# 1 Introduction

This is the solution of HW4 of Network Data, which consists of 3 parts.

First, we need to repeat the analysis/results described in Slides #31-81. Then, we need to create a graph similar as Slide #55 for the actor Bradley Cooper. Finally, we need to create a graph similar as Slide #81 for all the movies related to "Star Wars: Episode III - Revenge of the Sith".

# 2 Repeat the analysis/results described in Slides #31-81

3 Create a graph similar as Slide #55 for the actor Bradley Cooper
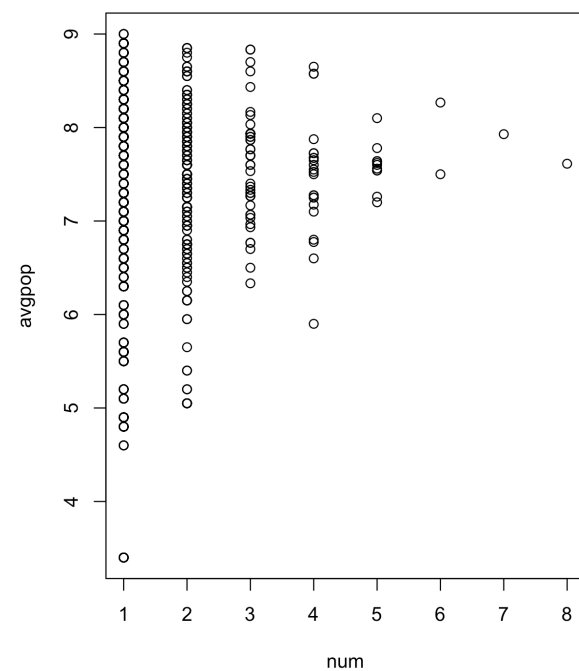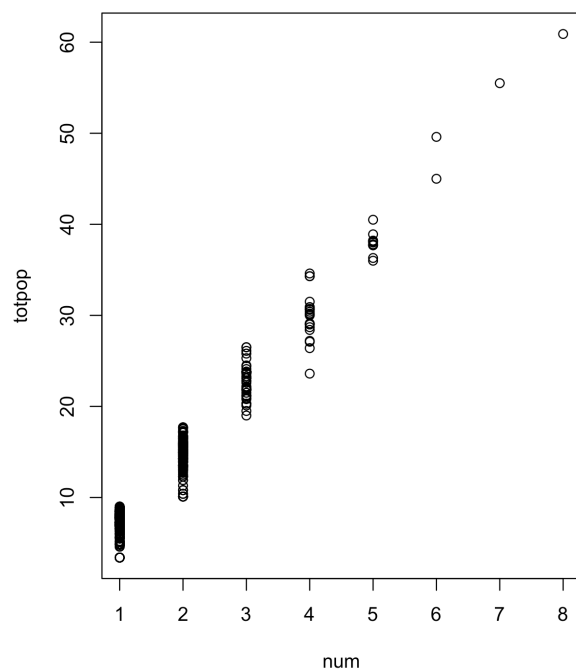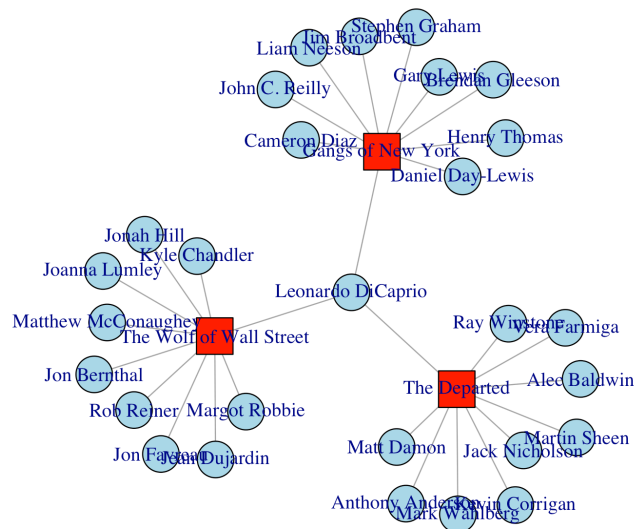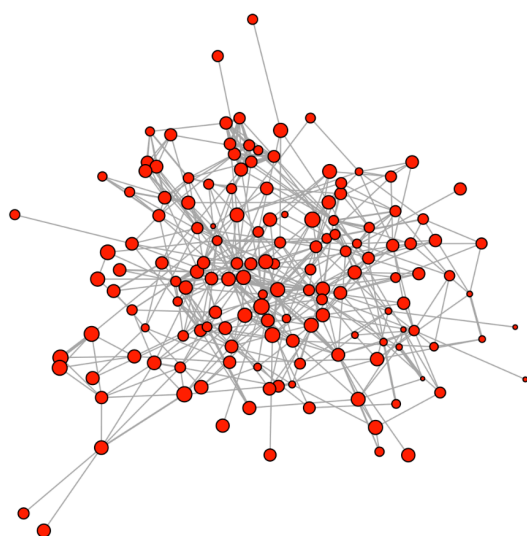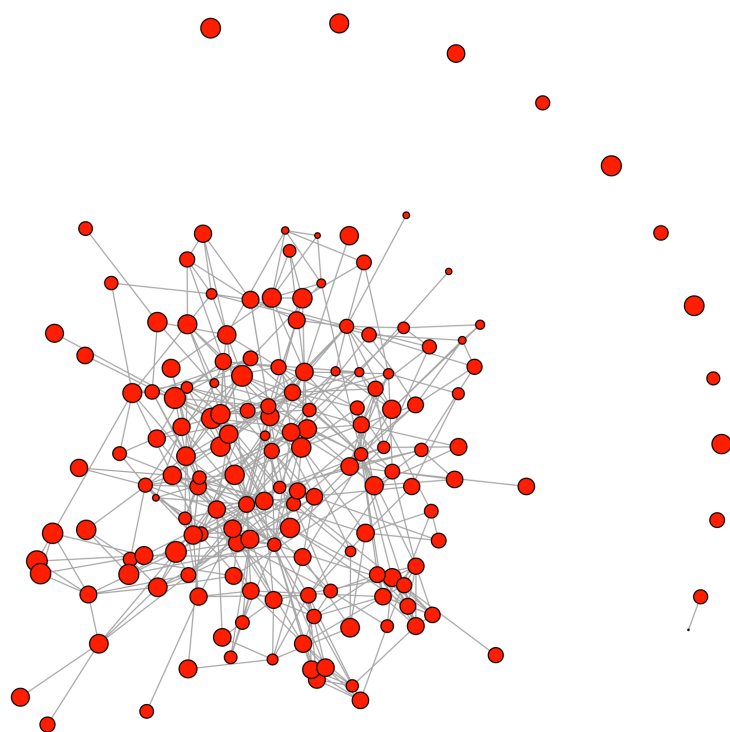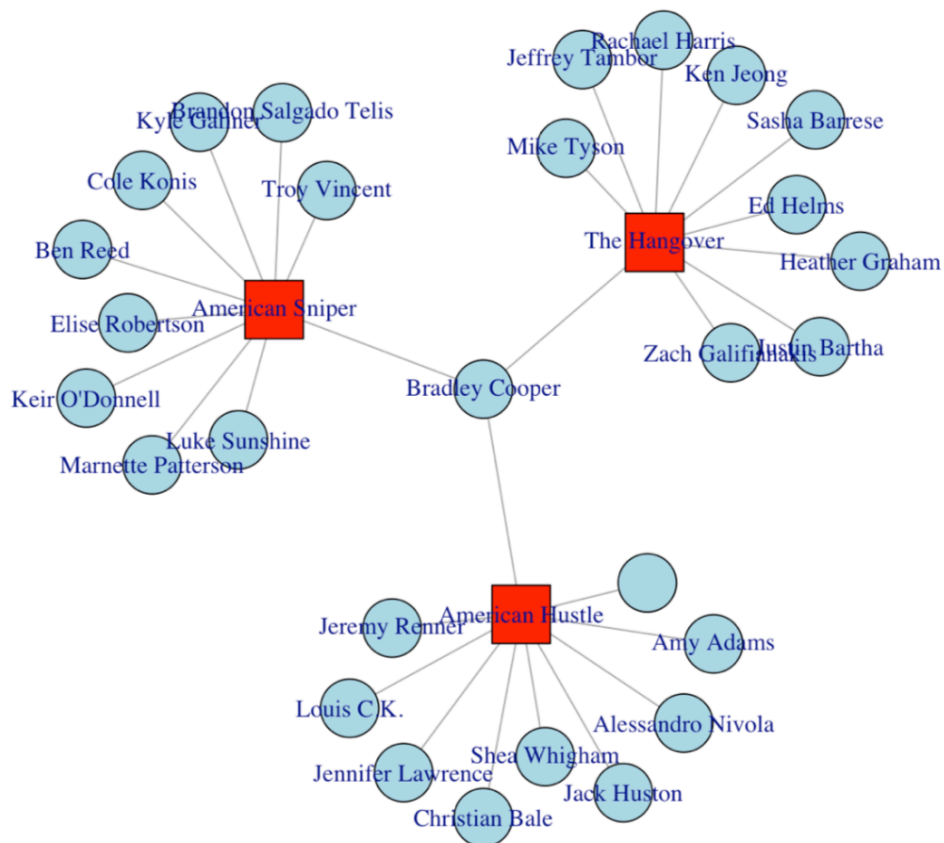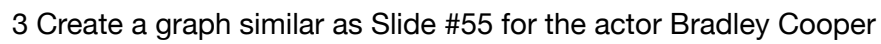
4 Create a graph similar as Slide #81 for all the movies related to "Star Wars: Episode III - Revenge of the Sith"
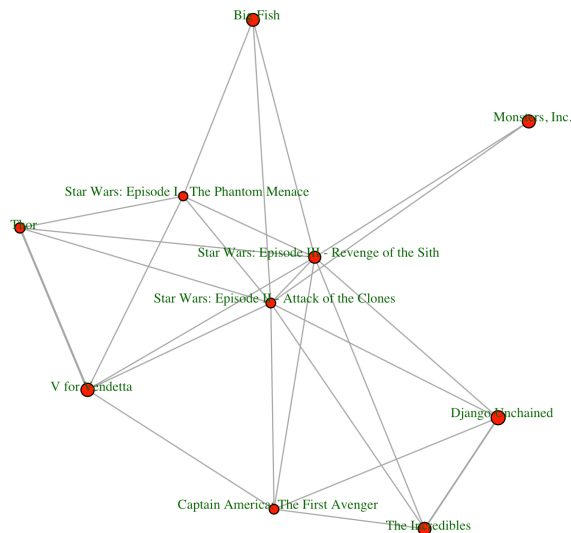


