MSD 2019 Final Project

A replication and extension of < PAPER TITLE > by < ORIGINAL AUTHORS >, < PUBLISHED IN >

Your Names (your unis)

2019-05-05 15:35:42

Contents

```r
library(tidyverse)
```

```
## -- Attaching packages -------------------------------------------------------------
## v ggplot2 2.2.1      v purrr   0.3.0
## v tibble  2.0.1      v dplyr   0.8.0.1
## v tidyr   0.8.1      v stringr 1.3.1
## v readr   1.1.1      v forcats 0.3.0

## Warning: package 'tibble' was built under R version 3.4.4

## Warning: package 'tidyr' was built under R version 3.4.4

## Warning: package 'purrr' was built under R version 3.4.4

## Warning: package 'dplyr' was built under R version 3.4.4

## -- Conflicts ----------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(modelr)
```

```
## Warning: package 'modelr' was built under R version 3.4.4
```

```r
library(ggplot2)
library(igraph)
```

```
## Warning: package 'igraph' was built under R version 3.4.4

##
## Attaching package: 'igraph'

## The following object is masked from 'package:modelr':
##
##     permute

## 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
```

Make a loop that uses the same code for the three departments

```r
top15 = data.frame()
rest = data.frame()


for( dep in c('ComputerScience','Business', 'History' ) ){

#Reading the tables

edgelist = read.table( paste(dep, '_edgelist.txt', sep = ""), header = FALSE,  col.names = c('u', 'v',

vertex = read.table(file  = paste(dep, '_vertexlist.txt', sep = ""), sep = '\t', header = FALSE, col.nam



#making a table that includes the weight of each edge
weighted_edgelist = edgelist %>%
  group_by(v, u) %>%
  summarize( count = n()) %>%
  ungroup() %>%
  left_join( vertex, by = c('v'= 'u')) %>%
  select(v, u, count, institution)


#fitlering the weighted edgelist to make an easier to look at plot (like in Fig. 1)
smaller = weighted_edgelist %>%
  filter( u <= 10 , v <= 10 )

#Then plotting this network of the top schools
smaller_graph = smaller %>%
