Hierarchy in a TC, Facility 1 Men’s Unit 1

Benjamin W. Campbell

Table of Contents

# Introduction

This notebook is for a project related to modeling the hierarchy within a TC clinical setting, looking specifically at corrections and whether many of our prior expectations about hierarchy within the TC environment hold. For example, does maximum position within the hierarchy correlate with outcomes, such as graduation?

# Create Network Objects

The first step is to load in the data. We want weighted, directed networks of corrections at the weekly level.

## set up working directory  
wd <- getwd()  
setwd(wd)  
  
## load corrections edgelist  
edgelist <- read.table(paste0(wd,"/data/F1-ledge"), stringsAsFactors = FALSE)  
  
## process data  
library(tidyverse)

## Warning: package 'tibble' was built under R version 3.6.2

## Warning: package 'tidyr' was built under R version 3.6.2

## Warning: package 'dplyr' was built under R version 3.6.2

library(lubridate)  
  
edgelist\_cleaned <- edgelist %>%  
 dplyr::rename(Date = V1, Sender = V2, Reciever = V3, Weight = V4) %>%  
 filter(Sender != 0) %>%  
 group\_by(Date, Sender, Reciever) %>%  
 summarize(Weight = sum(Weight))  
  
summary(edgelist\_cleaned$Weight)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.000 1.000 1.000 1.092 1.000 7.000

edgelist\_cleaned$Date <- mdy(edgelist\_cleaned$Date)  
  
## get t time stamp  
edgelist\_cleaned$t <- as.numeric(round(difftime(edgelist\_cleaned$Date, min(edgelist\_cleaned$Date), units = "weeks"))+1)  
# get in rank  
edgelist\_cleaned$t <- match(edgelist\_cleaned$t, sort(unique(edgelist\_cleaned$t)))  
  
## aggregate to week  
edgelist\_weekly <- edgelist\_cleaned %>%  
 group\_by(Sender, Reciever, t) %>%  
 summarize(Weight = sum(Weight)) %>%  
 arrange(t)  
  
head(edgelist\_weekly)

## # A tibble: 6 x 4  
## # Groups: Sender, Reciever [6]  
## Sender Reciever t Weight  
## <chr> <chr> <int> <int>  
## 1 "" "" 1 1  
## 2 "" "6780" 1 1  
## 3 "" "6781" 1 2  
## 4 "" "6970" 1 1  
## 5 "" "7040" 1 1  
## 6 "" "7045" 1 2

summary(edgelist\_weekly$Weight)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.000 1.000 1.000 1.201 1.000 15.000

## get into network format  
library(igraph)  
t\_steps <- sort(unique(edgelist\_weekly$t))  
net\_list <- as.list(rep(NA, length(t\_steps)))  
index = 0  
  
## function to make network for time slice  
create\_network <- function(edgelist, t){  
 # reduce edgelist to time t  
 t\_slice <- edgelist[edgelist$t == t,]  
 t\_graph <- graph.data.frame(t\_slice, directed = TRUE)  
 # return network  
 return(t\_graph)  
}  
  
## populate list  
for(t in t\_steps){  
 # increment index  
 index = index+1  
 # create network  
 net <- create\_network(edgelist\_weekly, t)  
 # insert into index'ed element of list  
 net\_list[[index]] <- net  
}

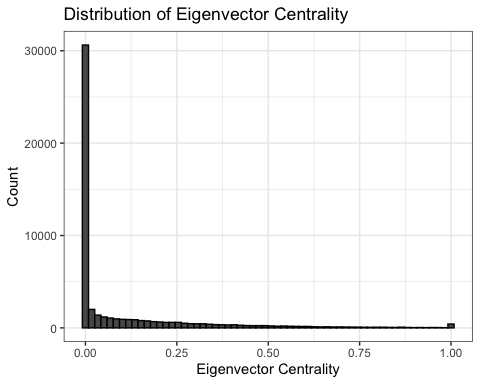
# Calculate Eigenvector Centrality

net\_list now contains a list of weighted and directed igraph objects. With this list, we can then go on to compute eigenvector centrality at the weekly level for every node.

# make function  
get\_eigen\_table <- function(graph, t){  
 # get weighted eigenvector centrality  
 scores <- eigen\_centrality(graph, weights = E(graph)$Weight, directed = TRUE)$vector  
 # put in table  
 t\_df <- tibble(  
 Id = as.character(names(scores)),  
 t = as.integer(t),  
 eigen\_cent = as.numeric(scores)  
 )  
 # return table  
 return(t\_df)  
}  
  
# initialize empty dataframe  
eigen\_df <- tibble()  
  
# loop through  
for(t in 1:length(net\_list)){  
 # get one network  
 net <- net\_list[[t]]  
 # get dataframe  
 t\_df <- get\_eigen\_table(net, t)  
 # bind to original dataframe  
 eigen\_df <- bind\_rows(eigen\_df, t\_df)  
}  
  
head(eigen\_df)

## # A tibble: 6 x 3  
## Id t eigen\_cent  
## <chr> <int> <dbl>  
## 1 "" 1 2.04e-17  
## 2 "6780" 1 1.86e-17  
## 3 "6781" 1 6.03e- 1  
## 4 "6829" 1 1.63e-17  
## 5 "6832" 1 1.63e-17  
## 6 "6833" 1 1.63e-17

ggplot(eigen\_df, aes(x = eigen\_cent)) +  
 geom\_histogram(colour="black", bins = 60) +  
 theme\_bw() +  
 ggtitle("Distribution of Eigenvector Centrality") +  
 xlab("Eigenvector Centrality") +  
 ylab("Count")

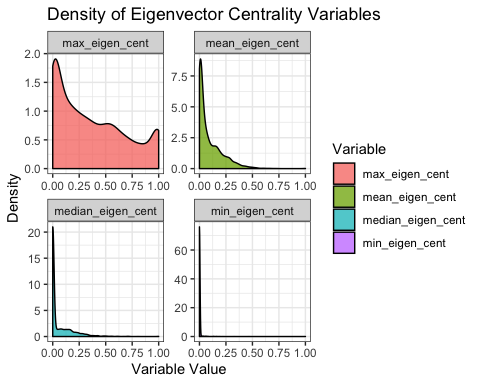


# Create Variables

So, given that we have an eigenvector centrality that is measured longitudinally, but only have a single observation of the outcome, how do we collapse this measure?

