radcliffe	8
11	Christian Bale	7
1	Leonardo DiCaprio	6
2	Emma Watson	6
3	Richard Griffiths	5
4	Harry Melling	5
6	Rupert Grint	5
7	James Franco	5
8	Ian McKellen	5
9	Martin Freeman	5
10	Bradley Cooper	5
12	Samuel L. Jackson	5
13	Natalie Portman	5
14	Brad Pitt	5
15	Liam Neeson	5

Table 3: IMDBrating

index	Name	Popularity
3	Daniel Radcliffe	60.9
4	Christian Bale	55.5
1	Leonardo DiCaprio	49.6
2	Emma Watson	45
5	Brad Pitt	40.5

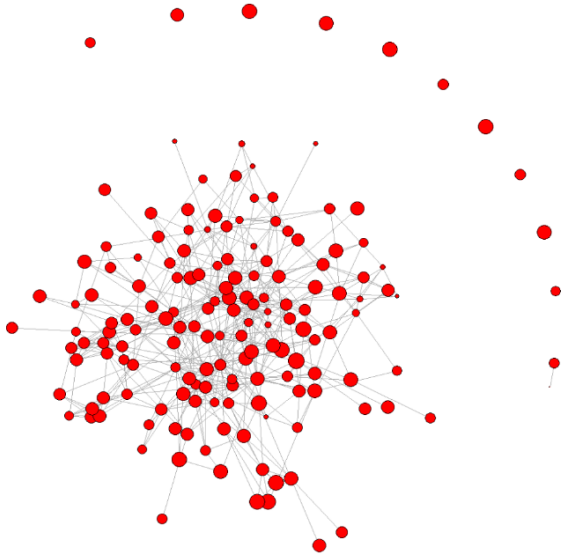
So this two analysis index, which one is more useful? Busier actors or more popular movies? In my opinion, the busier actor can show that he or she is hardworking and in the same time, it is sure that he or she is famous so that a lot of movie find them to perform. But busier do not mean popular movies. People would trend to like the movie which is more meaningful and attractive.
The reason we choose the average popularity score is because it is a effective measurement that can show the result fairly.

Figure 2: Total vs average



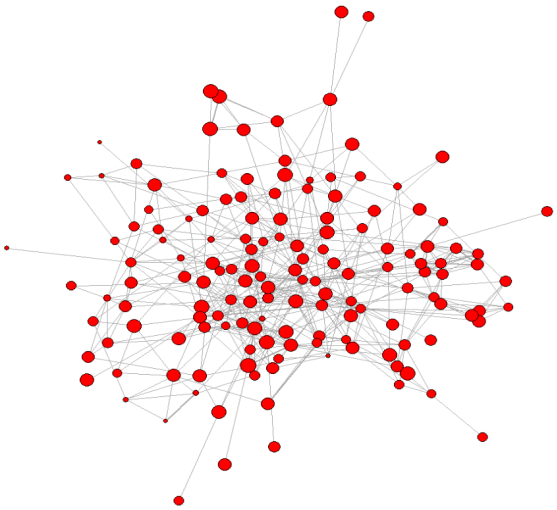
From now, we analysis the relation between actors and the movie. As we know, this is a affiliation network with two group one is actors and one is movie. So we use the projection to separation the network into actor projection and movie projection.
First, let's consider the whole picture of the movie without any reduction.

Figure 3: Projection of Movie



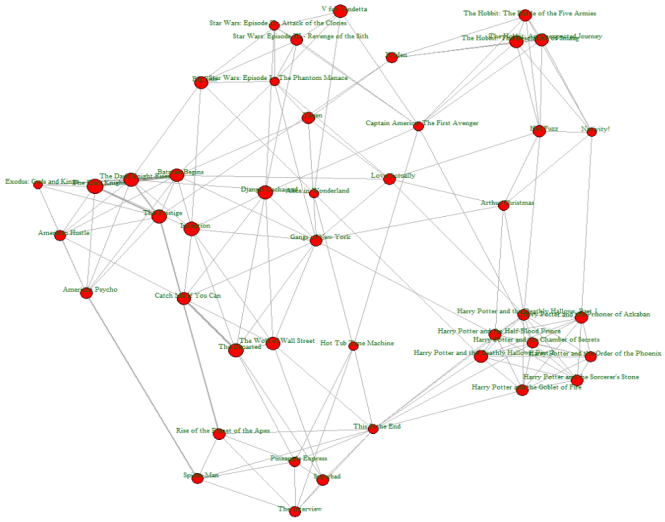
however, this figure is hard to identified, so in the next step, we use the cluster to reduce the network.

