Homework 3 (CSC 495)

Kai Chung, Ying

04-18-2017

```
library(knitr)
setwd("/Users/KevQuant/Desktop/Depaul/csc495/wk3/hwk3")
read_chunk("hwk3_2.R")
knitr::opts_chunk$set(echo = TRUE)
```

Centrality measures on networks

Step 0: Load the libraries

```
library("ggplot2")
library("GGally")
# Must load other packages first
library("sand")
library("intergraph")
```

Graded portion *(17 points)*

Step 1: Load the dolphin data *(1 pt)*

```
#Setting up working directory
setwd("/Users/KevQuant/Desktop/Depaul/csc495/wk3/hwk3")
dolphin <- read.graph("dolphin.graphml", format="graphml")
```
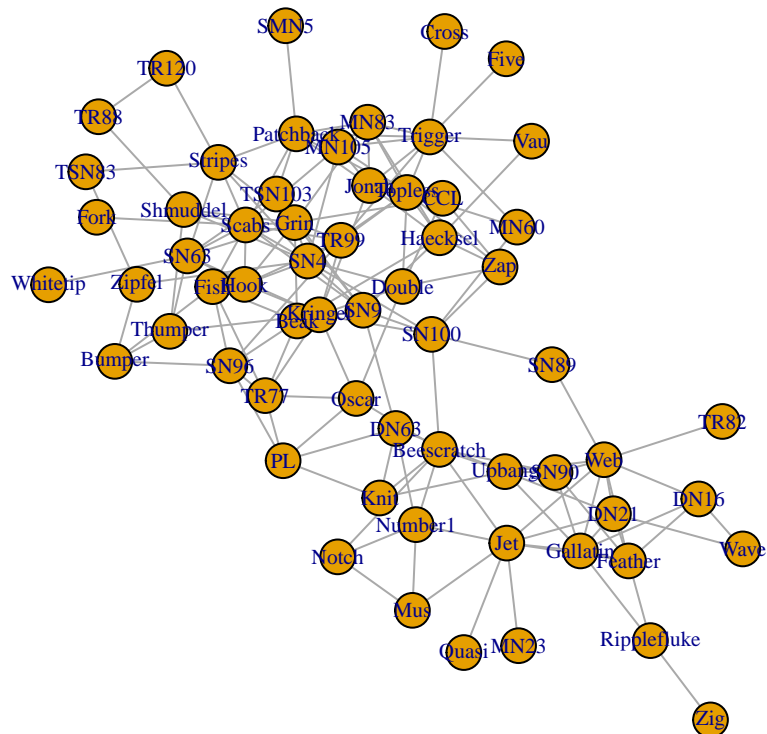
Step 2: Plot the network *(1 pt)*

Compute a Kamada-Kawai layout (`layout_with_kk`) in a separate step so that we can create comparable visualizations later.

```
par(mar=c(1,1,1,1))
plot(dolphin,vertex.label.cex=0.7)
```
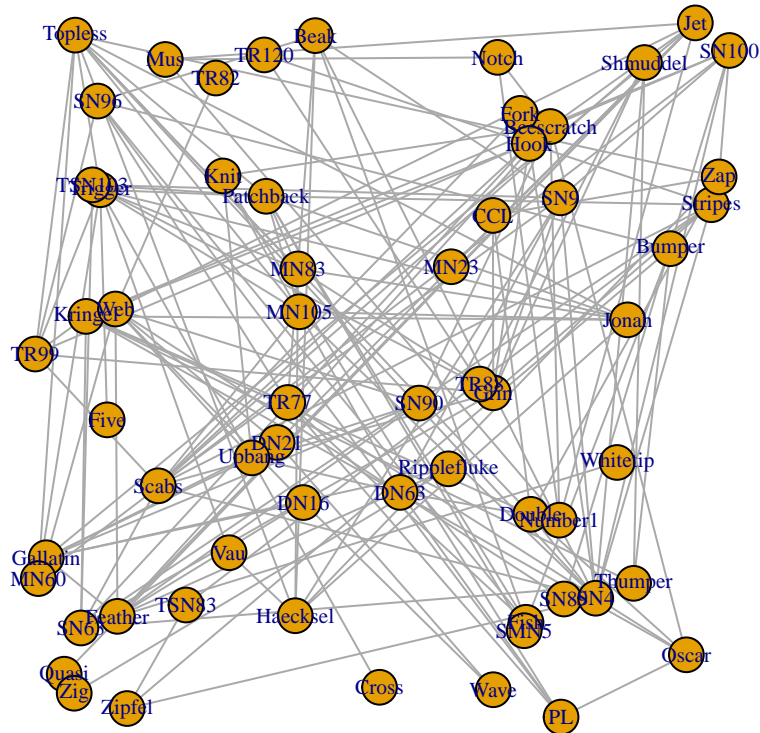
```
plot(dolphin,layout=layout_with_kk,vertex.label.cex=0.7)
```



```
summary(dolphin)
```

```
## IGRAPH U-W- 62 159 --
## + attr: label (v/c), r (v/n), g (v/n), b (v/n), x (v/n), y (v/n),
## | size (v/n), Sex (v/n), id (v/c), Edge Label (e/c), weight (e/n),
```

```
## | Edge Id (e/c)
```
```r
diameter(dolphin)
```
```
## [1] 8
```