* We could look at minimum eigenvector centrality, which would tell us about the highest position in the hierarchy that anyone ever achieves.
* We could look at maximum eigenvector centrality, which would tell us about the lowest position in the hierarchy that anyone ever achieves.
* we could look at average or median eigenvector centrality, which would tell us something about the central tendency of someone in the networ with respect to where they are in the hierarchy.
* We could look at any of the prior measures over their last month or something there. This would tell us in general how well they do towards the end of their tenure.

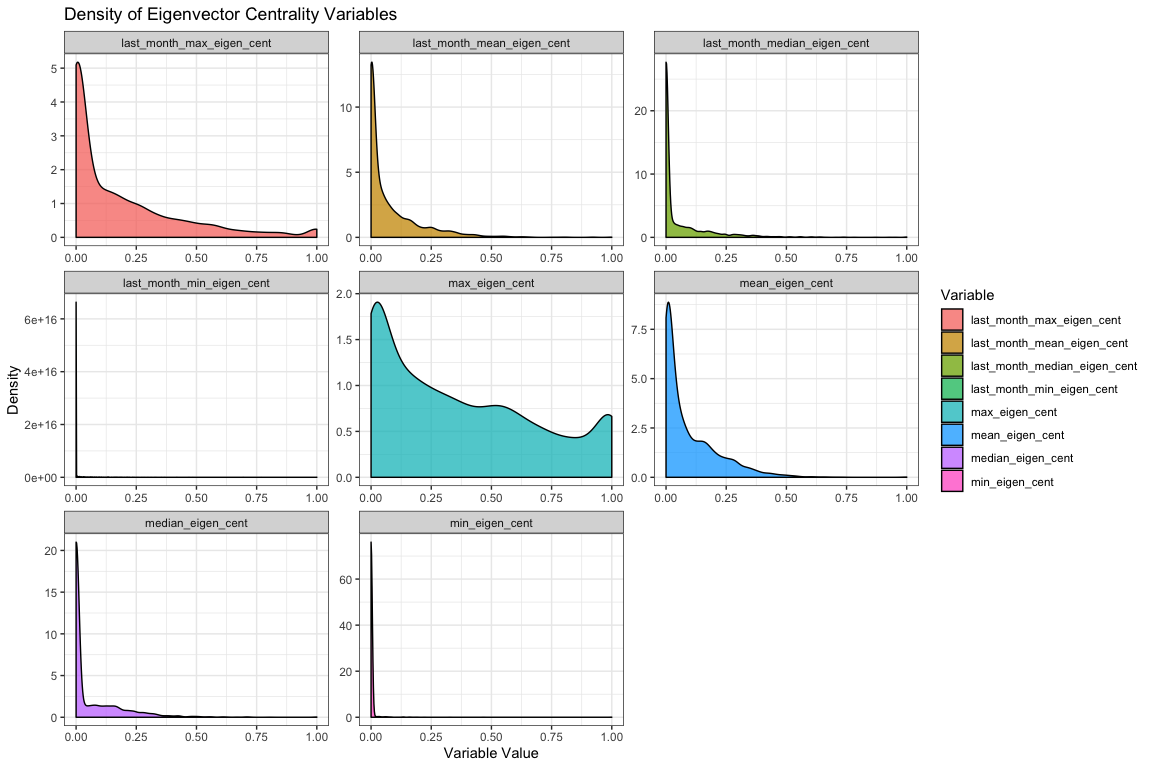
# min. eigen centrality -- highest position ever achieved in hierarchy  
# max. eigen centrality -- lowest position ever achieved in hierarchy  
# mean eigen centrality -- average position in the hierarchy  
# median eigen centrality -- another measure of central tendency  
nodal\_eigen <- eigen\_df %>%  
 group\_by(Id) %>%  
 summarize(min\_eigen\_cent = min(eigen\_cent),  
 max\_eigen\_cent = max(eigen\_cent),  
 mean\_eigen\_cent = mean(eigen\_cent),  
 median\_eigen\_cent = median(eigen\_cent))  
  
# plot df  
plot\_df <- nodal\_eigen %>%   
 gather("Variable", "Value",-Id)  
   
  
ggplot(plot\_df, aes(x = Value, fill = Variable)) +  
 geom\_density(colour="black", alpha = 0.75) +  
 theme\_bw() +  
 ggtitle("Density of Eigenvector Centrality Variables") +  
 xlab("Variable Value") +  
 ylab("Density") +   
 facet\_wrap(vars(Variable), scales = 'free')



# Summaries  
summary(nodal\_eigen)

## Id min\_eigen\_cent max\_eigen\_cent   
## Length:3579 Min. :0.000000 Min. :0.00000   
## Class :character 1st Qu.:0.000000 1st Qu.:0.06395   
## Mode :character Median :0.000000 Median :0.29678   
## Mean :0.002297 Mean :0.36996   
## 3rd Qu.:0.000000 3rd Qu.:0.60811   
## Max. :1.000000 Max. :1.00000   
## mean\_eigen\_cent median\_eigen\_cent  
## Min. :0.000000 Min. :0.00000   
## 1st Qu.:0.007098 1st Qu.:0.00000   
## Median :0.043158 Median :0.00000   
## Mean :0.094296 Mean :0.06172   
## 3rd Qu.:0.151367 3rd Qu.:0.09225   
## Max. :1.000000 Max. :1.00000