Figure 4: Projection of Movie



It is still difficult to understand the network network structure, we can use coreness to further reduce the movie network and then add the labels.

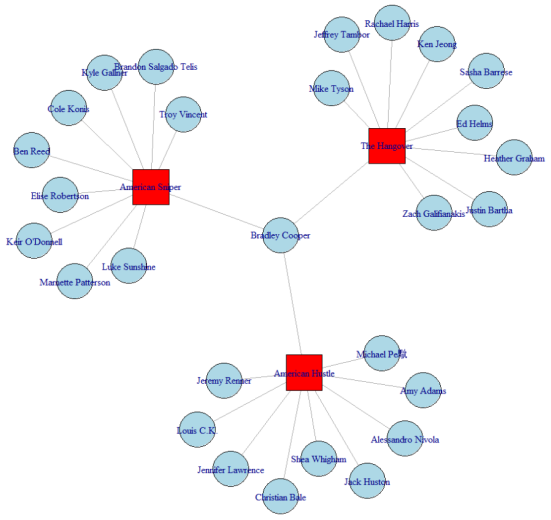
Figure 5: Projection of Movie



1.2.2 Bradley Cooper

In this section, we try to create a graph similar as figure 1. The first thing is to find the movie about Bradley Cooper. I choose three movie name "The Hangover","American Hustle" and "American Sipper"

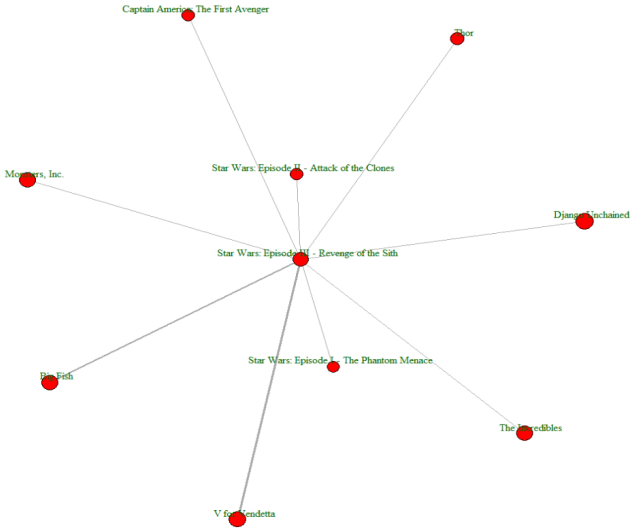
Figure 6: Closeness Centrality



1.2.3 Star WAR III

In this section, we try to create a graph for all the movies related to "Star Wars: Episode III - Revenge of the Sith". So in the first step, by using the subgraph.edges function to choose the node name "Star Wars: Episode III - Revenge of the Sith" and its edges in all movie connection network. The set the edge.width to be edge weight. and the vertex.size be the IMDBrating. So we get the result showing below.

