Hierarchy in a TC, Facility 3 Men

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/F3M-ledge"), stringsAsFactors = FALSE)  
  
## process data  
library(tidyverse)

## ── Attaching packages ──────────────────────────────────────────────────────────────────────── tidyverse 1.2.1 ──

## ✔ ggplot2 3.2.1 ✔ purrr 0.3.2  
## ✔ tibble 3.0.4 ✔ dplyr 1.0.2  
## ✔ tidyr 1.1.2 ✔ stringr 1.4.0  
## ✔ readr 1.3.1 ✔ forcats 0.4.0

## 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

## ── Conflicts ─────────────────────────────────────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(lubridate)

##   
## Attaching package: 'lubridate'

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

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

## `summarise()` regrouping output by 'Date', 'Sender' (override with `.groups` argument)

summary(edgelist\_cleaned$Weight)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.000 1.000 1.000 1.053 1.000 5.000

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)

## `summarise()` regrouping output by 'Sender', 'Reciever' (override with `.groups` argument)

head(edgelist\_weekly)

## # A tibble: 6 x 4  
## # Groups: Sender, Reciever [6]  
## Sender Reciever t Weight  
## <chr> <chr> <int> <int>  
## 1 09-12-051 08-51-339 1 1  
## 2 01-12-549 01-51-552 2 1  
## 3 08-21-089 08-80-066 2 1  
## 4 01-12-493 01-44-501 3 1  
## 5 00-21-433 01-12-493 4 1  
## 6 00-49-430 01-11-508 4 1

summary(edgelist\_weekly$Weight)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.000 1.000 1.000 1.141 1.000 9.000

## get into network format  
library(igraph)

##   
## Attaching package: 'igraph'

## The following objects are masked from 'package:lubridate':  
##   
## %--%, union

## The following objects are masked from 'package:dplyr':  
##   
## as\_data\_frame, groups, union

## The following objects are masked from 'package:purrr':  
##   
## compose, simplify

## The following object is masked from 'package:tidyr':  
##   
## crossing

## The following object is masked from 'package:tibble':  
##   
## as\_data\_frame

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

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

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  
}

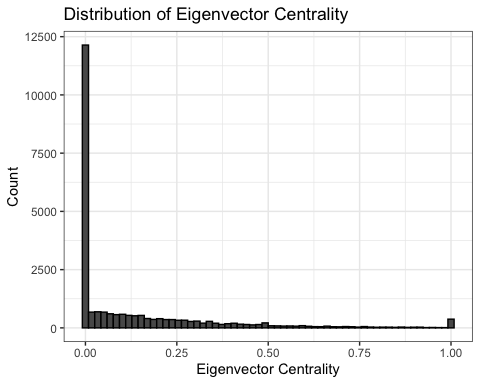
# Calculate Eigenvector Centrality

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 09-12-051 1 0  
## 2 08-51-339 1 0  
## 3 01-12-549 2 0  
## 4 08-21-089 2 0  
## 5 01-51-552 2 0  
## 6 08-80-066 2 0

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")

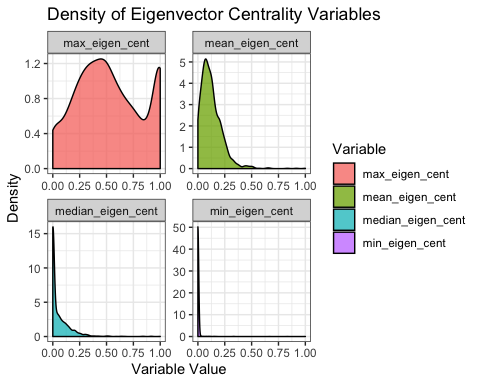


# Create Variables

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.

