Marvel Social Network Analysis

Laura Le

2/18/2021

DATASET:

The Marvel Universal Social Network : https://www.kaggle.com/csanhueza/the-marvel-universe-social-network.

The dataset contains heroes and comics, and the relationship between them. The dataset is divided into three files: node types (hero, comic), edges (which comic the heroes appear) and hero-edge (heroes which appear together in the comics).

PROBLEM DESCRIPTION:

There are hundreds of thousands of heroes in the Marvel Universe and they appear in an extensive list of comics. I'm interested in learning the relationship among heroes and how they appear in comics.

There are several questions that I will analyze this social network to figure out the answers:

- What are the most popular heroes in Marvel comics?
- Which heroes usually appear together?
- How the teams are formed and the connections between members?

GENERAL APPROACH:

I will explore this social network problem by evaluating the network size, density, centralization, reciprocity and hierarchy of different levels of network:

- Node: node level analysis to understand which nodes have higher degree centrality in the network. By determining the betweenness centrality and closeness centrality, I could define the importance or position of each actor in the Marvel universe social network.
- Sub-group level: sub-group level analysis to detect communities in Marvel network by finding dense subgraph because the team expects the graph to be relatively dense with high connectivity.

```r
library(ggplot2)
library(readr)
library(igraph)
```

```
##
## Attaching package: 'igraph'
```

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

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

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:igraph':
##
##     as_data_frame, groups, union
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
#### Import data

edges <- read.csv("~/Documents/Network class/Data Marvel/edges.csv")


hero.network <- read.csv("~/Documents/Network class/Data Marvel/hero-network.
csv", header=FALSE)


nodes <- read.csv("~/Documents/Network class/Data Marvel/nodes.csv")
```

```r
head(edges,10)
```

```
##                    hero    comic
## 1  24-HOUR MAN/EMMANUEL   AA2 35
## 2  3-D MAN/CHARLES CHAN    AVF 4
## 3  3-D MAN/CHARLES CHAN    AVF 5
## 4  3-D MAN/CHARLES CHAN    COC 1
## 5  3-D MAN/CHARLES CHAN   H2 251
## 6  3-D MAN/CHARLES CHAN   H2 252
## 7  3-D MAN/CHARLES CHAN M/PRM 35
## 8  3-D MAN/CHARLES CHAN M/PRM 36
## 9  3-D MAN/CHARLES CHAN M/PRM 37
## 10 3-D MAN/CHARLES CHAN    WI? 9
```

```r
head(hero.network,10)
```

```
##                      V1                  V2
## 1                 hero1               hero2
## 2         LITTLE, ABNER       PRINCESS ZANDA
## 3         LITTLE, ABNER BLACK PANTHER/T'CHAL
## 4  BLACK PANTHER/T'CHAL       PRINCESS ZANDA
## 5         LITTLE, ABNER       PRINCESS ZANDA
## 6         LITTLE, ABNER BLACK PANTHER/T'CHAL
## 7  BLACK PANTHER/T'CHAL       PRINCESS ZANDA
## 8  STEELE, SIMON/WOLFGA     FORTUNE, DOMINIC
## 9  STEELE, SIMON/WOLFGA  ERWIN, CLYTEMNESTRA
## 10 STEELE, SIMON/WOLFGA IRON MAN/TONY STARK
```

```r
head(nodes,10)
```

```
##                    node  type
## 1              2001 10 comic
## 2               2001 8 comic
## 3               2001 9 comic
## 4  24-HOUR MAN/EMMANUEL  hero
## 5  3-D MAN/CHARLES CHAN  hero
## 6      4-D MAN/MERCURIO  hero
## 7              8-BALL/  hero
## 8                A '00 comic
## 9                A '01 comic
## 10               A 100 comic
```

```r
# Get the dimension and see every column in dataframes

dim(edges) #there are 96104 observations of 2 variables
```

```
## [1] 96104     2
```

```r
glimpse(edges)
```

```
## Rows: 96,104
## Columns: 2
## $ hero  <chr> "24-HOUR MAN/EMMANUEL", "3-D MAN/CHARLES CHAN", "3-D MAN/CHA
RLE…
## $ comic <chr> "AA2 35", "AVF 4", "AVF 5", "COC 1", "H2 251", "H2 252", "M/
PRM…
```

```r
dim(hero.network) #there are 574468 observations of 2 variables
```

```
## [1] 574468      2
```

```r
glimpse(hero.network)
```

```
## Rows: 574,468
## Columns: 2
## $ V1 <chr> "hero1", "LITTLE, ABNER", "LITTLE, ABNER", "BLACK PANTHER/T'CHA
L",…
## $ V2 <chr> "hero2", "PRINCESS ZANDA", "BLACK PANTHER/T'CHAL", "PRINCESS ZA
NDA…
```

```r
dim(nodes) #there are 19090 observations of 2 variables
```

```
## [1] 19090     2
```

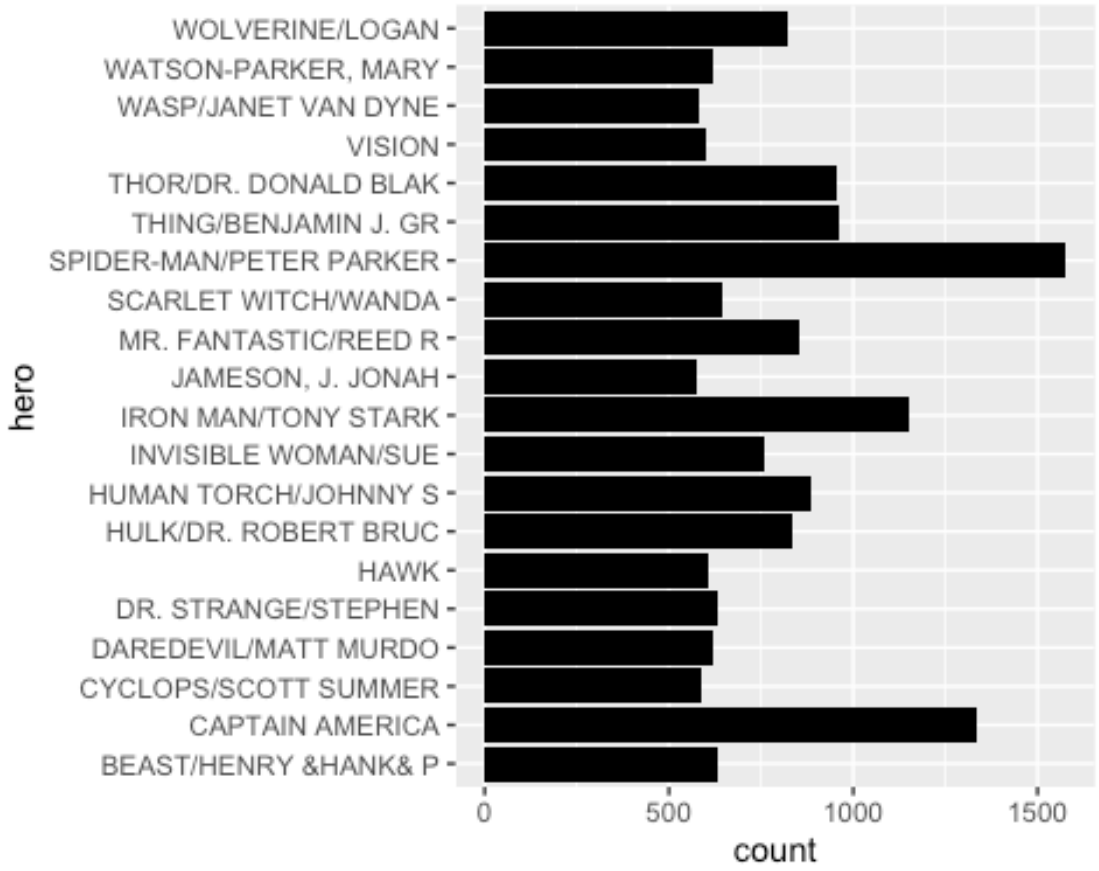
```r
glimpse(nodes)
```

```
## Rows: 19,090
## Columns: 2
## $ node <chr> "2001 10", "2001 8", "2001 9", "24-HOUR MAN/EMMANUEL", "3-D M
AN/…
## $ type <chr> "comic", "comic", "comic", "hero", "hero", "hero", "hero", "c
omi…
```

```r
# Top 5 heroes appear the most in all Marvel comics

edges_top<-edges%>%select(hero)%>%group_by(hero)%>%summarize(count=n())%>%arr
ange(desc(count))
edges_top<-as.data.frame(edges_top[1:20,])
head(edges_top)
```

```
##                         hero count
## 1 SPIDER-MAN/PETER PARKER  1577
## 2          CAPTAIN AMERICA  1334
## 3      IRON MAN/TONY STARK  1150
## 4     THING/BENJAMIN J. GR   963
## 5     THOR/DR. DONALD BLAK   956
## 6     HUMAN TORCH/JOHNNY S   886
```