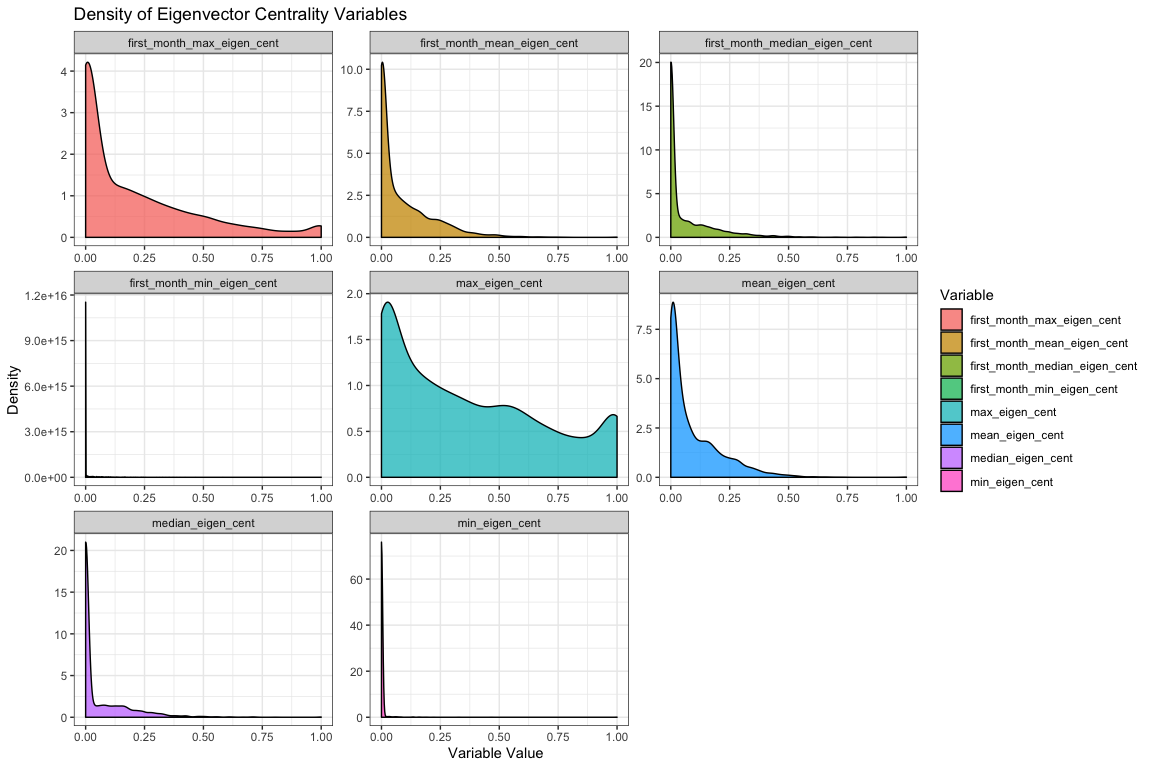
get\_last\_month\_variables <- function(id){  
 id\_df <- eigen\_df %>%   
 filter(Id == id) %>%   
 arrange(t) %>%   
 tail(4) %>%  
 summarize(Id = unique(Id),  
 last\_month\_min\_eigen\_cent = min(eigen\_cent),  
 last\_month\_max\_eigen\_cent = max(eigen\_cent),  
 last\_month\_mean\_eigen\_cent = mean(eigen\_cent),  
 last\_month\_median\_eigen\_cent = median(eigen\_cent))  
 return(id\_df)  
}  
  
ids <- unique(eigen\_df$Id)  
  
last\_month\_df <- tibble()  
  
for(i in ids){  
 id\_df <- get\_last\_month\_variables(i)  
 last\_month\_df <- bind\_rows(last\_month\_df, id\_df)  
}  
  
# plot df  
plot\_df\_full <- last\_month\_df %>%   
 gather("Variable", "Value",-Id) %>%  
 bind\_rows(plot\_df)  
   
  
ggplot(plot\_df\_full, aes(x = Value, fill = Variable)) +  
 geom\_density(colour="black", alpha = 0.75) +  
 theme\_bw() +  
 ggtitle("Density of Eigenvector Centrality Variables") +  
 xlab("Variable Value") +  
 ylab("Density") +   
 facet\_wrap(vars(Variable), scales = 'free', nrow = 3)



# Summaries  
summary(last\_month\_df)

## Id last\_month\_min\_eigen\_cent last\_month\_max\_eigen\_cent  
## Length:3579 Min. :0.00000 Min. :0.0000000   
## Class :character 1st Qu.:0.00000 1st Qu.:0.0000002   
## Mode :character Median :0.00000 Median :0.0694461   
## Mean :0.01291 Mean :0.1788988   
## 3rd Qu.:0.00000 3rd Qu.:0.2721853   
## Max. :1.00000 Max. :1.0000000   
## last\_month\_mean\_eigen\_cent last\_month\_median\_eigen\_cent  
## Min. :0.00000 Min. :0.0000000   
## 1st Qu.:0.00000 1st Qu.:0.0000000   
## Median :0.02211 Median :0.0000065   
## Mean :0.07708 Mean :0.0581374   
## 3rd Qu.:0.10946 3rd Qu.:0.0681954   
## Max. :1.00000 Max. :1.0000000

get\_first\_month\_variables <- function(id){  
 id\_df <- eigen\_df %>%   
 filter(Id == id) %>%   
 arrange(t) %>%   
 head(4) %>%  
 summarize(Id = unique(Id),  
 first\_month\_min\_eigen\_cent = min(eigen\_cent),  
 first\_month\_max\_eigen\_cent = max(eigen\_cent),  
 first\_month\_mean\_eigen\_cent = mean(eigen\_cent),  
 first\_month\_median\_eigen\_cent = median(eigen\_cent))  
 return(id\_df)  
}  
  
ids <- unique(eigen\_df$Id)  
  
first\_month\_df <- tibble()  
  
for(i in ids){  
 id\_df <- get\_first\_month\_variables(i)  
 first\_month\_df <- bind\_rows(first\_month\_df, id\_df)  
}  
  