Figure 7: Star Wars: Episode III - Revenge of the Sith



2 Appendix

```
library(UserNetR)
library(igraph)

data(hwd)

h1 <- hwd
h1

V(h1)$name[1:10]
V(h1)$name[155:165]
V(h1)$type[1:10]
V(h1)$type[155:165]


V(h1)$shape <- ifelse(
    V(h1)$type==TRUE,
    "square","circle")
V(h1)$color <- ifelse(
    V(h1)$type==TRUE,
    "red","lightblue")
h2 <- subgraph.edges(h1,
E(h1)[inc(V(h1)[name %in%
c("The Wolf of Wall Street",
  "Gangs of New York",
  "The Departed")])])
plot(h2, layout = layout_with_kk)


table(degree(h1,
    v=V(h1)[type==FALSE]))
mean(degree(h1,
    v=V(h1)[type==FALSE]))

V(h1)$deg <- degree(h1)

V(h1)[type==FALSE & deg > 4]$name

busy_actor <- data.frame(cbind(
Actor = V(h1)[type==FALSE
 & deg > 4]$name,
Movies = V(h1)[type==FALSE
 & deg > 4]$deg
))
busy_actor[order(
```

```
    busy_actor$Movies,
decreasing=TRUE),]


for (i in 161:1365) {
V(h1)[i]$totrating <-
    sum(V(h1)[nei(i)]$IMDBrating)
}


for (i in 161:1365) {
V(h1)[i]$avgrating <-
    mean(V(h1)[nei(i)]$IMDBrating)
}
pop_actor <- data.frame(cbind(
Actor = V(h1)[type==FALSE &
totrating > 40]$name,
Popularity = V(h1)[type==FALSE &
totrating > 40]$totrating
))

pop_actor[order(
    pop_actor$Popularity,
    decreasing=TRUE),
]


num <- V(h1)[type==FALSE]$deg

avgpop <- V(h1)[type==FALSE]$avgrating
totpop <- V(h1)[type==FALSE]$totrating

op <- par(mfrow=c(1,2))
plot(num,totpop)
plot(num,avgpop)
par(op)

summary(lm(avgpop~num))


h1.pr <- bipartite.projection(h1)
h1.act <- h1.pr$proj1
h1.mov <- h1.pr$proj2
h1.act
h1.mov

op <- par(mar = rep(0, 4))
plot(h1.mov,vertex.color="red",
```

```
    vertex.shape="circle",
    vertex.size=
        (V(h1.mov)$IMDBrating)-3,
vertex.label=NA)

par(op)

graph.density(h1.mov)
no.clusters(h1.mov)
clusters(h1.mov)$csize
table(E(h1.mov)$weight)

h2.mov <- induced.subgraph(h1.mov,
vids=clusters(h1.mov)$membership==1)

plot(h2.mov, vertex.color="red",
edge.width=sqrt(E(h1.mov)$weight),
vertex.shape="circle",
vertex.size=(V(h2.mov)$IMDBrating)-3,
vertex.label=NA)

table(graph.coreness(h2.mov))
h3.mov <- induced.subgraph(h2.mov,
    vids=graph.coreness(h2.mov)>4)
h3.mov

plot(h3.mov, vertex.color="red",
vertex.shape="circle",
edge.width=sqrt(E(h1.mov)$weight),
vertex.label.cex=0.7,
vertex.label.color="darkgreen",
vertex.label.dist=0.3,
vertex.size=
    (V(h3.mov)$IMDBrating)-3)


#Bradley Cooper

#V(h1)[V(h1)[type==F]$name=="Bradley Cooper" &type==T]$name
V(h1)$shape <- ifelse(
    V(h1)$type==TRUE,
    "square","circle")
V(h1)$color <- ifelse(
    V(h1)$type==TRUE,
    "red","lightblue")
h2 <- subgraph.edges(h1,
E(h1)[inc(V(h1)[name %in%
c("American_Sniper",
```

```
  "American Hustle",
  "The Hangover")])])
plot(h2, layout = layout_with_kk)

h1.pr <- bipartite.projection(h1)
h1.act <- h1.pr$proj1
h1.mov <- h1.pr$proj2
h1.act
h1.mov


h6.mov <- subgraph.edges(h1.mov,
E(h1.mov)[inc(V(h1.mov)[name %in%
c("Star Wars: Episode III – Revenge of the Sith")])])

plot(h6.mov, vertex.color="red",
vertex.shape="circle",
edge.width=(E(h1.mov)$weight),
vertex.label.cex=1,
vertex.label.color="darkgreen",
vertex.label.dist=0.5,
vertex.size=
    (V(h6.mov)$IMDBrating)-2)
```