```r
# Plot top 20 characters that have highest appearancy in Marvel comics

edges_top_plot<-edges%>%filter(hero%in%edges_top$hero)
g <- ggplot(edges_top_plot, aes(hero))
g + geom_bar(fill = "#000000")+coord_flip()
```



Create 3 sample graphs from the Marvel Social network

```
# Sample Social Network graph (1)

set.seed(1234)
herodf1 <- head(hero.network,100)
hero_g1<-graph_from_data_frame(herodf1, directed = F)

# Sample Social Network graph (2)

N <- 1600
hero.network2 <- hero.network[-(1:N), , drop = FALSE]
herodf2 <- head(hero.network2,100)
hero_g2<-graph_from_data_frame(herodf2, directed = F)



# Sample Social Network graph (3)

herodf3 <- tail(hero.network,100)
hero_g3<-graph_from_data_frame(herodf3, directed = F)
```

Analyze Social Network graph (1)

NODE ANALYSIS

```
# Check graph edges and vertices

V(hero_g1) #contents in vertices
```

```
## + 25/25 vertices, named, from a16951f:
##  [1] hero1                LITTLE, ABNER        BLACK PANTHER/T'CHAL
##  [4] STEELE, SIMON/WOLFGA RAVEN, SABBATH II/EL IRON MAN IV/JAMES R.
##  [7] IRON MAN/TONY STARK  ERWIN, CLYTEMNESTRA  PRINCESS ZANDA
## [10] CARNIVORE/COUNT ANDR GHOST                ZIMMER, ABE
## [13] FU MANCHU           SHANG-CHI             SMITH, SIR DENIS NAY
## [16] STARSHINE II/BRANDY MAN-THING/THEODORE T TARR, BLACK JACK
## [19] WU, LEIKO           JACKSON, STEVE        RESTON, CLIVE
## [22] ROM, SPACEKNIGHT    hero2                 FORTUNE, DOMINIC
## [25] DOCTOR DREDD
```

```r
gorder(hero_g1) # Count number of vertices
```

```
## [1] 25
```

```r
E(hero_g1) #contents in edges
```

```
## + 100/100 edges from a16951f (vertex names):
##  [1] hero1                --hero2
##  [2] LITTLE, ABNER        --PRINCESS ZANDA
##  [3] LITTLE, ABNER        --BLACK PANTHER/T'CHAL
##  [4] BLACK PANTHER/T'CHAL--PRINCESS ZANDA
##  [5] LITTLE, ABNER        --PRINCESS ZANDA
##  [6] LITTLE, ABNER        --BLACK PANTHER/T'CHAL
##  [7] BLACK PANTHER/T'CHAL--PRINCESS ZANDA
##  [8] STEELE, SIMON/WOLFGA--FORTUNE, DOMINIC
##  [9] STEELE, SIMON/WOLFGA--ERWIN, CLYTEMNESTRA
## [10] STEELE, SIMON/WOLFGA--IRON MAN/TONY STARK
## + ... omitted several edges
```

```r
gsize(hero_g1)# Count number of edges
```

```
## [1] 100
```

```r
# Measure the size of network
```

```r
diameter(hero_g1, directed=FALSE, weights=NA) #the length of the longest path between two nodes is 4
```

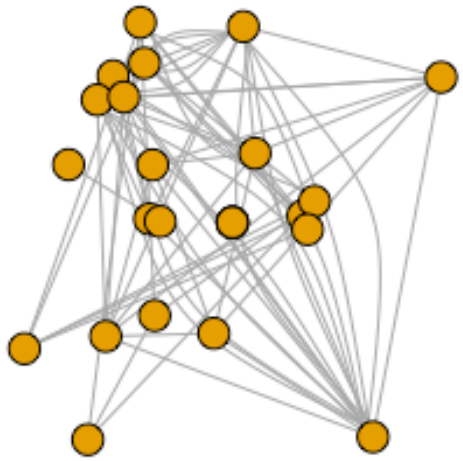
```
## [1] 2
```

```r
get_diameter(hero_g1, directed=FALSE, weights=NA) # identify the longest path
```

```
## + 3/25 vertices, named, from a16951f:
## [1] STEELE, SIMON/WOLFGA IRON MAN IV/JAMES R. GHOST
```

```r
# Plot social networks
```

```r
plot(hero_g1, layout = layout_with_lgl(hero_g1), vertex.label=NA)
```



```r
# Compute edge_density

edge_density(hero_g1)
```

```
## [1] 0.3333333
```

```r
# Compute mean_distance of graph

mean_distance(hero_g1, directed = FALSE)
```

```
## [1] 1.3
```

```r
# Compute clustering coefficient to find the probability that the adjacent ve
rtices of a vertex are connected

transitivity(hero_g1, type = "average")
```

```
## [1] 0.8881643
```

```r
# Calculate the degree

hero_deg <- degree(hero_g1, mode = c("all"))
which.max(hero_deg)
```

```
## SMITH, SIR DENIS NAY
##                   15
```

```r
# Top 3 most popular

top<-mean(hero_deg)+ 1.5*sd(hero_deg)
length(hero_deg[hero_deg>top])
```

```
## [1] 3
```

```r
hero_deg[hero_deg>top]
```

```
##        LITTLE, ABNER BLACK PANTHER/T'CHAL SMITH, SIR DENIS NAY
##                   16                   16                   18
```

```r
# Calculate betweenness of each vertex to find the degree of which heroes sta
nd between each other

betw <- betweenness(hero_g1, directed = F)
which.max(betw)
```

```
## SMITH, SIR DENIS NAY
##                   15
```

```r
# Betweeness of top most popular heroes

top<-mean(betw)+ 0.8*sd(betw)
length(betw[betw>top])
```

```
## [1] 3
```

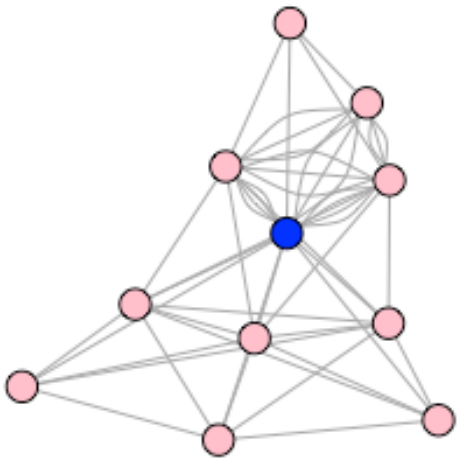
```r
betw[betw>top]
```

```
## IRON MAN IV/JAMES R. IRON MAN/TONY STARK  SMITH, SIR DENIS NAY
##              4.00000              4.00000             11.62857
```

```r
# Identify key nodes using eigenvector centrality to measure the influence of
a node in a network

g.ec <- eigen_centrality(hero_g1)
which.max(g.ec$vector)
```

```
## LITTLE, ABNER
##             2
```

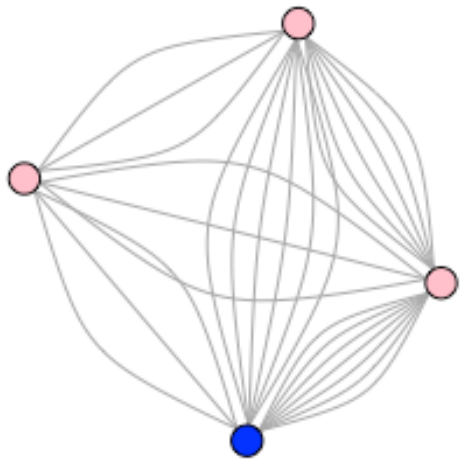
```r
# Measure the influence of top most popular heroes

top<-mean(g.ec$vector)+ 1.8*sd(g.ec$vector)
length(g.ec$vector[g.ec$vector>top])

## [1] 3

g.ec$vector[g.ec$vector>top]

##        LITTLE, ABNER BLACK PANTHER/T'CHAL       PRINCESS ZANDA
##            1.0000000            1.0000000            0.9513032



# Sir Denis Nayland Smith is having the most connections and control over the
network.
# Find who is around Sir Denis Nayland Smith ?

g_sdennis <- make_ego_graph(hero_g1, diameter(hero_g1), nodes = 'SMITH, SIR D
ENIS NAY', mode = c("all"))[[1]]
V(g_sdennis)$color <- ifelse(V(g_sdennis)$name=="SMITH, SIR DENIS NAY","blue"
,"pink")
plot(g_sdennis, vertex.label=NA)
```