# plot df  
plot\_df\_full <- first\_month\_df %>%   
 gather("Variable", "Value",-Id) %>%  
 bind\_rows(plot\_df)  
   
  
ggplot(plot\_df\_full, aes(x = Value, fill = Variable)) +  
 geom\_density(colour="black", alpha = 0.75) +  
 theme\_bw() +  
 ggtitle("Density of Eigenvector Centrality Variables") +  
 xlab("Variable Value") +  
 ylab("Density") +   
 facet\_wrap(vars(Variable), scales = 'free', nrow = 3)



# Summaries  
summary(first\_month\_df)

## Id first\_month\_min\_eigen\_cent first\_month\_max\_eigen\_cent  
## Length:3579 Min. :0.00000 Min. :0.0000000   
## Class :character 1st Qu.:0.00000 1st Qu.:0.0000172   
## Mode :character Median :0.00000 Median :0.0937946   
## Mean :0.01379 Mean :0.2054786   
## 3rd Qu.:0.00000 3rd Qu.:0.3267918   
## Max. :1.00000 Max. :1.0000000   
## first\_month\_mean\_eigen\_cent first\_month\_median\_eigen\_cent  
## Min. :0.0000000 Min. :0.0000000   
## 1st Qu.:0.0000044 1st Qu.:0.0000000   
## Median :0.0312856 Median :0.0005118   
## Mean :0.0874034 Mean :0.0650609   
## 3rd Qu.:0.1392061 3rd Qu.:0.0896701   
## Max. :1.0000000 Max. :1.0000000

# Join to Node Data

With the measures of hierarchy created, we can now process the node data and join these variables to it. Once all of that is taken care of we can move on to analysis!

# read node data  
nodes <- read.table(paste0(wd,"/data/F1M1"), stringsAsFactors = FALSE, header = TRUE)  
  
sum(!is.na(nodes$mdid))

## [1] 964

# get total unique nodes  
length(unique(nodes$mdid))

## [1] 924

# get days in program  
nodes$days\_in\_program <- as.Date(as.character(nodes$exit), format="%m/%d/%Y")-as.Date(as.character(nodes$enter), format="%m/%d/%Y")  
  
# process recidivism  
nodes$recidFlag <- rep(0, times = nrow(nodes))  
nodes$recidFlag[!(is.na(nodes$recidate1))] <- 1  
   
nodes$recidDate <- as.character(nodes$recidate1)  
nodes[is.na(nodes$recidDate),]$recidDate <- "09/04/2009"  
  
nodes$gap <- as.Date(as.character(nodes$recidDate), format="%m/%d/%Y")-as.Date(as.character(nodes$exit), format="%m/%d/%Y")  
  
# remove folks who visit multiple times  
repeat\_visitors <- names(which(table(nodes$mdid) > 1))  
  
nodes <- nodes[!(nodes$mdid %in% repeat\_visitors),]  
  
nodes$mdid <- as.character(nodes$mdid)  
  
# join network variables  
# first rename Id to wcid  
nodes$lsir <- as.numeric(nodes$lsir)

## Warning: NAs introduced by coercion

dat <- nodes %>%  
 rename(Id = mdid) %>%  
 select(Id, age, lsir, black, success, recidFlag, recidDate, gap, days\_in\_program) %>%  
 inner\_join(last\_month\_df, by = "Id") %>%  
 inner\_join(first\_month\_df, by = "Id") %>%  
 inner\_join(nodal\_eigen, by = "Id")  
  
dat$days\_in\_program <- as.numeric(dat$days\_in\_program)  
  
head(dat)

## Id age lsir black success recidFlag recidDate gap  
## 1 7806 29 NA 0 1 0 09/04/2009 2801 days  
## 2 8095 30 NA 0 1 0 09/04/2009 2713 days  
## 3 8251 32 NA 1 1 0 09/04/2009 2626 days  
## 4 8333 28 12 1 1 0 09/04/2009 2654 days  
## 5 8357 19 22 0 1 0 09/04/2009 2490 days  
## 6 8747 20 34 1 1 0 09/04/2009 2417 days  
## days\_in\_program last\_month\_min\_eigen\_cent last\_month\_max\_eigen\_cent  
## 1 150 1.991768e-17 1.407778e-03  
## 2 102 0.000000e+00 5.485533e-05  
## 3 142 0.000000e+00 1.338031e-01  
## 4 105 6.787844e-18 4.910991e-02  
## 5 179 0.000000e+00 2.500000e-01  
## 6 163 0.000000e+00 2.646906e-01  
## last\_month\_mean\_eigen\_cent last\_month\_median\_eigen\_cent  
## 1 3.544116e-04 4.933940e-06  
## 2 1.371383e-05 1.315708e-17  
## 3 3.345078e-02 2.393528e-17  
## 4 1.227748e-02 1.861745e-17  
## 5 6.250000e-02 1.920873e-17  
## 6 7.789513e-02 2.344498e-02  
## first\_month\_min\_eigen\_cent first\_month\_max\_eigen\_cent  
## 1 8.598704e-17 7.093579e-04  
## 2 0.000000e+00 1.633249e-01  
## 3 0.000000e+00 5.315433e-16  
## 4 0.000000e+00 1.161352e-01  
## 5 0.000000e+00 1.829422e-02  
## 6 0.000000e+00 1.142043e-15  
## first\_month\_mean\_eigen\_cent first\_month\_median\_eigen\_cent min\_eigen\_cent  
## 1 2.250246e-04 9.537032e-05 0  
## 2 4.083486e-02 7.281200e-06 0  
## 3 1.441407e-16 2.250977e-17 0  
## 4 2.903380e-02 2.097020e-17 0  
## 5 4.573554e-03 0.000000e+00 0  
## 6 2.975587e-16 2.409590e-17 0  
## max\_eigen\_cent mean\_eigen\_cent median\_eigen\_cent  
## 1 1.0000000 0.06691632 1.341905e-15  
## 2 0.1633249 0.01633943 1.065665e-17  
## 3 0.1338031 0.01933487 4.390176e-17  
## 4 0.1161352 0.01383209 1.939259e-17  
## 5 0.2500000 0.01388080 0.000000e+00  
## 6 0.5894750 0.10192553 1.438834e-07