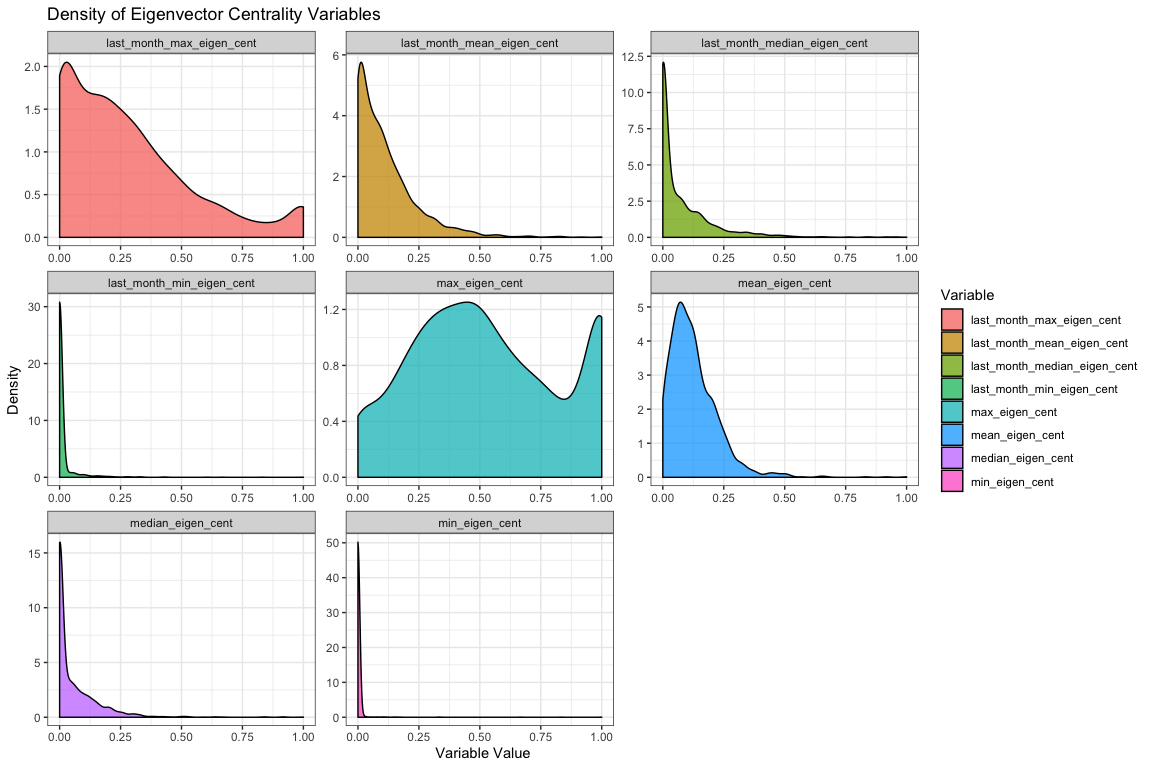
# min. eigen centrality -- highest position ever achieved in hierarchy  
# max. eigen centrality -- lowest position ever achieved in hierarchy  
# mean eigen centrality -- average position in the hierarchy  
# median eigen centrality -- another measure of central tendency  
nodal\_eigen <- eigen\_df %>%  
 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 mean\_eigen\_cent   
## Length:1727 Min. :0.000000 Min. :0.0000 Min. :0.00000   
## Class :character 1st Qu.:0.000000 1st Qu.:0.2995 1st Qu.:0.05773   
## Mode :character Median :0.000000 Median :0.5000 Median :0.10641   
## Mean :0.002926 Mean :0.5319 Mean :0.12546   
## 3rd Qu.:0.000000 3rd Qu.:0.7719 3rd Qu.:0.17196   
## Max. :1.000000 Max. :1.0000 Max. :1.00000   
## median\_eigen\_cent   
## Min. :0.000000   
## 1st Qu.:0.000000   
## Median :0.006291   
## Mean :0.058551   
## 3rd Qu.:0.088390   
## Max. :1.000000