```r
# Neighbors of Sir Denis Nayland Smith

unique(neighbors(hero_g1, v=which(V(hero_g1)$name=="SMITH, SIR DENIS NAY")))
```

```
## + 10/25 vertices, named, from a16951f:
##  [1] FU MANCHU            SHANG-CHI            STARSHINE II/BRANDY
##  [4] MAN-THING/THEODORE T TARR, BLACK JACK     WU, LEIKO
##  [7] JACKSON, STEVE       RESTON, CLIVE        ROM, SPACEKNIGHT
## [10] DOCTOR DREDD
```

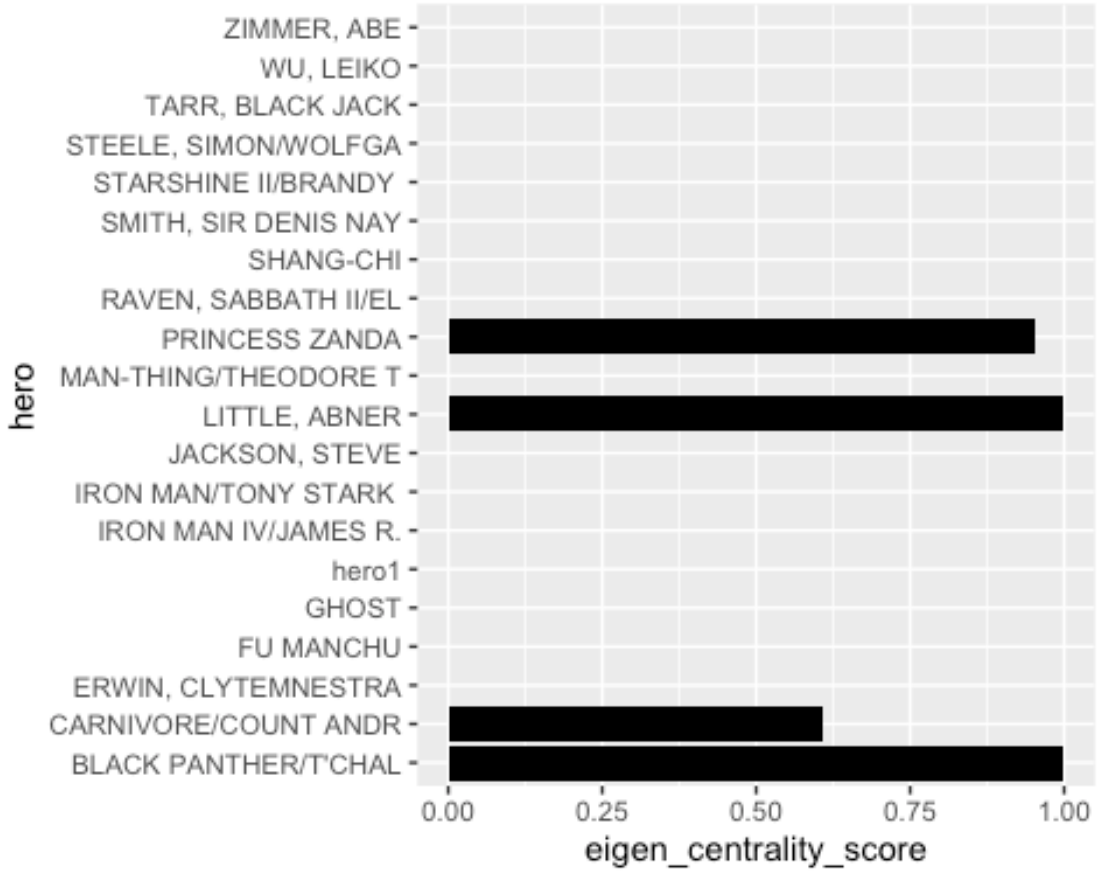
```r
# Black Panther is the most influence character
# Find who is around Black Panther?

g_blackpanther <- make_ego_graph(hero_g1, diameter(hero_g1), nodes = "BLACK P
ANTHER/T'CHAL", mode = c("all"))[[1]]
V(g_blackpanther)$color <- ifelse(V(g_blackpanther)$name=="BLACK PANTHER/T'CH
AL","blue","pink")
plot(g_blackpanther, vertex.label=NA)
```



```r
# Neighbors of Black Panther

unique(neighbors(hero_g1, v=which(V(hero_g1)$name=="BLACK PANTHER/T'CHAL")))
```

```
## + 3/25 vertices, named, from a16951f:
## [1] LITTLE, ABNER        PRINCESS ZANDA       CARNIVORE/COUNT ANDR
```

```r
# Use centrality to summarize which Marvel characteristics have more connecti
ons than others

hero_g_eigen_centrality_people=as.data.frame(eigen_centrality(hero_g1)$vector
)
hero_g_eigen_centrality_people$hero=rownames(hero_g_eigen_centrality_people)
rownames(hero_g_eigen_centrality_people)<-1:nrow(hero_g_eigen_centrality_peop
le)
colnames(hero_g_eigen_centrality_people)<-c("eigen_centrality_score","hero")


# Identify which Marvel characteristics are more important than others in sel
ected first 20 characters

hero_g_eigen_centrality_people_20<-hero_g_eigen_centrality_people[1:20,]



# According to eigen centrality score, Black Panther and Li'l Abner are the m
ost influence nodes within this network

herro_connection <- ggplot(hero_g_eigen_centrality_people_20, aes(x=hero,y=ei
gen_centrality_score))
herro_connection + geom_bar(stat="identity", fill = "#000000")+coord_flip()
```

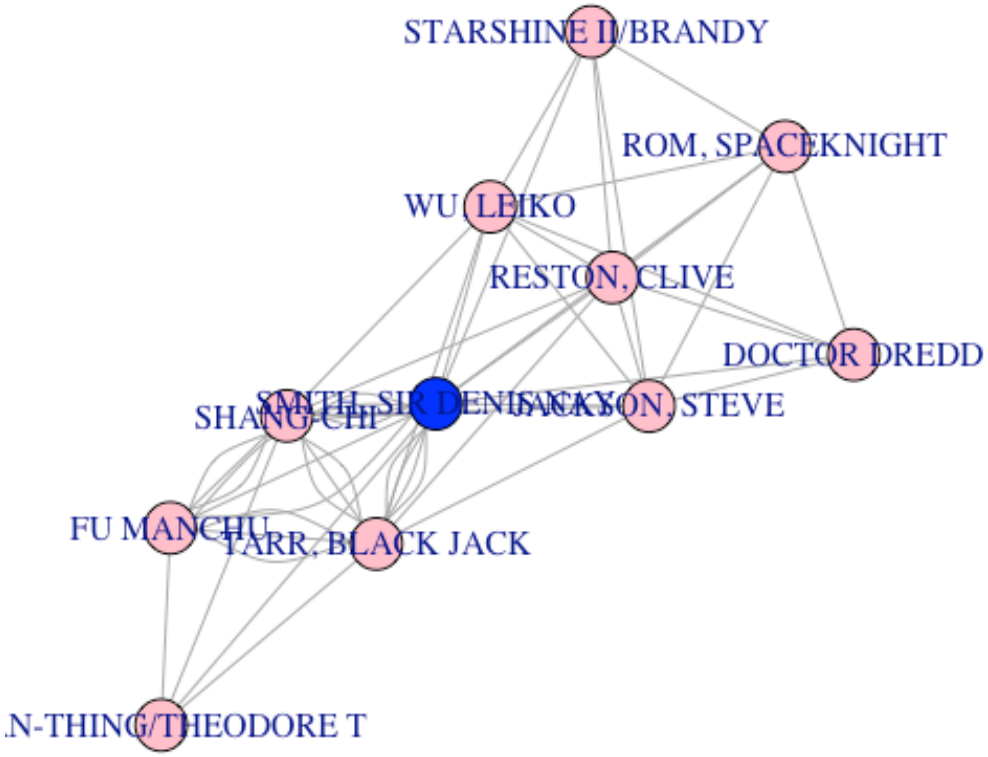


SUBGROUP ANALYSIS

```
# Identify clusters or communities of nodes in hero network

components(hero_g1) #this network has 4 components
```

```
## $membership
##                 hero1        LITTLE, ABNER BLACK PANTHER/T'CHAL
##                     1                    2                     2
## STEELE, SIMON/WOLFGA RAVEN, SABBATH II/EL IRON MAN IV/JAMES R.
##                     3                    3                     3
## IRON MAN/TONY STARK   ERWIN, CLYTEMNESTRA       PRINCESS ZANDA
##                     3                    3                     2
## CARNIVORE/COUNT ANDR               GHOST          ZIMMER, ABE
##                     2                    3                     3
##            FU MANCHU            SHANG-CHI SMITH, SIR DENIS NAY
##                     4                    4                     4
## STARSHINE II/BRANDY  MAN-THING/THEODORE T     TARR, BLACK JACK
##                     4                    4                     4
##            WU, LEIKO       JACKSON, STEVE        RESTON, CLIVE
##                     4                    4                     4
##     ROM, SPACEKNIGHT               hero2     FORTUNE, DOMINIC
##                     4                    1                     3
##         DOCTOR DREDD
##                     4
##
## $csize
## [1]  2  4  8 11
##
## $no
## [1] 4
```