# Exploratory Data Analysis

We’ve got the data put together, now is time to think about the relationship between these key variables and TC outcomes like graduation or recidivism. The following network variables we think might matter most based upon their distributions:

* last\_month\_max\_eigen\_cent
* last\_month\_mean\_eigen\_cent
* max\_eigen\_cent
* mean\_eigen\_cent

# Neg, sig - as lower in hierarchy towards the end, less likely to be successful  
cor.test(dat$last\_month\_max\_eigen\_cent, dat$success)

##   
## Pearson's product-moment correlation  
##   
## data: dat$last\_month\_max\_eigen\_cent and dat$success  
## t = -4.6547, df = 875, p-value = 3.746e-06  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.21938853 -0.09017448  
## sample estimates:  
## cor   
## -0.1554463

# Neg, sig - as lower in hierarchy towards the end, less likely to be successful  
cor.test(dat$last\_month\_mean\_eigen\_cent, dat$success)

##   
## Pearson's product-moment correlation  
##   
## data: dat$last\_month\_mean\_eigen\_cent and dat$success  
## t = -5.205, df = 875, p-value = 2.419e-07  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.2367821 -0.1083419  
## sample estimates:  
## cor   
## -0.1732987

# Neg, sig - as lower in hierarchy, less likely to be successful  
cor.test(dat$max\_eigen\_cent, dat$success)

##   
## Pearson's product-moment correlation  
##   
## data: dat$max\_eigen\_cent and dat$success  
## t = -1.0018, df = 875, p-value = 0.3167  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.09982282 0.03242579  
## sample estimates:  
## cor   
## -0.03384668

# Neg, sig - as lower in hierarchy, less likely to be successful  
cor.test(dat$mean\_eigen\_cent, dat$success)

##   
## Pearson's product-moment correlation  
##   
## data: dat$mean\_eigen\_cent and dat$success  
## t = -3.8329, df = 875, p-value = 0.0001357  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.19305913 -0.06283636  
## sample estimates:  
## cor   
## -0.1285016

# Modeling Graduation

The first and most easy thing we could do is use simple linear modeling to examine the effect of some of these covariates on graduation, while controlling for the confounding effects of other variables. Here we will fit those models and show results:

last\_month\_max\_model <- glm(success ~  
 age +  
 lsir +  
 black +  
 days\_in\_program +  
 last\_month\_max\_eigen\_cent,  
 data = dat,  
 family = binomial(link = 'logit'))  
  
last\_month\_mean\_model <- glm(success ~  
 age +  
 lsir +  
 black +  
 days\_in\_program +  
 last\_month\_mean\_eigen\_cent,  
 data = dat,  
 family = binomial(link = 'logit'))  
  
max\_model <- glm(success ~  
 age +  
 lsir +  
 black +  
 days\_in\_program +  
 max\_eigen\_cent,  
 data = dat,  
 family = binomial(link = 'logit'))  
  
mean\_model <- glm(success ~  
 age +  
 lsir +  
 black +  
 days\_in\_program +  
 mean\_eigen\_cent,  
 data = dat,  
 family = binomial(link = 'logit'))  
  
library(texreg)

## Warning: package 'texreg' was built under R version 3.6.2

screenreg(l = list(last\_month\_max\_model, last\_month\_mean\_model, max\_model, mean\_model))

##   
## ==============================================================================  
## Model 1 Model 2 Model 3 Model 4   
## ------------------------------------------------------------------------------  
## (Intercept) 6.36 \*\*\* 6.42 \*\*\* 6.90 \*\*\* 6.86 \*\*\*  
## (1.35) (1.36) (1.39) (1.38)   
## age -0.00 -0.00 -0.01 -0.01   
## (0.02) (0.02) (0.02) (0.02)   
## lsir -0.36 \*\*\* -0.36 \*\*\* -0.37 \*\*\* -0.36 \*\*\*  
## (0.04) (0.04) (0.04) (0.04)   
## black 0.25 0.24 0.24 0.23   
## (0.31) (0.32) (0.32) (0.32)   
## days\_in\_program 0.05 \*\*\* 0.05 \*\*\* 0.06 \*\*\* 0.05 \*\*\*  
## (0.01) (0.01) (0.01) (0.01)   
## last\_month\_max\_eigen\_cent -1.64 \*   
## (0.81)   
## last\_month\_mean\_eigen\_cent -6.30 \*   
## (2.57)   
## max\_eigen\_cent -1.63 \*\*   
## (0.55)   
## mean\_eigen\_cent -8.11 \*   
## (3.20)   
## ------------------------------------------------------------------------------  
## AIC 301.78 299.88 296.92 299.46   
## BIC 329.97 328.08 325.11 327.66   
## Log Likelihood -144.89 -143.94 -142.46 -143.73   
## Deviance 289.78 287.88 284.92 287.46   
## Num. obs. 812 812 812 812   
## ==============================================================================  
## \*\*\* p < 0.001; \*\* p < 0.01; \* p < 0.05

get\_pred\_prob\_plot <- function(model, xvar, xlab){  
 std <- qnorm(0.95 / 2 + 0.5)  
   