# 5 Appendix
Q1
```
> library(igraph)
> library(UserNetR)
> data(hwd)
>
> h1 <- hwd
> h1
IGRAPH 9cdab39 UN-B 1365 1600 --
+ attr: name (v/c), type (v/l), year (v/n), IMDBrating (v/n), MPAArating (v/c)
+ edges from 9cdab39 (vertex names):
 [1] Inception       --Leonardo DiCaprio    Inception        --Joseph Gordon-Levitt
 [3] Inception       --Ellen Page        Inception       --Tom Hardy
 [5] Inception       --Ken Watanabe       Inception        --Dileep Rao
 [7] Inception       --Cillian Murphy      Inception        --Tom Berenger
 [9] Inception       --Marion Cotillard    Inception       --Pete Postlethwaite
[11] Alice in Wonderland--Johnny Depp        Alice in Wonderland--Mia Wasikowska
[13] Alice in Wonderland--Helena Bonham Carter Alice in Wonderland--Anne Hathaway
[15] Alice in Wonderland--Crispin Glover      Alice in Wonderland--Matt Lucas
+ ... omitted several edges
>
> V(h1)$name[1:10]
 [1] "Inception"
 [2] "Alice in Wonderland"
 [3] "Kick-Ass"
```

```
 [4] "Toy Story 3"
 [5] "How to Train Your Dragon"
 [6] "Despicable Me"
 [7] "Scott Pilgrim vs. the World"
 [8] "Hot Tub Time Machine"
 [9] "Harry Potter and the Deathly Hallows: Part 1"
[10] "Tangled"
> V(h1)$name[155:165]
 [1] "Notting Hill"              "Eyes Wide Shut"          "The Green Mile"
 [4] "10 Things I Hate About You" "American Pie"           "Girl, Interrupted"
 [7] "Leonardo DiCaprio"         "Joseph Gordon-Levitt"    "Ellen Page"
[10] "Tom Hardy"                 "Ken Watanabe"
> V(h1)$type[1:10]
 [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
> V(h1)$type[155:165]
 [1]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE
>
>
> V(h1)$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]))

  1   2   3   4   5   6   7   8
955 165  47  23  11   2   1   1
> mean(degree(h1,
+         v=V(h1)[type==FALSE]))
[1] 1.327801
>
> V(h1)$deg <- degree(h1)
>
> V(h1)[type==FALSE & deg > 4]$name
 [1] "Leonardo DiCaprio" "Emma Watson"       "Richard Griffiths" "Harry Melling"
 [5] "Daniel Radcliffe"  "Rupert Grint"      "James Franco"      "Ian McKellen"
 [9] "Martin Freeman"    "Bradley Cooper"    "Christian Bale"    "Samuel L. Jackson"
[13] "Natalie Portman"   "Brad Pitt"         "Liam Neeson"
>
>
> 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),]
              Actor Movies
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
>
>
>
> 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),
+    ]
          Actor 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
>
>
```

```
>
> 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))

Call:
lm(formula = avgpop ~ num)

Residuals:
    Min    1Q  Median    3Q    Max
-3.9858 -0.4330  0.1977  0.6170  1.6142

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 7.33868    0.05440 134.911   <2e-16 ***
num         0.04714    0.03527   1.337    0.182
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9605 on 1203 degrees of freedom
Multiple R-squared:  0.001483,      Adjusted R-squared:  0.0006528
F-statistic: 1.786 on 1 and 1203 DF,  p-value: 0.1816


>
>
> h1.pr <- bipartite.projection(h1)
> h1.act <- h1.pr$proj1
> h1.mov <- h1.pr$proj2
> h1.act
IGRAPH 9f34fd4 UNW- 1205 6903 --
+ attr: name (v/c), year (v/n), IMDBrating (v/n), MPAArating (v/c), shape (v/c),
| color (v/c), deg (v/n), totrating (v/n), avgrating (v/n), weight (e/n)
+ edges from 9f34fd4 (vertex names):
 [1] Leonardo DiCaprio--Joseph Gordon-Levitt Leonardo DiCaprio--Ellen Page
 [3] Leonardo DiCaprio--Tom Hardy         Leonardo DiCaprio--Ken Watanabe
 [5] Leonardo DiCaprio--Dileep Rao        Leonardo DiCaprio--Cillian Murphy
 [7] Leonardo DiCaprio--Tom Berenger      Leonardo DiCaprio--Marion Cotillard
 [9] Leonardo DiCaprio--Pete Postlethwaite   Leonardo DiCaprio--Jonah Hill
[11] Leonardo DiCaprio--Matthew McConaughey  Leonardo DiCaprio--Margot Robbie
[13] Leonardo DiCaprio--Kyle Chandler     Leonardo DiCaprio--Rob Reiner
+ ... omitted several edges
> h1.mov
IGRAPH b888ac1 UNW- 160 472 --
+ attr: name (v/c), year (v/n), IMDBrating (v/n), MPAArating (v/c), shape (v/c),
| color (v/c), deg (v/n), totrating (v/n), avgrating (v/n), weight (e/n)
+ edges from b888ac1 (vertex names):
```

```
 [1] Inception--The Wolf of Wall Street    Inception--Django Unchained
 [3] Inception--The Departed              Inception--Gangs of New York
 [5] Inception--Catch Me If You Can       Inception--The Dark Knight Rises
 [7] Inception--10 Things I Hate About You Inception--Batman Begins
 [9] Inception--The Dark Knight           Inception--Training Day
[11] Inception--Big Fish
+ ... omitted several edges
>
>
>
> op <- par(mar = rep(0, 4))
> plot(h1.mov,vertex.color="red",
+     vertex.shape="circle",
+     vertex.size=
+      (V(h1.mov)$IMDBrating)-3,
+     vertex.label=NA)
>
> par(op)
>
>
>
> graph.density(h1.mov)
[1] 0.03710692
> no.clusters(h1.mov)
[1] 12
> clusters(h1.mov)$csize
 [1] 148   1   1   1   1   1   1   2   1   1   1   1
> table(E(h1.mov)$weight)


  1   2   3   4   5   6   7  10
411  21  12  16   6   1   2   3
>
> 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))


 1 2 3 4 5 6 7
11 5 23 65 29 7 8
> h3.mov <- induced.subgraph(h2.mov,
+                 vids=graph.coreness(h2.mov)>4)
> h3.mov
IGRAPH 81bece2 UNW- 44 158 --
+ attr: name (v/c), year (v/n), IMDBrating (v/n), MPAArating (v/c), shape (v/c),
| color (v/c), deg (v/n), totrating (v/n), avgrating (v/n), weight (e/n)
+ edges from 81bece2 (vertex names):
[1] Alice in Wonderland                 --Hot Tub Time Machine
```