```
# We will analyze the component 4 which have the largest size

hero_subgroup1 <- decompose(hero_g1)[[4]]

par(mar=c(0,0,0,0))
V(hero_subgroup1)$color <- ifelse(V(hero_subgroup1)$name=="SMITH, SIR DENIS N
AY","blue","pink")
plot(hero_subgroup1,cex=0.5)
```



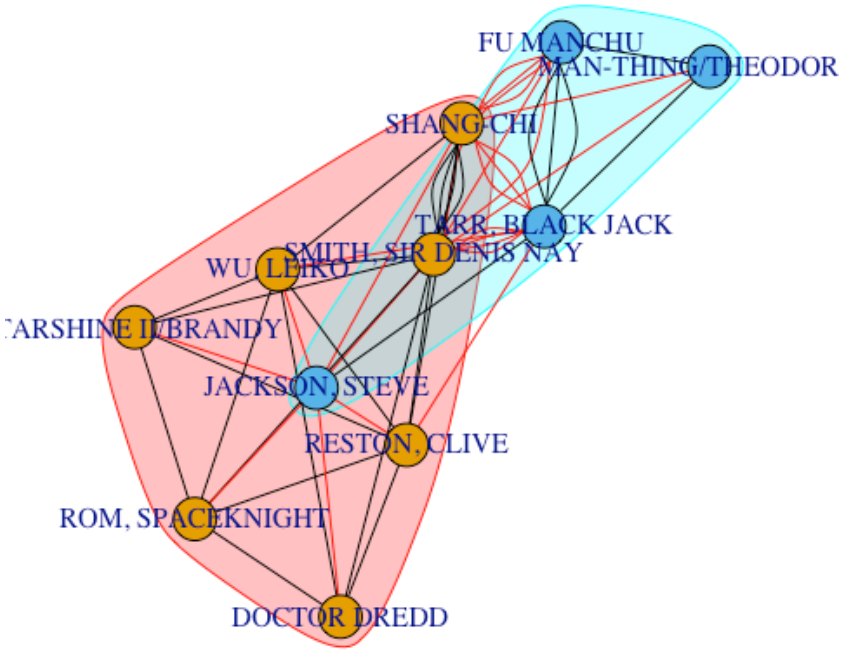

```
cluster_infomap(hero_subgroup1)
```

```
## IGRAPH clustering infomap, groups: 2, mod: -0.045
## + groups:
##   $`1`
##   [1] "SHANG-CHI"            "SMITH, SIR DENIS NAY" "STARSHINE II/BRANDY "
##   [4] "MAN-THING/THEODORE T" "WU, LEIKO"            "JACKSON, STEVE"
##   [7] "ROM, SPACEKNIGHT"     "DOCTOR DREDD"
##
##   $`2`
##   [1] "FU MANCHU"        "TARR, BLACK JACK" "RESTON, CLIVE"
##
```

```
# Map the flow of information in hero network, and the different clusters in
which information may get remain for longer periods

comm <- cluster_infomap(hero_subgroup1)
modularity(comm) # modularity score
```

```
## [1] -0.0192
```

```
# Plot the resulting communities

par(mar=c(0,0,0,0))
plot(comm, hero_subgroup1)
```



Analyze Social Network graph (2)

NODE ANALYSIS

```
# Check graph edges and vertices

V(hero_g2) #contents in vertices
```

```
## + 27/27 vertices, named, from c4b7128:
##  [1] LEEDS, BETTY BRANT   MAXWELL, MORRIS      THORSON, DR. WALTER
##  [4] SPIDER-MAN/PETER PAR THOMPSON, EUGENE FLA WATSON-PARKER, MARY
##  [7] ICEMAN/ROBERT BOBBY  OVERRIDE/DR. GREGORY KWAN, TERRY
## [10] URICH, BEN           DOLMAN               THOR/DR. DONALD BLAK
## [13] TOKKOTS              MCCORMICK, BARRY     JAMESON, J. JONAH
## [16] PARKER, MAY          FAIRMONT, HANNAH     ANGEL/WARREN KENNETH
## [19] MANSLAUGHTER         GRANT, GLORIA GLORY  NORRISS, SISTER BARB
## [22] STAR THIEF II        GARGOYLE II/ISAAC CH KUBIK
## [25] CLOUD                BEAST/HENRY &HANK& P ANDROMEDA/ANDROMEDA
```

```r
gorder(hero_g2) # Count number of vertices
```

```
## [1] 27
```

```r
E(hero_g2) #contents in edges
```

```
## + 100/100 edges from c4b7128 (vertex names):
##  [1] LEEDS, BETTY BRANT --OVERRIDE/DR. GREGORY
##  [2] LEEDS, BETTY BRANT --ICEMAN/ROBERT BOBBY
##  [3] LEEDS, BETTY BRANT --WATSON-PARKER, MARY
##  [4] LEEDS, BETTY BRANT --THOMPSON, EUGENE FLA
##  [5] LEEDS, BETTY BRANT --SPIDER-MAN/PETER PAR
##  [6] LEEDS, BETTY BRANT --THORSON, DR. WALTER
##  [7] LEEDS, BETTY BRANT --MAXWELL, MORRIS
##  [8] MAXWELL, MORRIS    --JAMESON, J. JONAH
##  [9] MAXWELL, MORRIS    --DOLMAN
## [10] MAXWELL, MORRIS    --URICH, BEN
## + ... omitted several edges
```

```r
gsize(hero_g2)# Count number of edges
```

```
## [1] 100
```

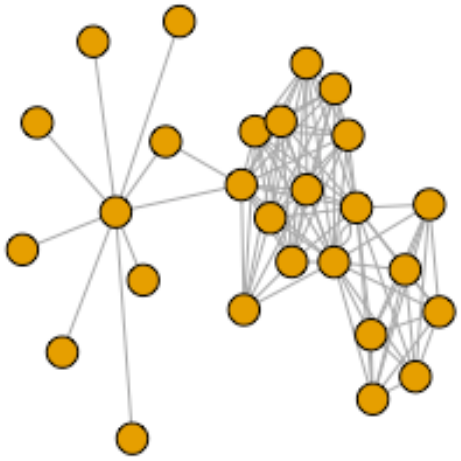
```r
# Measure the size of network
diameter(hero_g2, directed=FALSE, weights=NA) #the length of the longest path
between two nodes is 4
```

```
## [1] 4
```

```r
get_diameter(hero_g2, directed=FALSE, weights=NA) # identify the longest path
```

```
## + 5/27 vertices, named, from c4b7128:
## [1] THOR/DR. DONALD BLAK SPIDER-MAN/PETER PAR ICEMAN/ROBERT BOBBY
## [4] ANGEL/WARREN KENNETH NORRISS, SISTER BARB
```

```r
# Plot social networks

plot(hero_g2, layout = layout_with_lgl(hero_g2), vertex.label=NA)
```



```
# Compute edge_density

edge_density(hero_g2)

## [1] 0.2849003


# Compute mean_distance of graph

mean_distance(hero_g2, directed = FALSE)

## [1] 2.210826


# Compute clustering coefficient to find the probability that the adjacent ve
rtices of a vertex are connected

transitivity(hero_g2, type = "average")

## [1] 0.870511


# Calculate the degree

hero_deg <- degree(hero_g2, mode = c("all"))
which.max(hero_deg)

## SPIDER-MAN/PETER PAR
##
```

```r
# Top most popular

top<-mean(hero_deg)+ 0.8*sd(hero_deg)
length(hero_deg[hero_deg>top])
```

```
## [1] 3
```

```r
hero_deg[hero_deg>top]
```

```
## SPIDER-MAN/PETER PAR ICEMAN/ROBERT BOBBY     JAMESON, J. JONAH
##                   18                   13                   17
```

```r
# Calculate betweenness of each vertex to find the degree of which heroes sta
nd between each other

betw <- betweenness(hero_g2, directed = F)
which.max(betw)
```

```
## ANGEL/WARREN KENNETH
##                   18
```

```r
# Betweeness of top most popular heroes

top<-mean(betw)+ 0.9*sd(betw)
length(betw[betw>top])
```

```
## [1] 3
```

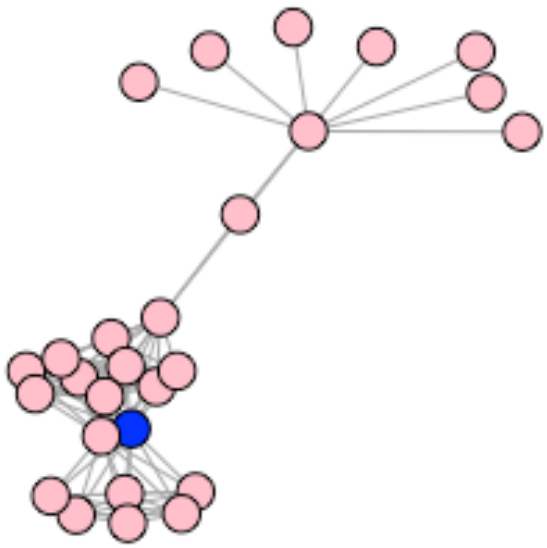
```r
betw[betw>top]
```

```
## SPIDER-MAN/PETER PAR ICEMAN/ROBERT BOBBY  ANGEL/WARREN KENNETH
##             60.67857            153.55357             154.00000
```