 #last\_month\_max\_eigen\_cent\_plot  
 data = model$data  
 new\_data <- data.frame(  
 age = rep(mean(data$age, na.rm = TRUE), nrow(data)),  
 lsir = rep(mean(data$lsir, na.rm = TRUE), nrow(data)),  
 black = rep(0, nrow(data)),  
 days\_in\_program = rep(mean(data$days\_in\_program, na.rm = TRUE), nrow(data)),  
 stupid\_placeholder = data[,xvar]  
 )  
   
 colnames(new\_data)[5] <- xvar  
   
 predicted\_data <- as.data.frame(predict(model, newdata = new\_data,  
 type="link", se=TRUE))  
   
 new\_data <- cbind(new\_data, predicted\_data)  
 new\_data$ymin <- model$family$linkinv(new\_data$fit - std \* new\_data$se)  
 new\_data$ymax <- model$family$linkinv(new\_data$fit + std \* new\_data$se)  
 new\_data$fit <- model$family$linkinv(new\_data$fit)  
   
 library(ggplot2)  
 p <- ggplot(new\_data, aes(x=new\_data[,xvar])) +  
 geom\_ribbon(data = new\_data, aes(y=fit, ymin=ymin, ymax=ymax), alpha = 0.5) +  
 geom\_line(data = new\_data, aes(x = new\_data[,xvar], y=fit), size = 1.5, colour = "firebrick4") +  
 scale\_y\_continuous(limits=c(0,1)) +  
 theme\_bw() +   
 theme(legend.position = c(0.2, 0.8),  
 axis.text=element\_text(size=12),  
 axis.title=element\_text(size=14,face="bold"))+  
 labs(x=xlab, y="Probability of Graduation")   
   
 return(p)  
}  
  
last\_month\_max\_pred\_prob <- get\_pred\_prob\_plot(last\_month\_max\_model,   
 "last\_month\_max\_eigen\_cent",  
 "Last Month's Highest Eigenvector Centrality")

## Warning: Ignoring unknown aesthetics: y

last\_month\_mean\_pred\_prob <- get\_pred\_prob\_plot(last\_month\_mean\_model,   
 "last\_month\_mean\_eigen\_cent",  
 "Last Month's Average Eigenvector Centrality")

## Warning: Ignoring unknown aesthetics: y

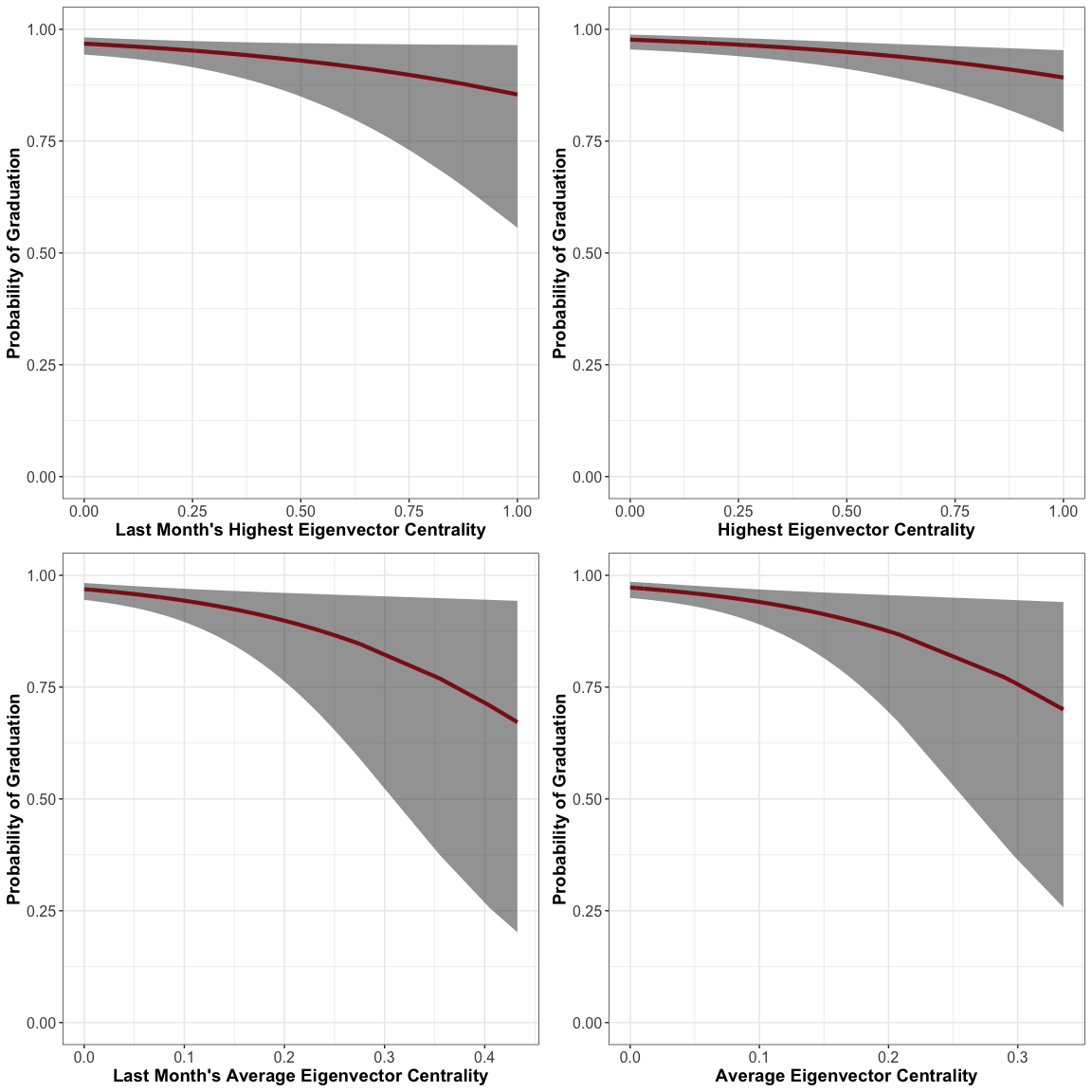
max\_eigen\_cent\_pred\_prob <- get\_pred\_prob\_plot(max\_model,   
 "max\_eigen\_cent",  
 "Highest Eigenvector Centrality")

