

**School of Social and Political Science – PG Feedback Form**

*PLEASE NOTE: This form* ***must*** *be completed and attached to the front of the submitted file prior to upload via ELMA. Failure to do so may result in a mark penalty.*

* *The coursework submitted* ***must*** *be the* ***final*** *version. There is no option to re-submit/make subsequent changes.*
* *All marks and penalties are provisional until ratified by the Board of Examiners in June.*

*To be completed by the student:*

|  |  |  |  |
| --- | --- | --- | --- |
| **Exam number** | B167501 | **Course code** | PGSP 11452 |
| **Course name** | Analysing Social Networks with Statistics | | |
| **Component name** | Practical Exercise 6 | **Session** | **Semester 2, 19/20** |
| **Marker** | (leave blank) | **Word Count** | **603** |

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

*For office use:*

|  |  |
| --- | --- |
| **Initial mark** |  |
| **Word count penalties** |  |
| **Lateness penalties** |  |
| **Adjusted provisional mark** |  |

**Marker’s comments**

*In line with the School’s* [*PG Marking Scheme*](http://www.sps.ed.ac.uk/gradschool/current_students/taught_msc_students/pg_marking_scheme)*, the comments might reflect on the quality of analysis, critical engagement with key concepts, structure and organisation, overall strength and cohesiveness of the argument, use of sources and evidence, breadth and relevance of reading, clarity of expression, presentation and referencing.*

|  |
| --- |
|  |

SNA\_prac6

Yidi Yan

04/03/2020

## Question (a)

## Network attributes:  
## vertices = 36  
## directed = FALSE  
## hyper = FALSE  
## loops = FALSE  
## multiple = FALSE  
## bipartite = FALSE  
## total edges = 115   
## missing edges = 0   
## non-missing edges = 115   
## density = 0.1825397   
##   
## Vertex attributes:  
##   
## age:  
## integer valued attribute  
## 36 values  
##   
## indegree:  
## numeric valued attribute  
## attribute summary:  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.000 3.000 6.000 6.389 9.000 15.000   
##   
## name:  
## integer valued attribute  
## 36 values  
##   
## office:  
## integer valued attribute  
## 36 values  
##   
## practice:  
## integer valued attribute  
## 36 values  
##   
## school:  
## integer valued attribute  
## 36 values  
##   
## senior:  
## integer valued attribute  
## 36 values  
##   
## sex:  
## integer valued attribute  
## 36 values  
##   
## status:  
## integer valued attribute  
## 36 values  
## vertex.names:  
## character valued attribute  
## 36 valid vertex names  
##   
## years:  
## integer valued attribute  
## 36 values  
##   
## No edge attributes

In the Lazega network, the data is undirected and binary. The network data has eight node attributes and all attributes are integer valued. The attributes name and seniority is identical for each person ranging from 1 to 36; sex and practice are coded as 1 and 2; office and school are classified as 1,2,3; years is integer ranging from 1 to 32; and status is 1 for all nodes. From the summary function, there are 115 edges and the overall density is 0.1825397.

## Question (b)

## 

## Question (c)

The density of the network is 0.1825, indicates that the network is not well-connected. From the first graph, lawyers in the network are not well-connected and two lawyers who are not connected with any other lawyers in the network. Three female lawyers are structured closely even though not all three female lawyers are connected to each other. There is not a clear pattern that the network has any structure in relation to the law school they went to. The node label indicates that lawyers are more connected within the same office. The node size indicates that lawyers with different age groups are well mixed.

From the second graph, still, three female lawyers are structured to be close to each other. The square shape nodes are centred in the network shows that lawyers working in the same practice are better connected. The size of the nodes indicates the years the lawyer has been working within the firm, and the graph shows that some nodes with smaller size are clusted into two groups.

In conclusion, collaboration is relatively structured by the lawyers’ attributes, especially their gender, office, years working within the firm and types of practice.

## Question (d)

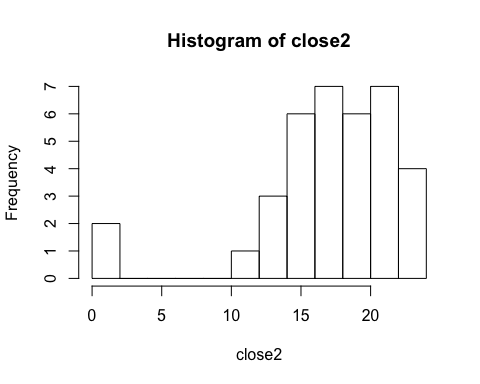
##   
## statnet: version 2019.6, created on 2019-06-13  
## Copyright (c) 2019, Mark S. Handcock, University of California -- Los Angeles  
## David R. Hunter, Penn State University  
## Carter T. Butts, University of California -- Irvine  
## Steven M. Goodreau, University of Washington  
## Pavel N. Krivitsky, University of Wollongong  
## Skye Bender-deMoll  
## Martina Morris, University of Washington  
## Based on "statnet" project software (statnet.org).  
## For license and citation information see statnet.org/attribution  
## or type citation("statnet").