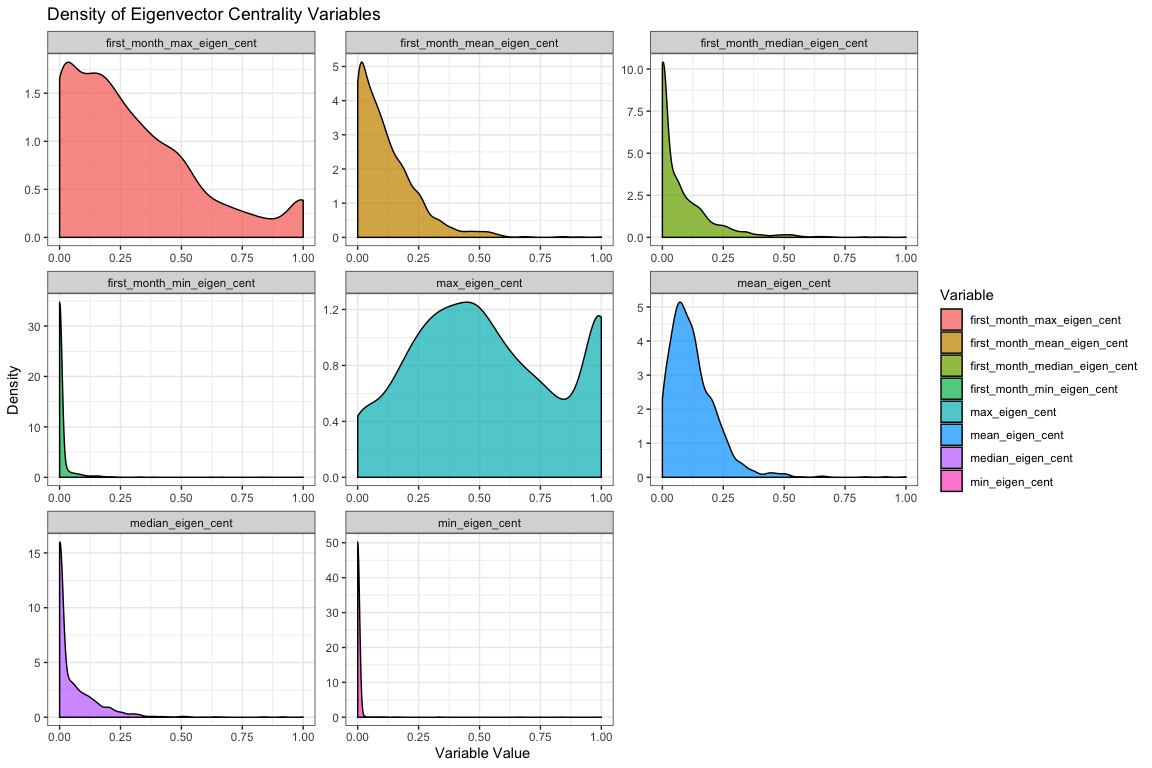
get\_last\_month\_variables <- function(id){  
 id\_df <- eigen\_df %>%   
 filter(Id == id) %>%   
 arrange(t) %>%   
 tail(4) %>%  
 summarize(Id = unique(Id),  
 last\_month\_min\_eigen\_cent = min(eigen\_cent),  
 last\_month\_max\_eigen\_cent = max(eigen\_cent),  
 last\_month\_mean\_eigen\_cent = 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:1727 Min. :0.00000 Min. :0.00000   
## Class :character 1st Qu.:0.00000 1st Qu.:0.06695   
## Mode :character Median :0.00000 Median :0.21580   
## Mean :0.01261 Mean :0.28215   
## 3rd Qu.:0.00000 3rd Qu.:0.41301   
## Max. :1.00000 Max. :1.00000   
## last\_month\_mean\_eigen\_cent last\_month\_median\_eigen\_cent  
## Min. :0.00000 Min. :0.00000   
## 1st Qu.:0.01948 1st Qu.:0.00000   
## Median :0.07894 Median :0.02172   
## Mean :0.11259 Mean :0.07755   
## 3rd Qu.:0.16128 3rd Qu.:0.10905   
## Max. :1.00000 Max. :1.00000