## Warning: Ignoring unknown aesthetics: y

mean\_eigen\_cent\_pred\_prob <- get\_pred\_prob\_plot(mean\_model,   
 "mean\_eigen\_cent",  
 "Average Eigenvector Centrality")

## Warning: Ignoring unknown aesthetics: y

library(Rmisc)  
multiplot(plotlist = list(last\_month\_max\_pred\_prob, last\_month\_mean\_pred\_prob, max\_eigen\_cent\_pred\_prob, mean\_eigen\_cent\_pred\_prob),  
 cols = 2)



get\_pred\_prob\_dist <- function(model, xvar){  
 data = model$data  
   
 mean <- mean(data[,xvar], na.rm = TRUE)  
 std <- sd(data[,xvar], na.rm = TRUE)  
 input\_vec <- c(mean-2\*std, mean-1\*std, mean, mean+1\*std, mean+2\*std)  
   
 new\_data <- data.frame(  
 age = rep(mean(data$age, na.rm = TRUE), length(input\_vec)),  
 lsir = rep(mean(data$lsir, na.rm = TRUE), length(input\_vec)),  
 black = rep(0, length(input\_vec)),  
 days\_in\_program = rep(mean(data$days\_in\_program, na.rm = TRUE), length(input\_vec)),  
 stupid\_placeholder = input\_vec  
 )  
   
 colnames(new\_data)[5] <- xvar  
   
 predicted\_data <- as.data.frame(predict(model, newdata = new\_data,  
 type="link", se=TRUE))  
   
 probs <- model$family$linkinv(predicted\_data$fit)   
   
 return(probs)  
}  
  
# 2 std below mean, 1 std below mean, mean, 1 std above, 2 std above  
get\_pred\_prob\_dist(last\_month\_max\_model,   
 "last\_month\_max\_eigen\_cent")

## [1] 0.9780391 0.9719140 0.9641433 0.9543236 0.9419765

# 2 std below mean, 1 std below mean, mean, 1 std above, 2 std above  
get\_pred\_prob\_dist(last\_month\_mean\_model,   
 "last\_month\_mean\_eigen\_cent")

## [1] 0.9805890 0.9738151 0.9647621 0.9527312 0.9368615

# 2 std below mean, 1 std below mean, mean, 1 std above, 2 std above  
get\_pred\_prob\_dist(max\_model,   
 "max\_eigen\_cent")

## [1] 0.9858734 0.9782223 0.9665677 0.9490011 0.9229403

# 2 std below mean, 1 std below mean, mean, 1 std above, 2 std above  
get\_pred\_prob\_dist(mean\_model,   
 "mean\_eigen\_cent")

## [1] 0.9822334 0.9752462 0.9656073 0.9523982 0.9344603

# Descriptive Stats

max(edgelist\_cleaned$Date)-min(edgelist\_cleaned$Date)

## Time difference of 2844 days

summary(dat)

## Id age lsir black   
## Length:877 Min. :18.00 Min. : 0.0 Min. :0.0000   
## Class :character 1st Qu.:21.00 1st Qu.:22.0 1st Qu.:0.0000   
## Mode :character Median :25.00 Median :26.0 Median :0.0000   
## Mean :28.06 Mean :25.8 Mean :0.4903   
## 3rd Qu.:34.00 3rd Qu.:29.0 3rd Qu.:1.0000   
## Max. :61.00 Max. :44.0 Max. :1.0000   
## NA's :65   
## success recidFlag recidDate gap   
## Min. :0.0000 Min. :0.0000 Length:877 Length:877   
## 1st Qu.:1.0000 1st Qu.:0.0000 Class :character Class :difftime   
## Median :1.0000 Median :0.0000 Mode :character Mode :numeric   
## Mean :0.8575 Mean :0.2258   
## 3rd Qu.:1.0000 3rd Qu.:0.0000   
## Max. :1.0000 Max. :1.0000   
##   
## days\_in\_program last\_month\_min\_eigen\_cent last\_month\_max\_eigen\_cent  
## Min. : 12.0 Min. :0.0000000 Min. :0.000000   
## 1st Qu.: 99.0 1st Qu.:0.0000000 1st Qu.:0.000000   
## Median :119.0 Median :0.0000000 Median :0.000427   
## Mean :117.4 Mean :0.0008504 Mean :0.069302   
## 3rd Qu.:136.0 3rd Qu.:0.0000000 3rd Qu.:0.058671   
## Max. :179.0 Max. :0.1886187 Max. :1.000000   
##   
## last\_month\_mean\_eigen\_cent last\_month\_median\_eigen\_cent  
## Min. :0.0000000 Min. :0.00000   
## 1st Qu.:0.0000000 1st Qu.:0.00000   
## Median :0.0001144 Median :0.00000   
## Mean :0.0209713 Mean :0.00666   
## 3rd Qu.:0.0176617 3rd Qu.:0.00000   
## Max. :0.4326199 Max. :0.39122   
##   
## first\_month\_min\_eigen\_cent first\_month\_max\_eigen\_cent  
## Min. :0.000000 Min. :0.000000   
## 1st Qu.:0.000000 1st Qu.:0.000000   
## Median :0.000000 Median :0.001838   
## Mean :0.001153 Mean :0.095992   
## 3rd Qu.:0.000000 3rd Qu.:0.099108   
## Max. :0.188619 Max. :1.000000   
##   
## first\_month\_mean\_eigen\_cent first\_month\_median\_eigen\_cent  
## Min. :0.0000000 Min. :0.000000   
## 1st Qu.:0.0000000 1st Qu.:0.000000   
## Median :0.0004894 Median :0.000000   
## Mean :0.0282743 Mean :0.007770   
## 3rd Qu.:0.0289823 3rd Qu.:0.000109   
## Max. :0.3190242 Max. :0.348163   
##   
## min\_eigen\_cent max\_eigen\_cent mean\_eigen\_cent   
## Min. :0.0000000 Min. :0.00000 Min. :0.000000   
## 1st Qu.:0.0000000 1st Qu.:0.01074 1st Qu.:0.001026   
## Median :0.0000000 Median :0.13066 Median :0.013527   
## Mean :0.0008004 Mean :0.23100 Mean :0.028777   
## 3rd Qu.:0.0000000 3rd Qu.:0.33744 3rd Qu.:0.038699   
## Max. :0.1886187 Max. :1.00000 Max. :0.335369   
##   
## median\_eigen\_cent   
## Min. :0.000000   
## 1st Qu.:0.000000   
## Median :0.000000   
## Mean :0.002574   
## 3rd Qu.:0.000000   
## Max. :0.212124   
##