## unable to reach CRAN

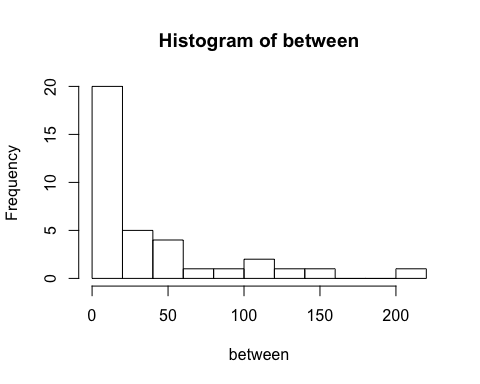
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00000 0.01304 0.02609 0.02778 0.03913 0.06522

## [1] 0.2530612

## [1] 13.75000 18.16667 14.61667 20.50000 17.75000 17.08333 14.25000 0.00000  
## [9] 15.00000 17.83333 13.75000 19.50000 13.70000 18.66667 20.83333 22.66667  
## [17] 23.50000 18.75000 20.83333 16.00000 10.73333 20.66667 0.00000 20.33333  
## [25] 17.41667 22.00000 15.50000 22.25000 19.66667 16.08333 22.08333 22.00000  
## [33] 16.41667 17.83333 18.91667 15.91667



## [1] 0.0000000 36.8857143 1.3888889 42.6140859 7.0897547 4.1698135  
## [7] 3.7095238 0.0000000 0.0000000 16.1947913 0.0000000 16.8605145  
## [13] 0.0000000 26.5583361 44.0896104 141.2988928 203.7686119 42.5324176  
## [19] 28.1733489 0.7166667 0.0000000 36.8639610 0.0000000 57.2120213  
## [25] 13.8249084 97.4550144 64.0000000 108.8873515 26.4872960 2.0151515  
## [31] 125.3248363 104.3455156 7.9666667 4.9303030 17.9541847 0.6818182



##   
## Attaching package: 'igraph'

## The following objects are masked from 'package:sna':  
##   
## betweenness, bonpow, closeness, components, degree, dyad.census,  
## evcent, hierarchy, is.connected, neighborhood, triad.census

## The following objects are masked from 'package:network':  
##   
## %c%, %s%, add.edges, add.vertices, delete.edges, delete.vertices,  
## get.edge.attribute, get.edges, get.vertex.attribute, is.bipartite,  
## is.directed, list.edge.attributes, list.vertex.attributes,  
## set.edge.attribute, set.vertex.attribute

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

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

## 1 2 3 4 5 6   
## 0.009917325 -1.440711010 -0.642983830 -0.829033809 -1.323180001 -0.973891310   
## 7 8 9 10 11 12   
## -1.189122998 0.000000000 -0.251688611 -1.614440399 0.009917325 -0.894411995   
## 13 14 15 16 17 18   
## 0.547118618 -0.916511046 -0.072539224 -1.093578715 -0.824393916 -1.417034469   
## 19 20 21 22 23 24   
## -1.067045905 -0.137179273 -1.006107650 -0.329184955 0.000000000 -0.794603676   
## 25 26 27 28 29 30   
## -0.380021969 -2.243666248 -1.840418890 -1.348861114 -0.766631623 -0.628478454   
## 31 32 33 34 35 36   
## -0.726459946 -1.647044323 -0.395043917 -1.254635110 -0.818828265 0.542212107

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -2.2437 -1.2055 -0.8216 -0.7711 -0.3098 0.5471

## [1] 4

**Node centrality** represents the direct connections one node has divided by the maximum possible connections. The maximum degree centrality in this network is 0.06522 and the mean degree centrality is only 0.02778, which is very low.

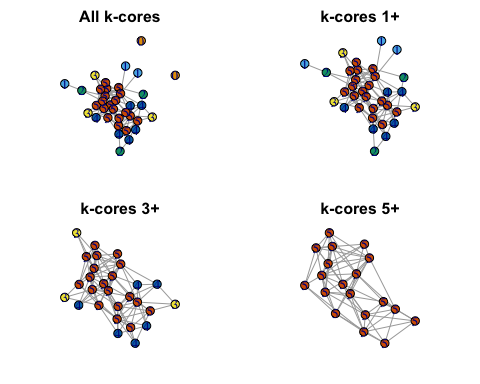
The **network centralization** is 0.2530612, is much higher than the node centrality.