Step 3: Compute 5 types of centrality and create a data frame *(2 pts)*

Degree, betweenness, closeness, eigenvector, page rank

Also include a column for the Sex attribute.

```r
dol.deg<-degree(dolphin)
dol.bet<-betweenness(dolphin, normalized = TRUE)
dol.clo<-closeness(dolphin)
dol.eig<-eigen_centrality(dolphin)$vector
dol.pgr<-page_rank(dolphin)$vector

#Check their vector length
#length(dol.deg)
#length(dol.bet)
#length(dol.clo)
#length(dol.eig)
#length(dol.pgr)
#length(V(dolphin)$Sex)

dol.df<-data.frame(Deg=dol.deg,
                   Bet=dol.bet,
                   Clo=dol.clo,
                   Eig=dol.eig,
                   Pgr=dol.pgr)
dol.df<-cbind(dol.df,Sex=V(dolphin)$Sex)
head(dol.df)
```
```
Deg   Bet   Clo   Eig   Pgr Sex
1   6 0.019082596 0.005681818 0.40693637 0.016965392   1
2   8 0.213324436 0.006097561 0.13324392 0.024650717   1
3   4 0.009072812 0.004629630 0.12590018 0.013338080   1
4   3 0.002373797 0.005050505 0.25123200 0.009629062   0
5   1 0.000000000 0.004081633 0.09274536 0.005079800   1
6   4 0.004380300 0.003906250 0.02077141 0.014428048   0
```

**Step 4: Display correlation plot via ggcorr** *(1 pt)*

You'll have to omit the sex attribute. Use `label=TRUE` to show the correlation values.

```r
ggcorr(dol.df[1:5],label = TRUE)
```
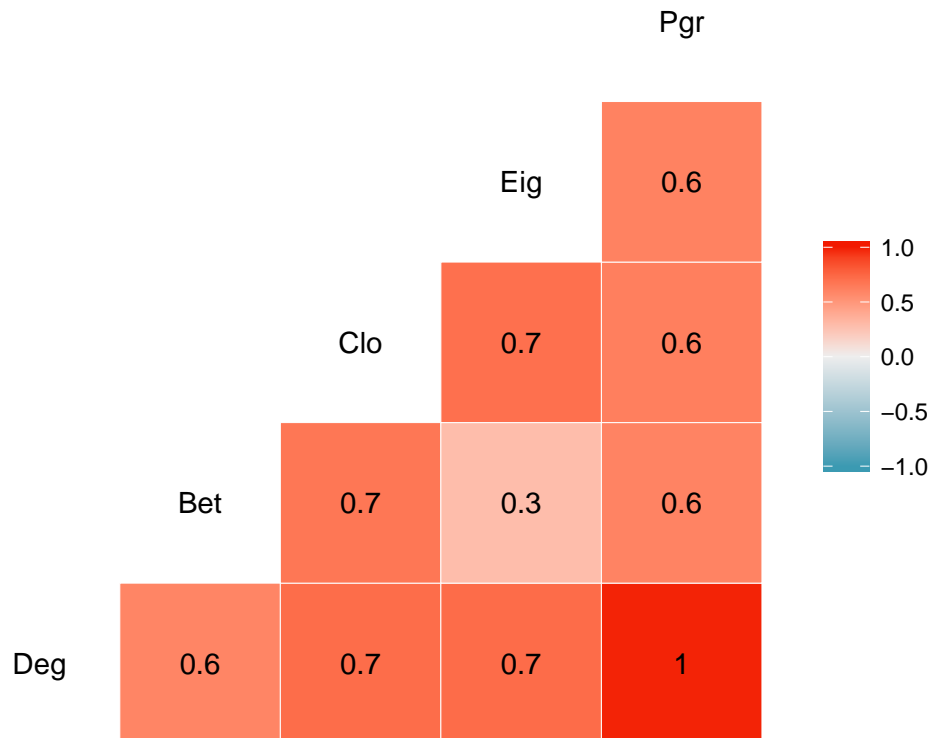
Step 5: Plot two versions of the network side by side *(2 pts)*