```r
# Identify key nodes using eigenvector centrality to measure the influence of
a node in a network

g.ec <- eigen_centrality(hero_g2)
which.max(g.ec$vector)
```

```
## SPIDER-MAN/PETER PAR
##                    4
```

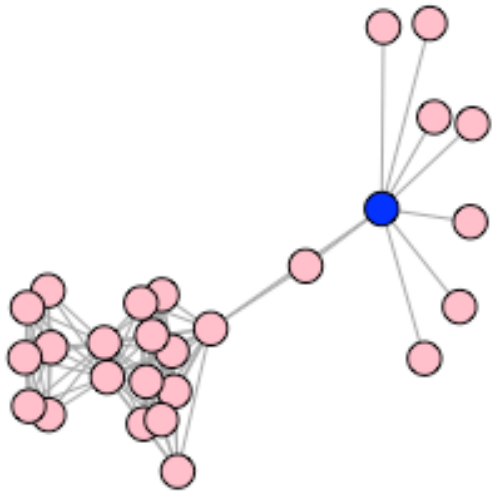
```r
# Measure the influence of top most popular heroes

top<-mean(g.ec$vector)+ 1.2*sd(g.ec$vector)
length(g.ec$vector[g.ec$vector>top])

## [1] 2

g.ec$vector[g.ec$vector>top]

## SPIDER-MAN/PETER PAR    JAMESON, J. JONAH
##            1.0000000            0.9615964



# Spider Man is the most influence character and has most connections in the
network.
# Find who is around Spider Man ?

g_spiderman <- make_ego_graph(hero_g2, diameter(hero_g2), nodes = 'SPIDER-MAN
/PETER PAR', mode = c("all"))[[1]]
V(g_spiderman)$color <- ifelse(V(g_spiderman)$name=="SPIDER-MAN/PETER PAR","b
lue","pink")
plot(g_spiderman, vertex.label=NA)
```



```r
# Neighbors of Spider Man

unique(neighbors(hero_g2, v=which(V(hero_g2)$name=="SPIDER-MAN/PETER PAR")))
```

```
## + 17/27 vertices, named, from c4b7128:
##  [1] LEEDS, BETTY BRANT   MAXWELL, MORRIS      THORSON, DR. WALTER
##  [4] THOMPSON, EUGENE FLA WATSON-PARKER, MARY  ICEMAN/ROBERT BOBBY
##  [7] OVERRIDE/DR. GREGORY KWAN, TERRY          URICH, BEN
## [10] DOLMAN               THOR/DR. DONALD BLAK TOKKOTS
## [13] MCCORMICK, BARRY     JAMESON, J. JONAH    PARKER, MAY
## [16] FAIRMONT, HANNAH     GRANT, GLORIA GLORY
```

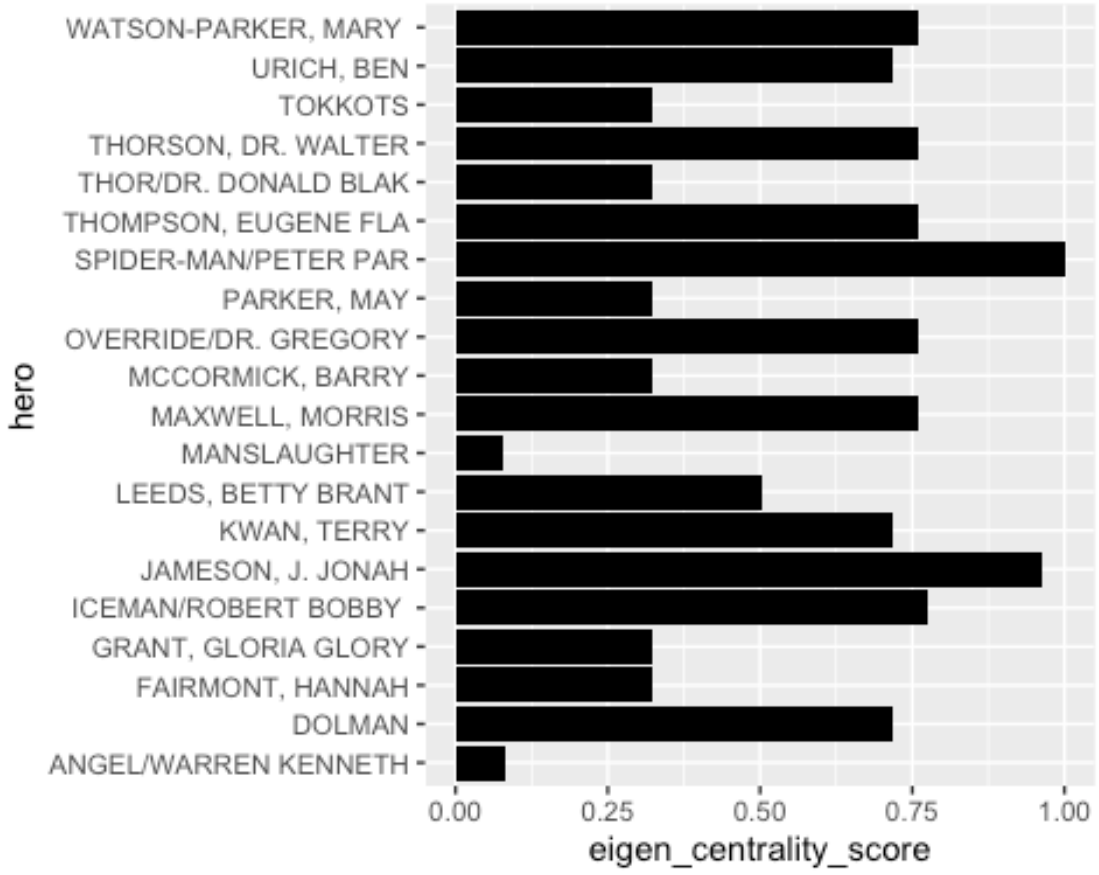
```r
# Warren Kenneth is having the most control over the network
# Find who is around Warren Kenneth ?

g_warren <- make_ego_graph(hero_g2, diameter(hero_g2), nodes = "ANGEL/WARREN
KENNETH", mode = c("all"))[[1]]
V(g_warren)$color <- ifelse(V(g_warren)$name=="ANGEL/WARREN KENNETH","blue","
pink")
plot(g_warren, vertex.label=NA)
```



```r
# Neighbors of Warren Kenneth

unique(neighbors(hero_g2, v=which(V(hero_g2)$name=="ANGEL/WARREN KENNETH")))
```

```
## + 9/27 vertices, named, from c4b7128:
## [1] ICEMAN/ROBERT BOBBY  MANSLAUGHTER         NORRISS, SISTER BARB
## [4] STAR THIEF II        GARGOYLE II/ISAAC CH KUBIK
## [7] CLOUD                BEAST/HENRY &HANK& P ANDROMEDA/ANDROMEDA
```

```r
# Use centrality to summarize which Marvel characteristics have more connecti
ons than others

hero_g_eigen_centrality_people=as.data.frame(eigen_centrality(hero_g2)$vector
)
hero_g_eigen_centrality_people$hero=rownames(hero_g_eigen_centrality_people)
rownames(hero_g_eigen_centrality_people)<-1:nrow(hero_g_eigen_centrality_peop
le)
colnames(hero_g_eigen_centrality_people)<-c("eigen_centrality_score","hero")


# Identify which Marvel characteristics are more important than others in sel
ected first 20 characters

hero_g_eigen_centrality_people_20<-hero_g_eigen_centrality_people[1:20,]



# According to eigen centrality score, Spider Man is the most influence node
within this network

herro_connection <- ggplot(hero_g_eigen_centrality_people_20, aes(x=hero,y=ei
gen_centrality_score))
herro_connection + geom_bar(stat="identity", fill = "#000000")+coord_flip()
```



SUBGROUP ANALYSIS

```
# Identify clusters or communities of nodes in hero network

components(hero_g2) #this network has 1 components
```

```
## $membership
##   LEEDS, BETTY BRANT      MAXWELL, MORRIS  THORSON, DR. WALTER
##                    1                    1                    1
## SPIDER-MAN/PETER PAR THOMPSON, EUGENE FLA WATSON-PARKER, MARY
##                    1                    1                    1
## ICEMAN/ROBERT BOBBY  OVERRIDE/DR. GREGORY          KWAN, TERRY
##                    1                    1                    1
##           URICH, BEN               DOLMAN THOR/DR. DONALD BLAK
##                    1                    1                    1
##              TOKKOTS     MCCORMICK, BARRY    JAMESON, J. JONAH
##                    1                    1                    1
##          PARKER, MAY     FAIRMONT, HANNAH ANGEL/WARREN KENNETH
##                    1                    1                    1
##         MANSLAUGHTER  GRANT, GLORIA GLORY NORRISS, SISTER BARB
##                    1                    1                    1
##        STAR THIEF II GARGOYLE II/ISAAC CH                KUBIK
##                    1                    1                    1
##                CLOUD BEAST/HENRY &HANK& P ANDROMEDA/ANDROMEDA
##                    1                    1                    1
##
## $csize
## [1] 27
##
## $no
## [1] 1
```

```
hero_subgroup2 <- decompose(hero_g2)[[1]]

par(mar=c(0,0,0,0))