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:1727 Min. :0.00000 Min. :0.00000   
## Class :character 1st Qu.:0.00000 1st Qu.:0.09183   
## Mode :character Median :0.00000 Median :0.23299   
## Mean :0.01055 Mean :0.30077   
## 3rd Qu.:0.00000 3rd Qu.:0.44605   
## Max. :1.00000 Max. :1.00000   
## first\_month\_mean\_eigen\_cent first\_month\_median\_eigen\_cent  
## Min. :0.00000 Min. :0.00000   
## 1st Qu.:0.02801 1st Qu.:0.00000   
## Median :0.08605 Median :0.03318   
## Mean :0.11840 Mean :0.08088   
## 3rd Qu.:0.17470 3rd Qu.:0.11753   
## Max. :1.00000 Max. :1.00000

# 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/F3M"), stringsAsFactors = FALSE, header = TRUE)  
  
sum(!is.na(nodes$wcid))

## [1] 1852

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

## [1] 1839

# 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$wcid) > 1))  
  
nodes <- nodes[!(nodes$wcid %in% repeat\_visitors),]  
  
# join network variables  
# first rename Id to wcid  
dat <- nodes %>%  
 rename(Id = wcid) %>%  
 select(Id, age, lsi, lsiExit, 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 lsi lsiExit black success recidFlag recidDate gap  
## 1 01-11-515 21 29 29 0 1 0 09/04/2009 2858 days  
## 2 04-59-028 22 35 30 0 1 1 6/20/2006 687 days  
## 3 05-11-011 20 29 32 0 0 1 9/8/2005 27 days  
## 4 02-80-908 19 27 19 0 1 0 09/04/2009 2361 days  
## 5 06-21-234 22 12 10 0 1 1 8/16/2007 185 days  
## 6 06-80-184 31 13 26 0 0 1 1/4/2007 41 days  
## days\_in\_program last\_month\_min\_eigen\_cent last\_month\_max\_eigen\_cent  
## 1 176 0.000000 8.607192e-03  
## 2 167 0.000000 3.432536e-01  
## 3 134 0.281806 6.405873e-01  
## 4 169 0.000000 2.606015e-17  
## 5 125 0.000000 2.136412e-01  
## 6 101 0.000000 4.313504e-01  
## last\_month\_mean\_eigen\_cent last\_month\_median\_eigen\_cent  
## 1 2.151798e-03 4.978814e-17  
## 2 1.744843e-01 1.773417e-01  
## 3 4.889577e-01 5.167187e-01  
## 4 1.204915e-17 1.106823e-17  
## 5 5.341029e-02 1.750826e-17  
## 6 1.078376e-01 0.000000e+00  
## first\_month\_min\_eigen\_cent first\_month\_max\_eigen\_cent  
## 1 0 4.462972e-16  
## 2 0 3.432536e-01  
## 3 0 3.791498e-01  
## 4 0 1.814168e-02  
## 5 0 7.788422e-17  
## 6 0 2.371643e-01  
## first\_month\_mean\_eigen\_cent first\_month\_median\_eigen\_cent min\_eigen\_cent  
## 1 1.115743e-16 0.00000000 0  
## 2 1.372990e-01 0.10297111 0  
## 3 1.705537e-01 0.15153240 0  
## 4 4.535421e-03 0.00000000 0  
## 5 1.947105e-17 0.00000000 0  
## 6 1.011923e-01 0.08380245 0  
## max\_eigen\_cent mean\_eigen\_cent median\_eigen\_cent  
## 1 0.6038762 0.03780514 0.000000e+00  
## 2 0.3432536 0.11632284 7.437058e-02  
## 3 0.6405873 0.24293459 2.520082e-01  
## 4 0.6000000 0.07134812 1.452356e-17  
## 5 0.2136412 0.02870847 0.000000e+00  
## 6 0.4313504 0.07136968 1.134048e-16

# 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 = -9.0683, df = 1675, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.2614870 -0.1702229  
## sample estimates:  
## cor   
## -0.2163274

# 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 = -11.702, df = 1675, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.3185792 -0.2300643  
## sample estimates:  
## cor   
## -0.2749041

# 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 = -2.4051, df = 1675, p-value = 0.01627  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.10623470 -0.01082896  
## sample estimates:  
## cor   
## -0.05866579

