Exercise 4

Konstantin Volodin

2023-04-03

Discussion

Without looking at interactions the following are results:

- Higher degree centrality decreases the time
- Higher betweenness centrality increases the time
- Higher closeness centrality decreases the time strongly
- Males tend to take increase the time
- “Other” and “White” races tend to decrease the time, while others increase.
- Having been in a position longer tends to decrease the time

When including interactions, the following are highlights:

- Higher degree increases time. However men with high degree take less time
- Betweenness now decreases the time, with men with high betweenness taking longer
- Closeness seems relatively similar, with men taking less time.
- Overall gender still has similar impact as before (when not looking at interactions)

```
model_complete
```

```
##
## Call:
## lm(formula = app_pro_time ~ degree + betweenness + closeness +
##     gender + race + tenure_days, data = applications_clean)
##
## Coefficients:
##  (Intercept)        degree   betweenness     closeness    gendermale
##     1276.298        -8.118        17.570       -44.719        23.439
##    raceblack  raceHispanic     raceother     racewhite   tenure_days
##       20.058         7.463       -10.638       -63.836       -24.799
```

```
model_interaction
```

```
##
## Call:
## lm(formula = app_pro_time ~ degree + betweenness + closeness +
##     gender + race + tenure_days + degree:gender + betweenness:gender +
##     closeness:gender, data = applications_clean)
##
## Coefficients:
```

```
##            (Intercept)                  degree             betweenness 
##               1274.768                   8.346                  -1.332 
##              closeness              gendermale               raceblack 
##                -37.324                  24.487                  18.430 
##           raceHispanic               raceother               racewhite 
##                  7.601                 -10.397                 -63.252 
##            tenure_days       degree:gendermale  betweenness:gendermale 
##                -24.364                 -20.742                  24.175 
##   closeness:gendermale 
##                 -9.813 
```

Code

```
library(here)
library(arrow)
library(lubridate)

library(tidyverse)
library(igraph)
library(tidygraph)
library(ggraph)
library(gridExtra)


### LOAD DATA
applications <- read_parquet(here('assignments','assignment_3',
                                  "clean_applications.parquet"))
edges <- read_parquet(here('assignments','assignment_3',
                           "clean_edges.parquet"))
examiner_data <- read_parquet(here('assignments','assignment_3',
                                   "clean_examiner.parquet"))


### CREATE NETWORK
edges <- edges %>%
  select(from = ego_examiner_id,
         to = alter_examiner_id, application_number) %>%
  drop_na()
nodes <- edges %>% gather() %>%
  filter(key %in% c('from', 'to')) %>%
  distinct(value) %>%
  select(name = value)
network <- graph_from_data_frame(edges, directed = TRUE, vertices=nodes) %>%
  as_tbl_graph()


### ESTIMATE CENTRALITY
network <- network %>%
  mutate(degree = centrality_degree(),
         betweenness = centrality_betweenness(),
         closeness = centrality_closeness())
centrality <- network %>%
  as.tibble() %>%
  mutate(name = as.numeric(name)) %>%
  rename(examiner_id = name)


### ADD CENTRALITY TO APPS
applications <- applications %>% left_join(centrality, on = 'examiner_id')
applications_clean <- applications %>%
  mutate(app_proc_time_issue = patent_issue_date - filing_date,
         app_proc_time_abandon = abandon_date - filing_date) %>%
  mutate(app_pro_time = ifelse(is.na(app_proc_time_issue),
                               app_proc_time_abandon, app_proc_time_issue)) %>%
  filter(app_pro_time > 0) %>%
  select(app_pro_time, gender, race, tenure_days, degree, betweenness, closeness) %>%
  drop_na() %>%
  mutate(degree = (degree - mean(degree)) / sd(degree)) %>%
  mutate(betweenness = (betweenness - mean(betweenness)) / sd(betweenness)) %>%
  mutate(closeness = (closeness  - mean(closeness)) / sd(closeness)) %>%
  mutate(tenure_days = (tenure_days - mean(tenure_days)) / sd(tenure_days))


### REGRESSION
model_complete <- lm(app_pro_time ~ degree + betweenness + closeness +
                     gender + race + tenure_days, data = applications_clean)
model_interaction <- lm(app_pro_time ~ degree + betweenness + closeness +
                          gender + race + tenure_days +
                          degree:gender +
                          betweenness:gender +
                          closeness:gender, data = applications_clean)
```