In the first, size the nodes by eigenvector centrality. In the second, size the nodes by betweenness centrality. Scale the values so the largest and smallest nodes are approximately the same size in each plot. Make sure the minimum size is not too small.

```r
#Divide the print window into 2 section and reset the plotting margin
par(mfrow=c(1,2))
par(mar=c(0.5,0.5,0.5,0.5))

#Rescale the value of eigenvalues for plotting purpose
dol.eig2<-dol.eig
min(dol.eig2)
```

```
## [1] 0.001697438
```

```r
max(dol.eig2)
```

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

```r
dol.eig2[dol.eig2<=0.3]<-0.3
dol.eig2[dol.eig2>0.5]<-0.5
plot(dolphin,
     vertex.size=dol.eig2*25,
     layout=layout_with_kk,
     vertex.label.cex=0.6)

#Rescale the value of betweenness for plotting purpose
dol.bet2<-dol.bet
min(dol.bet2)
```

4

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

```
max(dol.bet2)
```

```
## [1] 0.2482372
```

```
dol.bet2[dol.bet2<=0.1]<-0.1
dol.bet2[dol.bet2>0.2]<-0.2
plot(dolphin,
     vertex.size=dol.bet2*65,
     layout=layout_with_kk,
     vertex.label.cex=0.6)
```



```
par(mfrow=c(1,1))
```

Step 6: Interpretation *(5 pts)*

Question: Using the evidence from the correlation plot, the network visualizations, and our discussions in class, what can you say about the nodes with high betweenness in this network? What can you say about the nodes with high eigenvector centrality? Why are these measures only weakly correlated?

Answer: Regarding to the high betweenness values, we could conclude that the node with high betweenness values which are in the roles bridge out the nodes with dense connections, so that making most of the nodes needs to pass those nodes with high Betweenness in order to reach other side of the nodes.

Regarding ot the high eigenvector centrality values, we could conclude that those nodes most likely connect to other nodes with High Degree values which makes itself with high eigenvector centrality values.

Based on the the correlation graph, the Betweenness value is NOT quite correlated with the Eigenvalue values (corr=0.3). It might due to most of the node with High Betweenness value connects to the other nodes possessing very few connections (Or NOT connect to Network Center) which means those connected nodes are very isolated (Low Degree).So that even though the nodes high betweenness, but the Eigenvector Centrality value is still Low. So there these 2 values are weakly correlated.

Step 7: Plot betweenness distribution by Sex *(2 pts)*

Note that to use `fill=`, you will have to change the sex attribute into a factor. igraph does not support factors as node attributes. You will need to tune the `binwidth` parameter to get a good sense of what is

going on.

Use `scale_fill_manual` to create an appropriate legend. See included PNG file for an example.