**Closeness centrality** represents the distance of one node to other nodes. The histogram shows that except for two nodes which are isolated, other nodes have similar closeness.

**Betweenness centrality** represents the geopath between the node and other nodes. The histogram shows that the majority of nodes have small betweenness while some nodes have very large betweenness.

The **Bonacich Power Centrality** and its summary are shown above. Positive Bonacich power centrality indicates that the node’s connections have many connections, and it’s increasing the node’s power. Negative Bonacich power centrality indicates that the node’s connections have few connections, and it’s decreasing the node’s power. The Bonacich power centrality in this network ranges from -2.24 to 0.55, and on average, people in this network have negative Bonacich power centrality. Only 4 people out of 36 have positive Bonacich power centrality and they are considered as powerful people in this group.

## Question (e)



K-core analysis is a decomposition of the whole group into subgroups, uncovering the network’s hierarchical and structural properties.

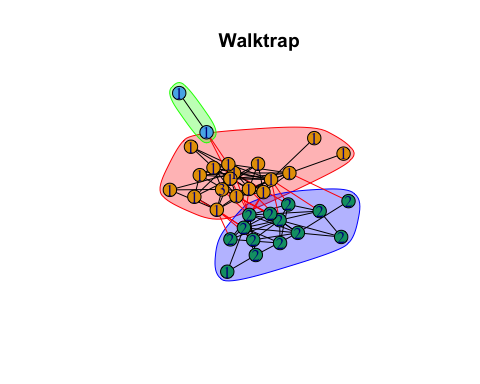
## Question (f)

## Sex Office Years Age Practice  
## Modularity 0.005141777 0.267637051 -0.019470699 -0.031190926 0.118676749  
## School  
## Modularity -0.023969754

The modularity statistic represents the extent to which a node grouping variable explains the observed clustering. The modularity statistic can range from -1/2 to +1. The closer to 1, the more the network clustering follows the given node grouping. The table above shows the modularities using six node attributes. From the table, the modularity using Office is closest to 1, thus which office does the lawyer work is the best predictor of the network clustering.

## Question (g)

## Walktrap Edge Betweenness Fastgreedy  
## Modularity 0.3160303 0.2896786 0.3125520



The walktrap community detection algorithms uses network data without two isolates, the edge betweenness and fast-greedy algorithms use the orginal netowrk data. From the modularity table, using walktrap algorithms shows the highest modularity, thus it’s the best solution.

## Appendix

library(statnet)

load("data\_lab6.RData")

##############################################################

### Question (a)

class(elist.lazega.df) #edge list

class(v.attr.lazega) #node attributes

#transform edge list to matrix

mat.lazega.df <- data.matrix(elist.lazega.df)

class(mat.lazega.df)

lazega.net <- network(mat.lazega.df, matrix.type = "edgelist", directed = FALSE, names.eval = "strength")

class(lazega.net)

summary(lazega.net, print.adj = FALSE)

plot(lazega.net)

as.sociomatrix(lazega.net,"strength") #binary adjacency matrix

plot(lazega.net, displaylabels = T)

#adding attributes

set.vertex.attribute(lazega.net, "name", c(1:36))

lazega.net %v% "sex" <- v.attr.lazega$Gender

lazega.net %v% "senior" <- v.attr.lazega$Seniority

lazega.net %v% "status" <- v.attr.lazega$Status

lazega.net %v% "office" <- v.attr.lazega$Office

lazega.net %v% "years" <- v.attr.lazega$Years

lazega.net %v% "age" <- v.attr.lazega$Age

lazega.net %v% "practice" <- v.attr.lazega$Practice

lazega.net %v% "school" <- v.attr.lazega$School

lazega.net %v% "indegree" <- degree(lazega.net, cmode = "indegree")

list.vertex.attributes(lazega.net)

get.vertex.attribute(lazega.net, "vertex.names")

library(igraph)

inet <- graph.edgelist(mat.lazega.df)

class(inet)

V(inet)$name <- paste("V",c(1:36))

E(inet)$val <- c(1:115)

summary(inet)

library(intergraph)

ilazega.net <- asIgraph(lazega.net)

class(ilazega.net)

library(netCoin)