apply(dat, 2, function(x) sd(x, na.rm = TRUE))

## Warning in var(if (is.vector(x) || is.factor(x)) x else as.double(x), na.rm  
## = na.rm): NAs introduced by coercion  
  
## Warning in var(if (is.vector(x) || is.factor(x)) x else as.double(x), na.rm  
## = na.rm): NAs introduced by coercion

## Id age   
## 1.582176e+03 8.616822e+00   
## lsir black   
## 5.606615e+00 5.001913e-01   
## success recidFlag   
## 3.497938e-01 4.183267e-01   
## recidDate gap   
## NA NA   
## days\_in\_program last\_month\_min\_eigen\_cent   
## 3.316707e+01 1.108080e-02   
## last\_month\_max\_eigen\_cent last\_month\_mean\_eigen\_cent   
## 1.540585e-01 4.859455e-02   
## last\_month\_median\_eigen\_cent first\_month\_min\_eigen\_cent   
## 3.105023e-02 1.256889e-02   
## first\_month\_max\_eigen\_cent first\_month\_mean\_eigen\_cent   
## 1.903624e-01 5.545243e-02   
## first\_month\_median\_eigen\_cent min\_eigen\_cent   
## 3.001569e-02 1.106210e-02   
## max\_eigen\_cent mean\_eigen\_cent   
## 2.709076e-01 4.175939e-02   
## median\_eigen\_cent   
## 1.750177e-02

# Session Info

sessionInfo()

## R version 3.6.0 (2019-04-26)  
## Platform: x86\_64-apple-darwin15.6.0 (64-bit)  
## Running under: macOS Mojave 10.14.6  
##   
## Matrix products: default  
## BLAS: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRblas.0.dylib  
## LAPACK: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRlapack.dylib  
##   
## locale:  
## [1] en\_US.UTF-8/en\_US.UTF-8/en\_US.UTF-8/C/en\_US.UTF-8/en\_US.UTF-8  
##   
## attached base packages:  
## [1] stats graphics grDevices utils datasets methods base   
##   
## other attached packages:  
## [1] Rmisc\_1.5 plyr\_1.8.4 lattice\_0.20-38 texreg\_1.37.5   
## [5] igraph\_1.2.4.1 lubridate\_1.7.4 forcats\_0.4.0 stringr\_1.4.0   
## [9] dplyr\_1.0.2 purrr\_0.3.2 readr\_1.3.1 tidyr\_1.1.2   
## [13] tibble\_3.0.4 ggplot2\_3.2.1 tidyverse\_1.2.1  
##   
## loaded via a namespace (and not attached):  
## [1] tidyselect\_1.1.0 xfun\_0.9 haven\_2.1.0 colorspace\_1.4-1  
## [5] vctrs\_0.3.4 generics\_0.0.2 htmltools\_0.3.6 yaml\_2.2.0   
## [9] utf8\_1.1.4 rlang\_0.4.8 pillar\_1.4.6 glue\_1.4.2   
## [13] withr\_2.1.2 modelr\_0.1.4 readxl\_1.3.1 lifecycle\_0.2.0   
## [17] munsell\_0.5.0 gtable\_0.3.0 cellranger\_1.1.0 rvest\_0.3.4   
## [21] evaluate\_0.14 labeling\_0.3 knitr\_1.24 fansi\_0.4.0   
## [25] broom\_0.7.2 Rcpp\_1.0.5 scales\_1.0.0 backports\_1.1.4   
## [29] jsonlite\_1.6 hms\_0.4.2 digest\_0.6.20 stringi\_1.4.3   
## [33] grid\_3.6.0 cli\_1.1.0 tools\_3.6.0 magrittr\_1.5   
## [37] lazyeval\_0.2.2 crayon\_1.3.4 pkgconfig\_2.0.2 ellipsis\_0.3.1   
## [41] xml2\_1.2.0 assertthat\_0.2.1 rmarkdown\_1.12 httr\_1.4.0   
## [45] rstudioapi\_0.10 R6\_2.4.0 compiler\_3.6.0