Add a density estimate overlay (alpha=0.3). Set the bandwith of the estimator equal to the binwidth.

```
dol.df$Sex<-factor(dol.df$Sex,c(0,1,2),labels = c("Female","Male","Unknown"))
dol.bet.Sex.df<-data.frame(Bet=dol.df$Bet,Sex=dol.df$Sex)
g2<-ggplot(dol.bet.Sex.df,aes(x=Bet,fill=Sex))
g2<-g2+geom_histogram(binwidth = 0.05,
                      position="dodge",
                      aes(y=..count..))
g2<-g2+ scale_fill_manual(values = c("Female"="red","Male"="blue","Unknown"="green"))
g2<-g2+geom_density(alpha=0.3,bw=0.05)
g2<-g2+scale_x_continuous(breaks=seq(0,0.25,0.05))
g2<-g2+scale_y_continuous(breaks=seq(0,20,5))
g2<-g2+ggtitle("Betweenness Histogram with Sex Attribute")
g2<-g2+xlab("Betweenness")
g2<-g2+ylab("Count")
print(g2)
```
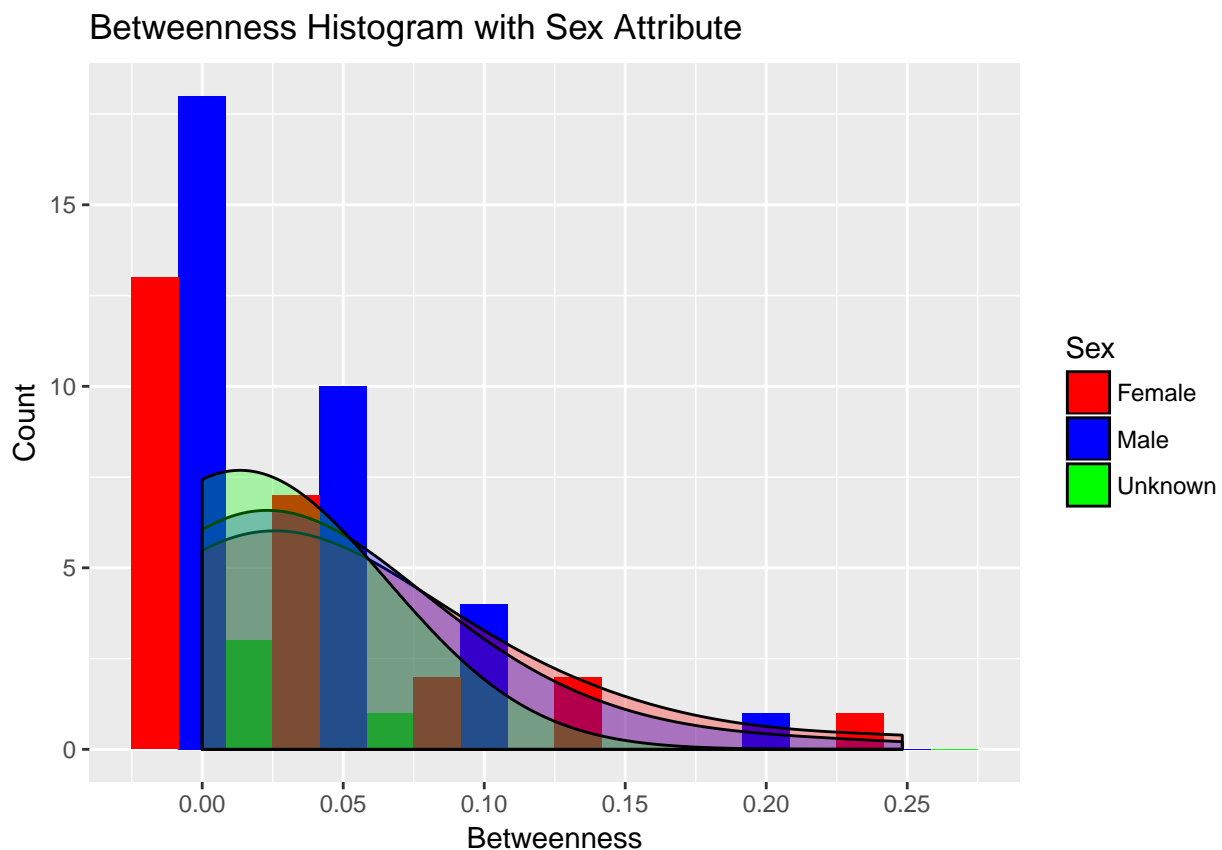

**Step 8: Plot eigenvector centrality distribution by Sex.** *(2 pts)*

Add a density estimator as in Step 7.

```
dol.eig.Sex.df<-data.frame(eig=dol.df$Eig,Sex=dol.df$Sex)
g3<-ggplot(dol.eig.Sex.df,aes(x=eig, fill=Sex))
g3<-g3+geom_histogram(binwidth = 0.2,
```

```
                          position="dodge")
g3<-g3+scale_fill_manual(values = c("Female"="red","Male"="blue","Unknown"="green"))
g3<-g3+geom_density(alpha=0.3,bw=0.2)
g3<-g3+ggtitle("Eigenvalues centrality distribution by Sex")
g3<-g3+xlab("Eigenvalue Centrality")
g3<-g3+ylab("Count")
g3<-g3+scale_x_continuous(breaks = seq(0,1.2,0.2))
print(g3)
```