# 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 = -11.94, df = 1675, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.3235943 -0.2353530  
## sample estimates:  
## cor   
## -0.2800651

# 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 +  
 lsiExit +  
 black +  
 days\_in\_program +  
 last\_month\_max\_eigen\_cent,  
 data = dat,  
 family = binomial(link = 'logit'))  
  
last\_month\_mean\_model <- glm(success ~  
 age +  
 lsiExit +  
 black +  
 days\_in\_program +  
 last\_month\_mean\_eigen\_cent,  
 data = dat,  
 family = binomial(link = 'logit'))  
  
max\_model <- glm(success ~  
 age +  
 lsiExit +  
 black +  
 days\_in\_program +  
 max\_eigen\_cent,  
 data = dat,  
 family = binomial(link = 'logit'))  
  
mean\_model <- glm(success ~  
 age +  
 lsiExit +  
 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) -6.00 \*\*\* -5.80 \*\*\* -6.33 \*\*\* -6.02 \*\*\*  
## (1.36) (1.36) (1.42) (1.44)   
## age 0.03 0.03 0.03 0.03   
## (0.02) (0.02) (0.02) (0.02)   
## lsiExit -0.24 \*\*\* -0.24 \*\*\* -0.24 \*\*\* -0.23 \*\*\*  
## (0.03) (0.03) (0.03) (0.03)   
## black 0.13 0.06 0.19 0.22   
## (0.42) (0.42) (0.44) (0.44)   
## days\_in\_program 0.10 \*\*\* 0.10 \*\*\* 0.11 \*\*\* 0.10 \*\*\*  
## (0.01) (0.01) (0.01) (0.01)   
## last\_month\_max\_eigen\_cent -2.10 \*\*\*   
## (0.54)   
## last\_month\_mean\_eigen\_cent -4.51 \*\*\*   
## (1.22)   
## max\_eigen\_cent -3.31 \*\*\*   
## (0.59)   
## mean\_eigen\_cent -9.58 \*\*\*  
## (1.68)   
## ------------------------------------------------------------------------------  
## AIC 346.57 347.79 324.19 326.50   
## BIC 377.73 378.95 355.35 357.66   
## Log Likelihood -167.28 -167.89 -156.09 -157.25   
## Deviance 334.57 335.79 312.19 314.50   
## Num. obs. 1331 1331 1331 1331   
## ==============================================================================  
## \*\*\* 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)),  
 lsiExit = rep(mean(data$lsiExit, na.rm = TRUE), nrow(data)),  
 black = rep(0, nrow(data)),  
 days\_in\_program = rep(mean(data$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