V(ilazega.net)$name <- V(ilazega.net)$vertex.names

visualize.broswer <- function(graph){

nodes <- as.data.frame(vertex\_attr(graph, "name"))

colnames(nodes) <- "name"

nodes$name <- as.character(nodes$name)

nodes$indegree <- degree(graph, mode = "in")

nodes$outdegree <- degree(graph, mode = "out")

net <- allNet(incidence=as\_adjacency\_matrix(graph),nodes=nodes)

plot(net)

}

visualize.broswer(ilazega.net)

##############################################################

### Question (b)

detach(package:igraph)

library(statnet)

library(ggplot2)

par(mar = rep(0.15, 4))

plot(lazega.net, displaylabels = T, vertex.cex=3, label.cex=0.5, label.pos=5, pad=0.1)

png(filename = "lazega.png", width = 10000, height = 5000, pointsize = 100)

plot(lazega.net, displaylabels = T, vertex.cex=3, label.cex=0.5, label.pos=5, pad=0.1)

dev.off()

# diagram 1

nodeColours <- ifelse(v.attr.lazega$Gender=="1", "dodgerblue", "hotpink")

lazega.school.factor <- as.factor(v.attr.lazega$School)

cex\_age <- v.attr.lazega$Age

side\_nb <- 3:7

par(mar = rep(0.2,4))

gplot(lazega.net, mode = "kamadakawai",

label = v.attr.lazega$Office, label.cex=0.6, pad=0.1, #label

arrowhead = F, vertex.col = nodeColours, #color

vertex.sides = side\_nb[lazega.school.factor], #node shape

vertex.cex = log(cex\_age-30)-1.2) #node size

legend("bottom", legend = c("Male", "Female"), col = c("dodgerblue","hotpink"), pch=19, pt.cex=1.5, title = "Gender")

legend("bottomleft", legend = c("Office 1", "Office 2", "Office 3"), title = "Node label")

title("Kamada-Kawai layout algorithm", line = -2)

# diagram 2

lazega.practice.factor <- as.factor(v.attr.lazega$Practice)

cex\_yrs <- v.attr.lazega$Years

side\_nb <- 3:7

par(mar = rep(0.2,4))

gplot(lazega.net, mode = "kamadakawai",

label = v.attr.lazega$School, label.cex=0.6, pad=0.1, #label

arrowhead = F, vertex.col = nodeColours, #color

vertex.sides = side\_nb[lazega.practice.factor], #node shape

vertex.cex = log(cex\_yrs)-1) #node size

legend("bottom", legend = c("Male", "Female"), col = c("dodgerblue","hotpink"), pch=19, pt.cex=1.5)

legend("bottomleft", legend = c("School 1", "School 2", "School 3"), title = "Node label")

legend("bottomright", legend = c("Practice 1", "Practice 2"), pch = c(0,2))

title("Kamada-Kawai layout algorithm", line = -2)

##############################################################

## Question (c)

gden(lazega.net)

is.connected(lazega.net)

is.connected(lazega.net, connected = "weak")

##############################################################

## Question (d)

#node centrality

detach(package:igraph)

library(statnet)

deg <- degree(lazega.net,rescale = TRUE)

summary(deg)

#closeness centrality

lazega.gdist <- geodist(lazega.net)$gdist

1/sum(lazega.gdist)[1,2:network.size(lazega.net)]

which(lazega.gdist[,1]=="Inf")

sum(1/lazega.gdist[1,2:network.size(lazega.net)])

closeness2 <- function(x){

geo <- 1/geodist(x)$gdist

diag(geo) <- 0

apply(geo,1,sum)

}

close2 <- closeness2(lazega.net)

#betweenness centrality

between <- betweenness(lazega.net)

#centralization

centralization(lazega.net, degree, cmode = "indegree")

cd <- centralization(lazega.net, degree)

cc <- centralization(lazega.net, closeness, cmode = "suminvdir")

cb <- centralization(lazega.net, betweenness)

#Bonacich power centrality

power\_centrality(ilazega.net)

##############################################################

## Question (e)

detach(package:statnet)

library(igraph)

library(intergraph)

#k-core structure analysis

lazega.coreness <- graph.coreness(ilazega.net)

table(lazega.coreness)

kc<-kcores(lazega.net, cmode="indegree")

#plot the nest of cores

gplot(lazega.net, displaylabels = TRUE, usearrows = FALSE,vertex.col=rainbow(max(kc)+1)[kc+1])

#plot only 2-core

gplot(lazega.net[kc>1,kc>1], displaylabels = TRUE, vertex.col=rainbow(max(kc)+1)[kc[kc>1]+1])

V(ilazega.net)$name <- get.vertex.attribute(ilazega.net, name = "vertex.names", index = V(ilazega.net))