Eigenvalues centrality distribution by Sex
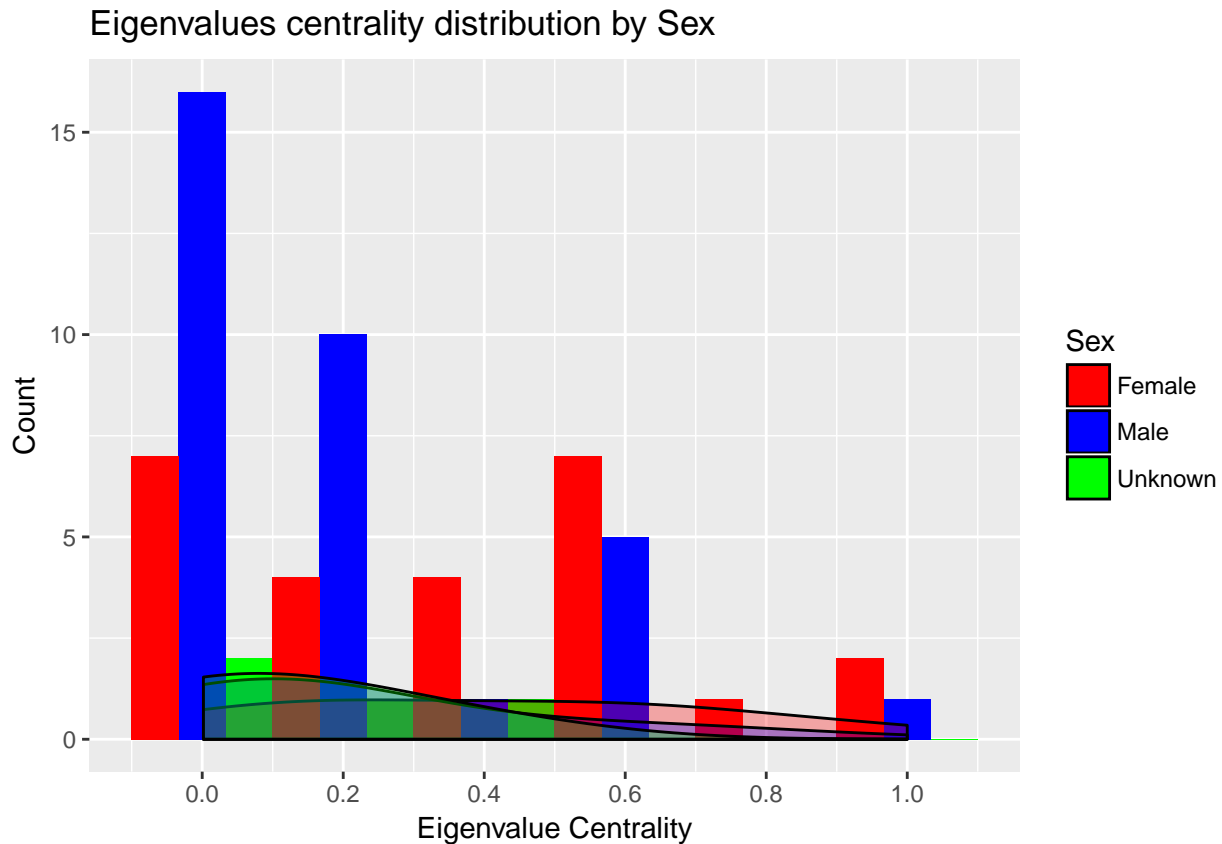
Step 9: If forced to choose, would you identify the "Unknown" group of dolphins as most likely male or female? Why? (1 pt)

Answer: Refering to the Betweenness and Eigenvalues graph, we could observe that Female dolphin tends to have greater values of these 2 Centralities. Male and Unknown Categories seems have less values. For this reason, there is one of the evidence I could have a good chance to classify the Unkown group mostly likely beloings to Male.