V(hero_subgroup2)$color <- ifelse(V(hero_subgroup2)$name=="ANGEL/WARREN KENNE
TH","blue","pink")


plot(hero_subgroup2,cex=0.005)
```

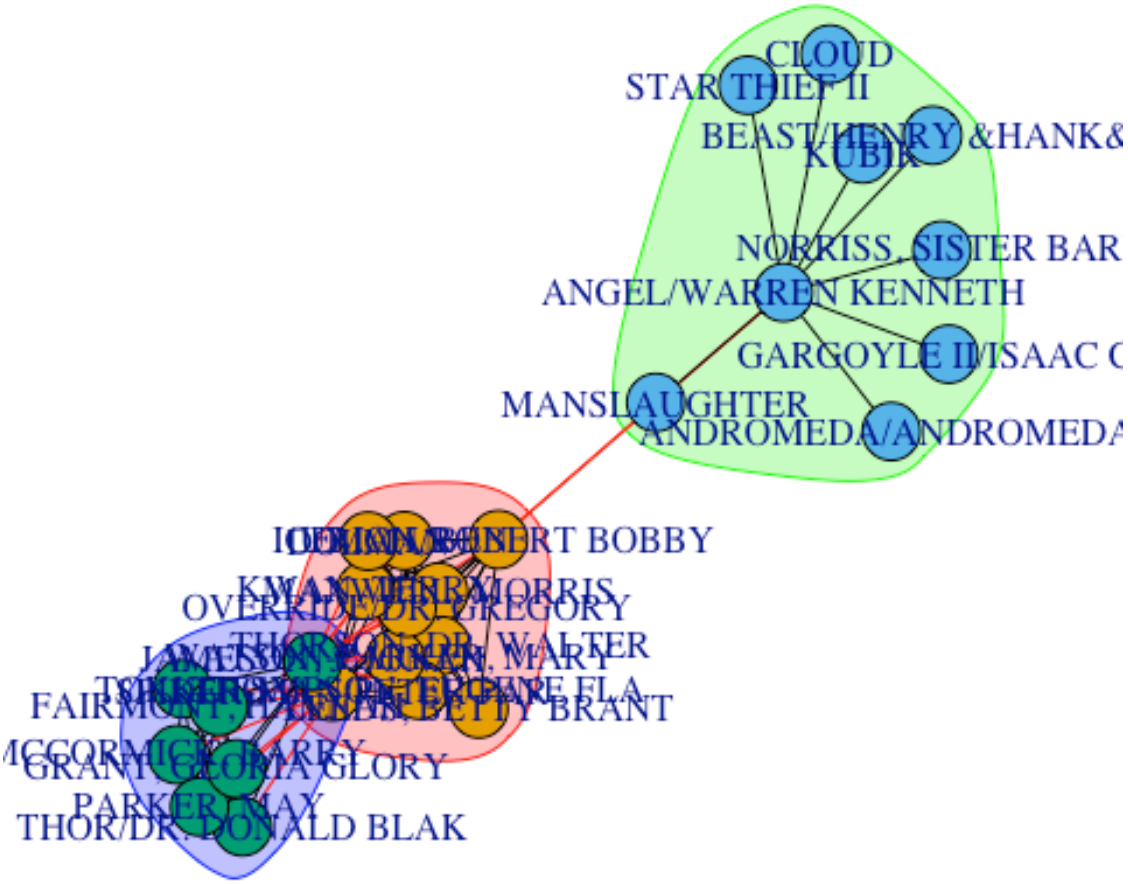

```
cluster_infomap(hero_subgroup2)
```

```
## IGRAPH clustering infomap, groups: 3, mod: 0.34
## + groups:
##   $`1`
##    [1] "LEEDS, BETTY BRANT"   "MAXWELL, MORRIS"      "THORSON, DR. WALTER"
##    [4] "SPIDER-MAN/PETER PAR" "THOMPSON, EUGENE FLA" "WATSON-PARKER, MARY
"
##    [7] "ICEMAN/ROBERT BOBBY " "OVERRIDE/DR. GREGORY" "KWAN, TERRY"
##   [10] "URICH, BEN"           "DOLMAN"
##
##   $`2`
##   [1] "ANGEL/WARREN KENNETH" "MANSLAUGHTER"         "NORRISS, SISTER BARB"
##   [4] "STAR THIEF II"        "GARGOYLE II/ISAAC CH" "KUBIK"
##   [7] "CLOUD"                "BEAST/HENRY &HANK& P" "ANDROMEDA/ANDROMEDA "
##   + ... omitted several groups/vertices
```

```
# Map the flow of information in hero network, and the different clusters in
which information may get remain for longer periods

comm <- cluster_infomap(hero_subgroup2)
modularity(comm) # modularity score
```

```
## [1] 0.33665
```

```
# Plot the resulting communities

par(mar=c(0,0,0,0))
plot(comm, hero_subgroup2,cex=0.0005)
```



Analyze Social Network graph 3

NODE ANALYSIS

```
# Check graph edges and vertices

V(hero_g3) #contents in vertices

## + 17/17 vertices, named, from 05882f2:
##  [1] WARLOCK III          CAPTAIN AMERICA      MAGIK/ILLYANA RASPUT
##  [4] SCARLET WITCH/WANDA  MAGMA/AMARA AQUILLA/ WOLFSBANE/RAHNE SINC
##  [7] CANNONBALL II/SAM GU WASP/JANET VAN DYNE  PHOENIX III/RACHEL S
## [10] PROFESSOR X/CHARLES  SELENE               COLOSSUS II/PETER RA
## [13] CALLISTO             CALIBAN/             HULK/DR. ROBERT BRUC
## [16] ROGUE /              MARKS, DR. SHIELA
```

```r
gorder(hero_g3) # Count number of vertices
```

```
## [1] 17
```

```r
E(hero_g3) #contents in edges
```

```
## + 100/100 edges from 05882f2 (vertex names):
##  [1] WARLOCK III    --PHOENIX III/RACHEL S WARLOCK III    --WASP/JANET VAN
DYNE
##  [3] WARLOCK III    --CANNONBALL II/SAM GU WARLOCK III    --WOLFSBANE/RAHN
E SINC
##  [5] WARLOCK III    --MAGMA/AMARA AQUILLA/ WARLOCK III    --SCARLET WITCH/
WANDA
##  [7] WARLOCK III    --MAGIK/ILLYANA RASPUT WARLOCK III    --CAPTAIN AMERIC
A
##  [9] CAPTAIN AMERICA--ROGUE /              CAPTAIN AMERICA--CALIBAN/
## [11] CAPTAIN AMERICA--CALLISTO             CAPTAIN AMERICA--COLOSSUS II/PE
TER RA
## [13] CAPTAIN AMERICA--SELENE               CAPTAIN AMERICA--PROFESSOR X/CH
ARLES
## [15] CAPTAIN AMERICA--PHOENIX III/RACHEL S CAPTAIN AMERICA--WASP/JANET VAN
DYNE
## [17] CAPTAIN AMERICA--CANNONBALL II/SAM GU CAPTAIN AMERICA--WOLFSBANE/RAHN
E SINC
## [19] CAPTAIN AMERICA--MAGMA/AMARA AQUILLA/ CAPTAIN AMERICA--SCARLET WITCH/
WANDA
## + ... omitted several edges
```

```r
gsize(hero_g3)# Count number of edges
```

```
## [1] 100
```

```r
# Measure the size of network
```

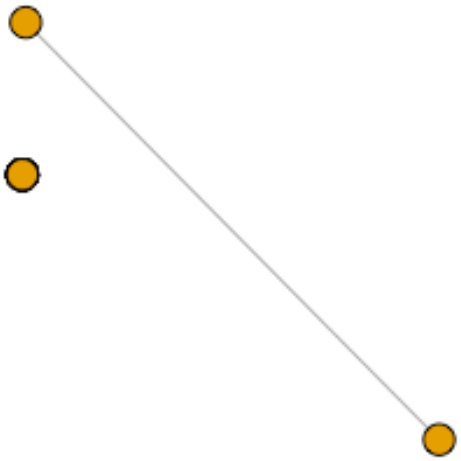
```r
diameter(hero_g3, directed=FALSE, weights=NA) #the length of the longest path
between two nodes is 4
```

```
## [1] 2
```

```r
get_diameter(hero_g3, directed=FALSE, weights=NA) # identify the longest path
```

```
## + 3/17 vertices, named, from 05882f2:
## [1] WARLOCK III          CAPTAIN AMERICA      PROFESSOR X/CHARLES
```

```r
# Plot social networks

plot(hero_g3, layout = layout_with_lgl(hero_g3), vertex.label=NA)
```



```r
# Compute edge_density

edge_density(hero_g3)
```

```
## [1] 0.7352941
```

```r
# Compute mean_distance of graph

mean_distance(hero_g3, directed = FALSE)
```

```
## [1] 1.056604
```

```r
# Compute clustering coefficient to find the probability that the adjacent ve
rtices of a vertex are connected.