[2] Hot Tub Time Machine                    --The Interview
[3] The Hobbit: The Battle of the Five Armies   --The Hobbit: The Desolation of Smaug
[4] Inception                    --The Wolf of Wall Street
[5] Exodus: Gods and Kings                 --American Hustle
[6] Hot Tub Time Machine                 --This Is the End
[7] Harry Potter and the Deathly Hallows: Part 1--This Is the End
+ ... omitted several edges
>
> 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)

Q2
> V(h1)[name %in% c("The Hangover","American Hustle","American Sniper")]
+ 3/1365 vertices, named, from 9cdab39:
[1] American Sniper American Hustle The Hangover
> h4 <- subgraph.edges(h1, E(h1)[inc(V(h1)[name %in% c("The Hangover","American
Hustle","American Sniper")])])])
> plot(h4, layout = layout_with_kk)

Q3
> sw3 <- induced.subgraph(h1.mov,vids=c(V(h1.mov)[nei("Star Wars: Episode III - Revenge of
the Sith")],V(h1.mov)[name%in%"Star Wars: Episode III - Revenge of the Sith"]))
> plot(sw3,vertex.color="red",
+     vertex.shape="circle",
+     edge.width=sqrt(E(h1.mov)$weight),
+     vertex.label.cex=0.7,
+     vertex.label.color="darkgreen",
+     vertex.label.dist=0.3,
+     vertex.size=
+      (V(sw3)$IMDBrating)-3)