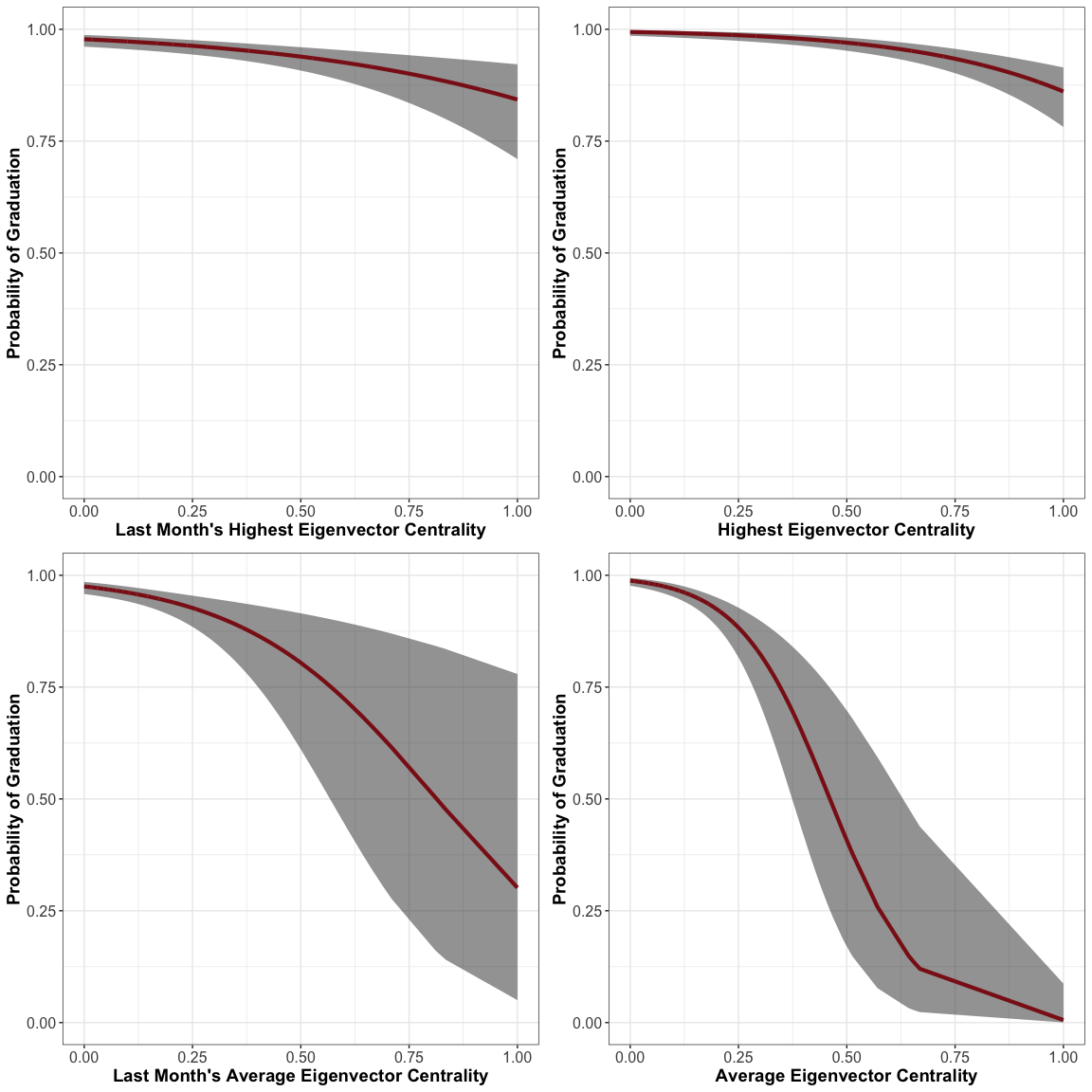
max\_eigen\_cent\_pred\_prob <- get\_pred\_prob\_plot(max\_model,   
 "max\_eigen\_cent",  
 "Highest Eigenvector Centrality")

## 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$age, na.rm = TRUE), length(input\_vec)),  
 lsiExit = rep(mean(data$lsiExit, 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] 0.9867252 0.9770179 0.9604962 0.9329128 0.8883092

# 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] 0.9860913 0.9761785 0.9594911 0.9319291 0.8878069

# 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] 0.9953066 0.9874153 0.9666999 0.9148267 0.7989536

# 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] 0.9935827 0.9842304 0.9617729 0.9102501 0.8034744

# Descriptive Stats

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

## Time difference of 3244 days

summary(dat)