transitivity(hero_g3, type = "average")
```

```
## [1] 0.9648352
```

```r
# Calculate the degree

hero_deg <- degree(hero_g3, mode = c("all"))
which.max(hero_deg)
```

```
## CAPTAIN AMERICA
##               2
```

```r
# Top most popular

top<-mean(hero_deg)+ 0.52*sd(hero_deg)
length(hero_deg[hero_deg>top])
```

```
## [1] 8
```

```r
hero_deg[hero_deg>top]
```

```
##      CAPTAIN AMERICA MAGIK/ILLYANA RASPUT SCARLET WITCH/WANDA
##                   14                   14                   14
## MAGMA/AMARA AQUILLA/ WOLFSBANE/RAHNE SINC CANNONBALL II/SAM GU
##                   14                   14                   14
## WASP/JANET VAN DYNE  PHOENIX III/RACHEL S
##                   14                   14
```

```r
# Calculate betweenness of each vertex to find the degree of which heroes sta
nd between each other.

betw <- betweenness(hero_g3, directed = F)
which.max(betw)
```

```
## CAPTAIN AMERICA
##               2
```

```r
# Betweeness of top most popular heroes

top<-mean(betw)+ 0.8*sd(betw)
length(betw[betw>top])
```

```
## [1] 8
```

```r
betw[betw>top]
```

```
##      CAPTAIN AMERICA MAGIK/ILLYANA RASPUT SCARLET WITCH/WANDA
##                 0.75                 0.75                 0.75
## MAGMA/AMARA AQUILLA/ WOLFSBANE/RAHNE SINC CANNONBALL II/SAM GU
##                 0.75                 0.75                 0.75
## WASP/JANET VAN DYNE  PHOENIX III/RACHEL S
##                 0.75                 0.75
```

```r
# Identify key nodes using eigenvector centrality to measure the influence of a node in a network.

g.ec <- eigen_centrality(hero_g3)
which.max(g.ec$vector)
```

```
## SCARLET WITCH/WANDA
##                   4
```

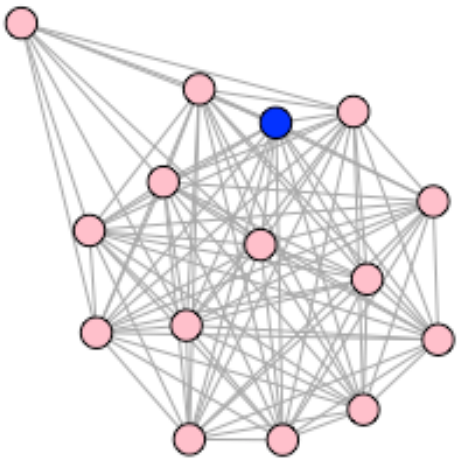
```r
# Measure the influence of top most popular heroes

top<-mean(g.ec$vector)+ 0.45*sd(g.ec$vector)
length(g.ec$vector[g.ec$vector>top])
```

```
## [1] 8
```

```r
g.ec$vector[g.ec$vector>top]
```

```
##      CAPTAIN AMERICA MAGIK/ILLYANA RASPUT SCARLET WITCH/WANDA
##                    1                    1                    1
## MAGMA/AMARA AQUILLA/ WOLFSBANE/RAHNE SINC CANNONBALL II/SAM GU
##                    1                    1                    1
## WASP/JANET VAN DYNE  PHOENIX III/RACHEL S
##                    1                    1
```

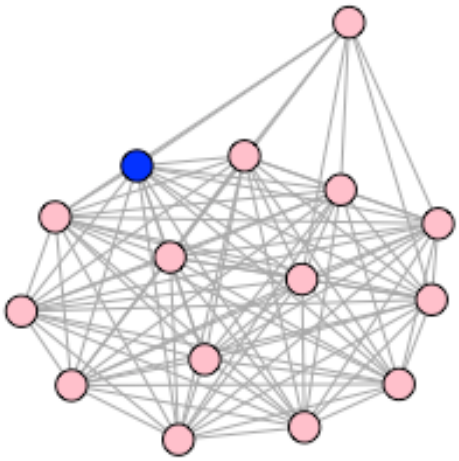
```r
# Captain America has most connections and control in the network.
# Find who is around Captain American ?

g_ca <- make_ego_graph(hero_g3, diameter(hero_g3), nodes = 'CAPTAIN AMERICA', mode = c("all"))[[1]]
V(g_ca)$color <- ifelse(V(g_ca)$name=="CAPTAIN AMERICA","blue","pink")
plot(g_ca, vertex.label=NA)
```



```
# Neighbors of Captain America
unique(neighbors(hero_g3, v=which(V(hero_g3)$name=="CAPTAIN AMERICA")))
```

```
## + 14/17 vertices, named, from 05882f2:
##  [1] WARLOCK III          MAGIK/ILLYANA RASPUT SCARLET WITCH/WANDA
##  [4] MAGMA/AMARA AQUILLA/ WOLFSBANE/RAHNE SINC CANNONBALL II/SAM GU
##  [7] WASP/JANET VAN DYNE  PHOENIX III/RACHEL S PROFESSOR X/CHARLES
## [10] SELENE               COLOSSUS II/PETER RA CALLISTO
## [13] CALIBAN/             ROGUE /
```

```
# Wolfsbane is having the most influence in the network
# Find who is around Wolfsbane?

g_scarlet <- make_ego_graph(hero_g3, diameter(hero_g3), nodes = 'WOLFSBANE/RA
HNE SINC', mode = c("all"))[[1]]
V(g_scarlet)$color <- ifelse(V(g_scarlet)$name=="WOLFSBANE/RAHNE SINC","blue"
,"pink")
plot(g_scarlet, vertex.label=NA)
```

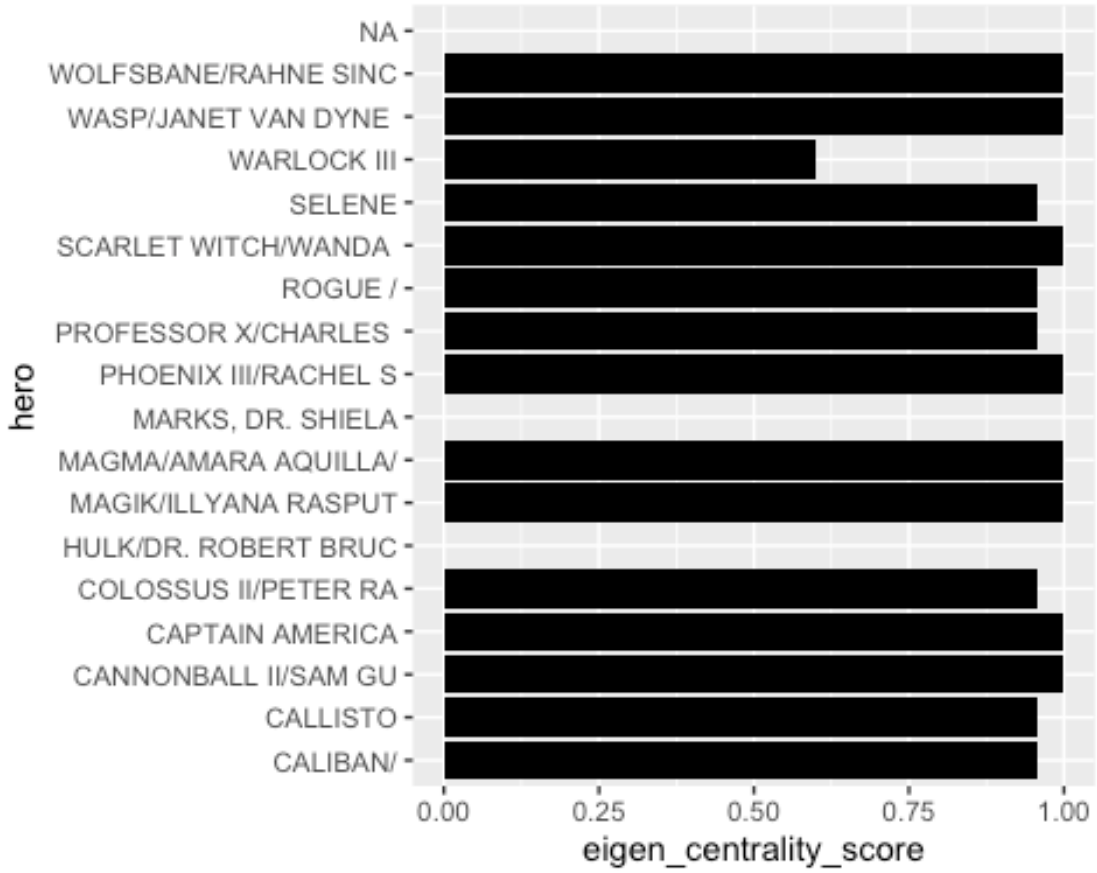