V(ilazega.net)$color <- lazega.coreness + 1

lay <- layout.fruchterman.reingold(ilazega.net)

op <- par(mfrow=c(1,2), mar = rep(0,4))

# In igraph, name attribute is used by default to label the nodes in a plot

plot(ilazega.net, vertex.size = 15, vertex.label.cex = 0.7, edge.arrow.size = 0.5, layout=lay)

# Label the nodes with their k-core membership value

plot(ilazega.net, vertex.label = lazega.coreness, vertex.label.cex = 0.7, edge.arrow.size = 0.5, layout=lay)

par(op)

ilazega.net\_core1 <- induced.subgraph(ilazega.net, vids = which(lazega.coreness >= 1))

ilazega.net\_core3 <- induced.subgraph(ilazega.net, vids = which(lazega.coreness >= 3))

ilazega.net\_core5 <- induced.subgraph(ilazega.net, vids = which(lazega.coreness >= 5))

lay <- layout.fruchterman.reingold(ilazega.net)

op <- par(mfrow=c(2,2), mar = c(2,0,2,0))

plot(ilazega.net, vertex.label = lazega.coreness, layout = lay, main = "All k-cores", edge.arrow.size = 0.1)

plot(ilazega.net\_core1, vertex.label = lazega.coreness[which(lazega.coreness >=1)], layout = lay[which(lazega.coreness >=1),], main = "k-cores 1+", edge.arrow.size = 0.1)

plot(ilazega.net\_core3, vertex.label = lazega.coreness[which(lazega.coreness >=3)], layout = lay[which(lazega.coreness >=3),], main = "k-cores 3+", edge.arrow.size = 0.1)

plot(ilazega.net\_core5, vertex.label = lazega.coreness[which(lazega.coreness >=5)], layout = lay[which(lazega.coreness >=5),], main = "k-cores 5+", edge.arrow.size = 0.1)

###############################################################

## Question (f)

mod\_sex <- modularity(ilazega.net, V(ilazega.net)$sex, weights = E(ilazega.net)$strength)

mod\_office <- modularity(ilazega.net, V(ilazega.net)$office, weights = E(ilazega.net)$strength)

mod\_years <- modularity(ilazega.net, V(ilazega.net)$years, weights = E(ilazega.net)$strength)

mod\_age <- modularity(ilazega.net, V(ilazega.net)$age, weights = E(ilazega.net)$strength)

mod\_practice <- modularity(ilazega.net, V(ilazega.net)$practice, weights = E(ilazega.net)$strength)

mod\_school <- modularity(ilazega.net, V(ilazega.net)$school, weights = E(ilazega.net)$strength)

mod\_table <- matrix(c(mod\_sex, mod\_office, mod\_years, mod\_age, mod\_practice,mod\_school),

ncol = 6, byrow = T)

colnames(mod\_table) <- c("Sex", "Office", "Years", "Age", "Practice", "School")

rownames(mod\_table) <- "Modularity"

as.table(mod\_table)

##############################################################

## Question (g)

# remove the isolates

lazega.iso.net <- lazega.net

delete.vertices(lazega.iso.net, isolates(lazega.iso.net))

lazega.iso <- lazega.iso.net

plot(lazega.iso)

ilazega.iso <- asIgraph(lazega.iso)

lazega.cw <- cluster\_walktrap(ilazega.iso)

class(lazega.cw)

membership(lazega.cw)

mod\_cw <- modularity(lazega.cw)

#using network with isolates

lazega.edge <- cluster\_edge\_betweenness(ilazega.net)

mod\_edge <- modularity(lazega.edge)

lazega.greedy <- cluster\_fast\_greedy(ilazega.net)

mod\_greedy <- modularity(lazega.greedy)

#plots

par(mfrow=c(2,2), mar = c(2,0,2,0))

lay2 <- layout.fruchterman.reingold(ilazega.iso)

lay3 <- layout.fruchterman.reingold(ilazega.net)

plot(lazega.cw, ilazega.iso, vertex.label = V(ilazega.iso)$office, layout = lay2, main = "Walktrap")

plot(lazega.edge, ilazega.net, vertex.label = V(ilazega.net)$office, layout = lay3, main = "Edge betweenness")

plot(lazega.greedy, ilazega.net, vertex.label = V(ilazega.net)$office, layout = lay3, main = "Fastgreedy")

mod\_table\_argo <- matrix(c(mod\_cw, mod\_edge, mod\_greedy), ncol = 3, byrow = T)

colnames(mod\_table\_argo) <- c("Walktrap", "Edge Betweenness", "Fastgreedy")

rownames(mod\_table\_argo) <- "Modularity"

as.table(mod\_table\_argo)

##############################################################