## Id age lsi lsiExit   
## Length:1677 Min. :17.00 Min. : 0.00 Min. : 5.00   
## Class :character 1st Qu.:21.00 1st Qu.:23.00 1st Qu.:18.00   
## Mode :character Median :25.00 Median :28.00 Median :23.00   
## Mean :27.87 Mean :27.51 Mean :23.18   
## 3rd Qu.:33.00 3rd Qu.:32.00 3rd Qu.:28.00   
## Max. :60.00 Max. :45.00 Max. :99.00   
## NA's :2 NA's :344   
## black success recidFlag recidDate   
## Min. :0.0000 Min. :0.0000 Min. :0.000 Length:1677   
## 1st Qu.:0.0000 1st Qu.:1.0000 1st Qu.:0.000 Class :character   
## Median :0.0000 Median :1.0000 Median :1.000 Mode :character   
## Mean :0.1497 Mean :0.8318 Mean :0.523   
## 3rd Qu.:0.0000 3rd Qu.:1.0000 3rd Qu.:1.000   
## Max. :1.0000 Max. :1.0000 Max. :1.000   
##   
## gap days\_in\_program last\_month\_min\_eigen\_cent  
## Length:1677 Min. :-196.0 Min. :0.0000   
## Class :difftime 1st Qu.: 139.0 1st Qu.:0.0000   
## Mode :numeric Median : 153.0 Median :0.0000   
## Mean : 146.5 Mean :0.0112   
## 3rd Qu.: 167.0 3rd Qu.:0.0000   
## Max. : 179.0 Max. :1.0000   
## NA's :4   
## last\_month\_max\_eigen\_cent last\_month\_mean\_eigen\_cent  
## Min. :0.00000 Min. :0.00000   
## 1st Qu.:0.06803 1st Qu.:0.02047   
## Median :0.21580 Median :0.07835   
## Mean :0.28171 Mean :0.11147   
## 3rd Qu.:0.41125 3rd Qu.:0.16022   
## Max. :1.00000 Max. :1.00000   
##   
## last\_month\_median\_eigen\_cent first\_month\_min\_eigen\_cent  
## Min. :0.00000 Min. :0.000000   
## 1st Qu.:0.00000 1st Qu.:0.000000   
## Median :0.02193 Median :0.000000   
## Mean :0.07621 Mean :0.009373   
## 3rd Qu.:0.10826 3rd Qu.:0.000000   
## Max. :1.00000 Max. :1.000000   
##   
## first\_month\_max\_eigen\_cent first\_month\_mean\_eigen\_cent  
## Min. :0.00000 Min. :0.00000   
## 1st Qu.:0.09195 1st Qu.:0.02848   
## Median :0.23299 Median :0.08546   
## Mean :0.30096 Mean :0.11733   
## 3rd Qu.:0.44605 3rd Qu.:0.17468   
## Max. :1.00000 Max. :1.00000   
##   
## first\_month\_median\_eigen\_cent min\_eigen\_cent max\_eigen\_cent   
## Min. :0.00000 Min. :0.000000 Min. :0.0000   
## 1st Qu.:0.00000 1st Qu.:0.000000 1st Qu.:0.3000   
## Median :0.03302 Median :0.000000 Median :0.5000   
## Mean :0.07923 Mean :0.001721 Mean :0.5328   
## 3rd Qu.:0.11574 3rd Qu.:0.000000 3rd Qu.:0.7718   
## Max. :1.00000 Max. :1.000000 Max. :1.0000   
##   
## mean\_eigen\_cent median\_eigen\_cent   
## Min. :0.00000 Min. :0.000000   
## 1st Qu.:0.05808 1st Qu.:0.000000   
## Median :0.10625 Median :0.006119   
## Mean :0.12421 Mean :0.056998   
## 3rd Qu.:0.17158 3rd Qu.:0.087385   
## Max. :1.00000 Max. :1.000000   
##

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  
  
## 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   
## NA 8.62205142   
## lsi lsiExit   
## 6.65736119 7.55673847   
## black success   
## 0.35685607 0.37411758   
## recidFlag recidDate   
## 0.49962165 NA   
## gap days\_in\_program   
## NA 33.34880580   
## last\_month\_min\_eigen\_cent last\_month\_max\_eigen\_cent   
## 0.04913857 0.26559222   
## last\_month\_mean\_eigen\_cent last\_month\_median\_eigen\_cent   
## 0.12161764 0.12263435   
## first\_month\_min\_eigen\_cent first\_month\_max\_eigen\_cent   
## 0.04208583 0.27147365   
## first\_month\_mean\_eigen\_cent first\_month\_median\_eigen\_cent   
## 0.11894266 0.11788778   
## min\_eigen\_cent max\_eigen\_cent   
## 0.02664112 0.30068423   
## mean\_eigen\_cent median\_eigen\_cent   
## 0.09485261 0.09464261

# 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