```r
# Neighbors of Wolfsbane

unique(neighbors(hero_g3, v=which(V(hero_g3)$name=="WOLFSBANE/RAHNE SINC")))
```

```
## + 14/17 vertices, named, from 05882f2:
##  [1] WARLOCK III          CAPTAIN AMERICA      MAGIK/ILLYANA RASPUT
##  [4] SCARLET WITCH/WANDA  MAGMA/AMARA AQUILLA/ CANNONBALL II/SAM GU
##  [7] WASP/JANET VAN DYNE  PHOENIX III/RACHEL S PROFESSOR X/CHARLES
## [10] SELENE               COLOSSUS II/PETER RA CALLISTO
## [13] CALIBAN/             ROGUE /
```

```r
# Use centrality to summarize which Marvel characteristics have more connecti
ons than others

hero_g_eigen_centrality_people=as.data.frame(eigen_centrality(hero_g3)$vector
)
hero_g_eigen_centrality_people$hero=rownames(hero_g_eigen_centrality_people)
rownames(hero_g_eigen_centrality_people)<-1:nrow(hero_g_eigen_centrality_peop
le)


# Identify which Marvel characteristics are more important than others in sel
ected first 20 characters

colnames(hero_g_eigen_centrality_people)<-c("eigen_centrality_score","hero")
hero_g_eigen_centrality_people_20<-hero_g_eigen_centrality_people[1:20,]



# According to eigen centrality score, Captain America is one of the most inf
luence node within this network

herro_connection <- ggplot(hero_g_eigen_centrality_people_20, aes(x=hero,y=ei
gen_centrality_score))
herro_connection + geom_bar(stat="identity", fill = "#000000")+coord_flip()
```

```
## Warning: Removed 3 rows containing missing values (position_stack).
```



SUBGROUP ANALYSIS

```
# Identify clusters or communities of nodes in hero network

components(hero_g3) #this network has 2 components
```

```
## $membership
##          WARLOCK III      CAPTAIN AMERICA MAGIK/ILLYANA RASPUT
##                    1                    1                    1
## SCARLET WITCH/WANDA  MAGMA/AMARA AQUILLA/ WOLFSBANE/RAHNE SINC
##                    1                    1                    1
## CANNONBALL II/SAM GU WASP/JANET VAN DYNE  PHOENIX III/RACHEL S
##                    1                    1                    1
## PROFESSOR X/CHARLES                SELENE COLOSSUS II/PETER RA
##                    1                    1                    1
##             CALLISTO             CALIBAN/ HULK/DR. ROBERT BRUC
##                    1                    1                    2
##              ROGUE /    MARKS, DR. SHIELA
##                    1                    2
##
```

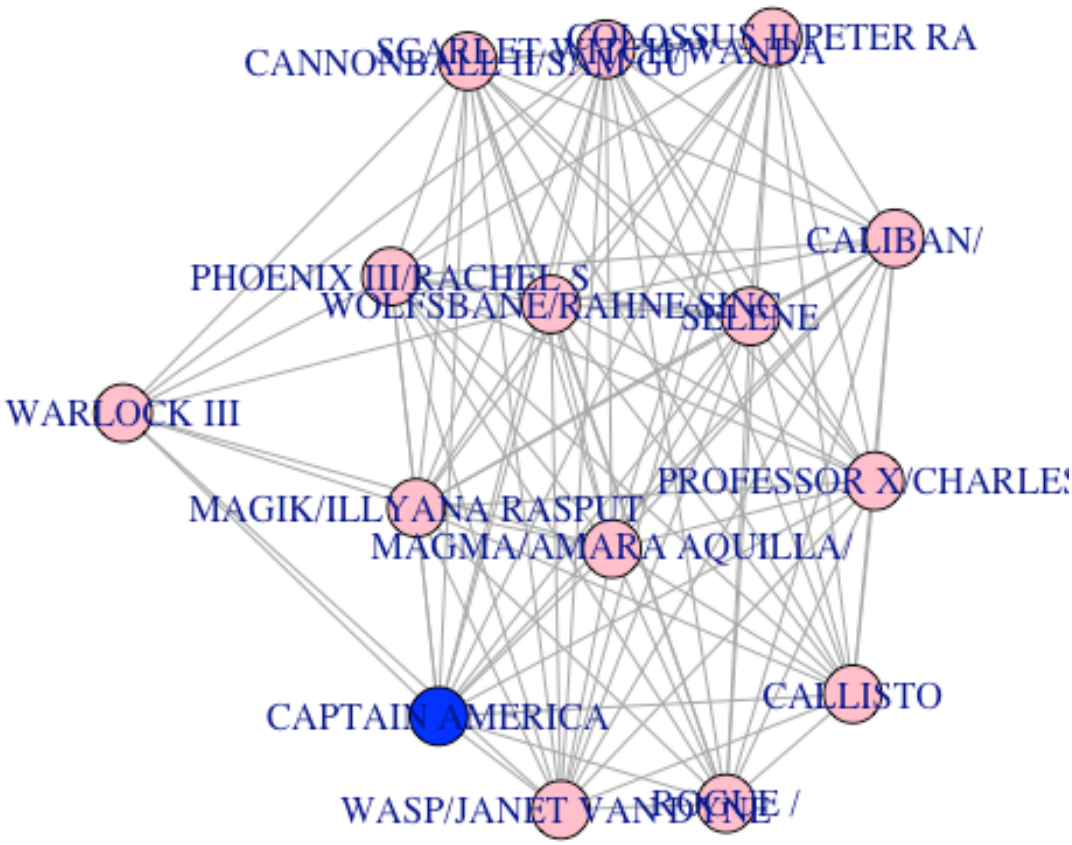
```
## $csize
## [1] 15  2
##
## $no
## [1] 2
```

```
# We will analyze the component 1 which have the largest size

hero_subgroup3 <- decompose(hero_g3)[[1]]
```

```
par(mar=c(0,0,0,0))
```

```
V(hero_subgroup3)$color <- ifelse(V(hero_subgroup3)$name=="CAPTAIN AMERICA","
blue","pink")
```

```
plot(hero_subgroup3,cex=0.05)
```



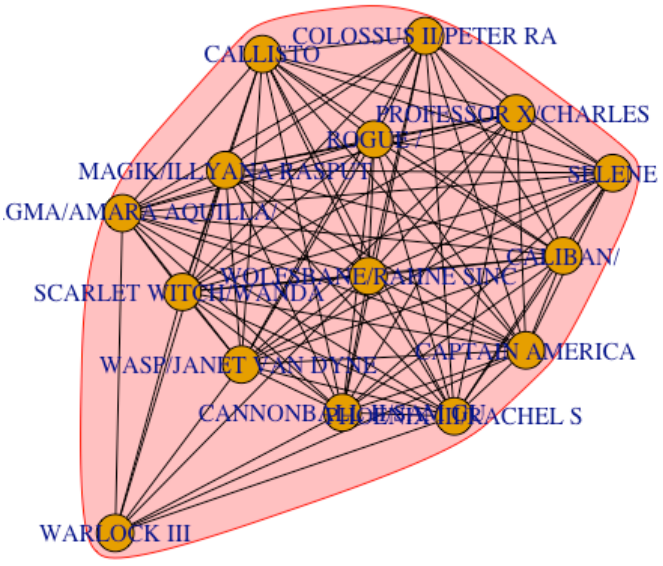
```
cluster_infomap(hero_subgroup3)
```

```
## IGRAPH clustering infomap, groups: 1, mod: 0
## + groups:
##   $`1`
##    [1] "WARLOCK III"          "CAPTAIN AMERICA"      "MAGIK/ILLYANA RASPUT
"
##    [4] "SCARLET WITCH/WANDA " "MAGMA/AMARA AQUILLA/" "WOLFSBANE/RAHNE SINC
"
##    [7] "CANNONBALL II/SAM GU" "WASP/JANET VAN DYNE " "PHOENIX III/RACHEL S
"
##   [10] "PROFESSOR X/CHARLES " "SELENE"               "COLOSSUS II/PETER RA
"
##   [13] "CALLISTO"             "CALIBAN/"             "ROGUE /"
##
```

```
# Map the flow of information in hero network, and the different clusters in
which information may get remain for longer periods

comm <- cluster_infomap(hero_subgroup3)
modularity(comm) # modularity score
```

```
## [1] 0
```

```
# Plot the resulting communities

par(mar=c(0,0,0,0))
plot(comm, hero_subgroup3,cex=0.05)
```



CONCLUSION:

Top 5 most popular characters of this social network are Spider man, Captain America, Iron man, Thing, Thor and Human Torch respectively.

- I selected random 3 subsets of this social network data to analyze the relationship between characters and looking for if there is any pattern between these subset datasets.
- In the Graph (1), Sir Denis Nayland Smith is the character having the most connections and control over the network (1).Black Panther is the most influence character.There are 2 different clusters in this network.
- In the Graph (2), Spider Man is the most influence character and has most connections in the network (2). Warren Kenneth is having the most control over the network. There are 3 different clusters in this network.
- In the Graph (3), Captain America is the character having the most connections and control over the network (3). Wolfsbane is the most influence character. This network only contain 1 community.
- We noticed that in 3 different subsets, we could see the appearances of top popular characters in the network such as Captain America, Spider Man, Iron Man, which means they the more popular of the characters, the more relationship they have with other heroes in the network.