   graph_from_data_frame(directed = TRUE)


plot(smaller_graph, vertex_size = 2 ,edge.width=E(smaller_graph)$count/2,
     layout = layout_in_circle(smaller_graph, order = V(smaller_graph)),
     vertex.label = unique( E(smaller_graph)$institution ),
     main = paste(dep, 'Department', sep = " ")
     )



num_schools = max(edgelist$u)
```

```r
#making another set of the full network to make a network plot like in Fig. 3)

power_list = weighted_edgelist %>%
  filter(v != num_schools, u != num_schools) %>%
  group_by(v) %>%
  summarize( top_school =  as.double( v <= 0.15*num_schools)[1],
  power = num_schools - v[1] ) %>%
  ungroup()


#Setting up the network to plot Fig. 3
graph = weighted_edgelist %>%
  filter( u != v, u %in%power_list$v, v %in% power_list$v  ) %>%
  graph_from_data_frame(directed = FALSE, vertices = power_list    )


plot( graph, vertex.size = 2 + 3*V(graph)$top_school + V(graph)$power/15,
      vertex.color = 3 + V(graph)$top_school,
      vertex.label = NA,
      main = paste(dep, 'Department', sep = " ")
      )




#making dataframes of the top 15 of institutions with the differences in prestige from phd to faculty s
#This is to make the density plots in Fig. 3
#Am doing rbing to keep data from all the departments, but addign the label of department first

edgelist$dep = dep


top15 = rbind(top15,  edgelist %>%
  filter(u <= .15*num_schools) %>%
  mutate(diff = (v - u)/num_schools) %>%
  select(diff, dep)
  )

#doing the same thing for the rest of the institutions
rest = rbind(rest,  edgelist %>%
  filter(u > .15*num_schools) %>%
  filter( u < num_schools ) %>%
  mutate(diff = (v - u)/num_schools) %>%
  select(diff, dep)
  )


}
```
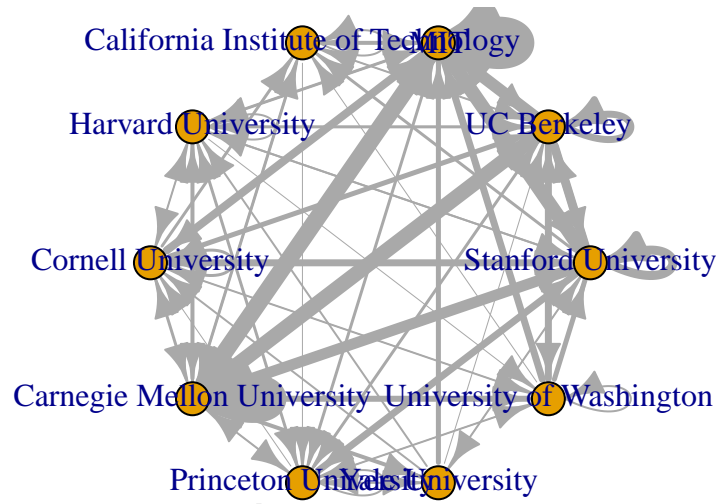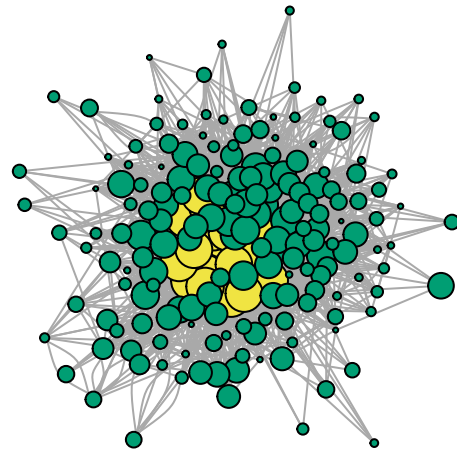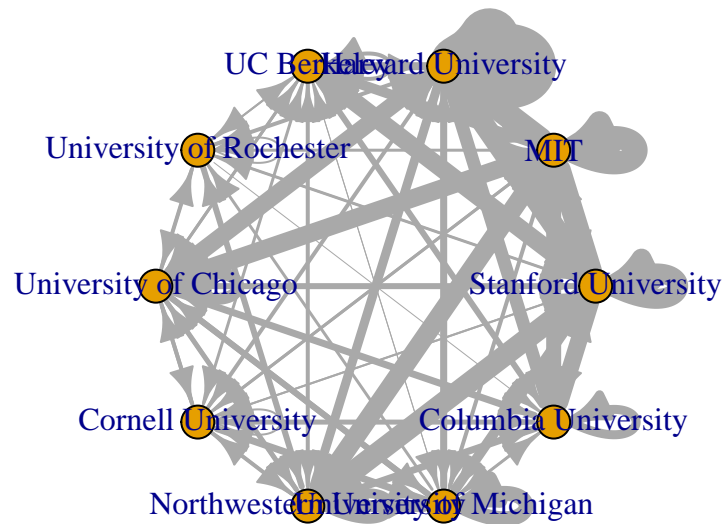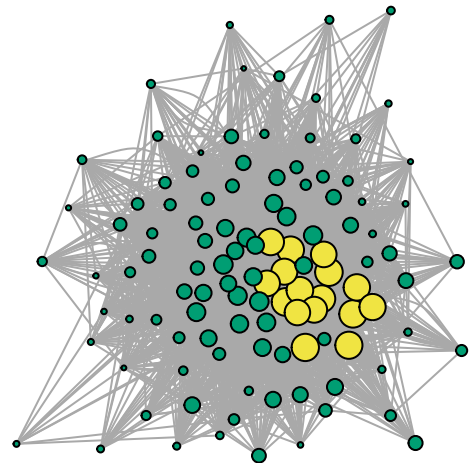
ComputerScience Department
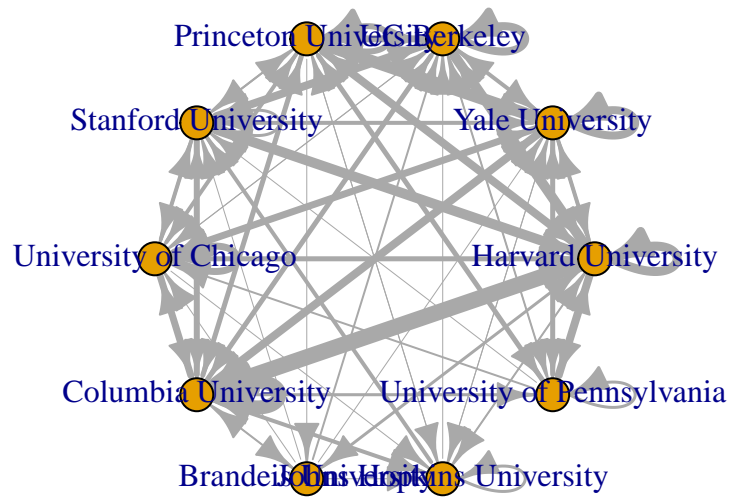

ComputerScience Department
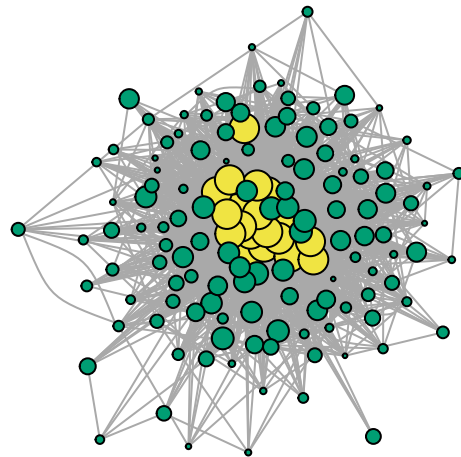

Business Department

Business Department
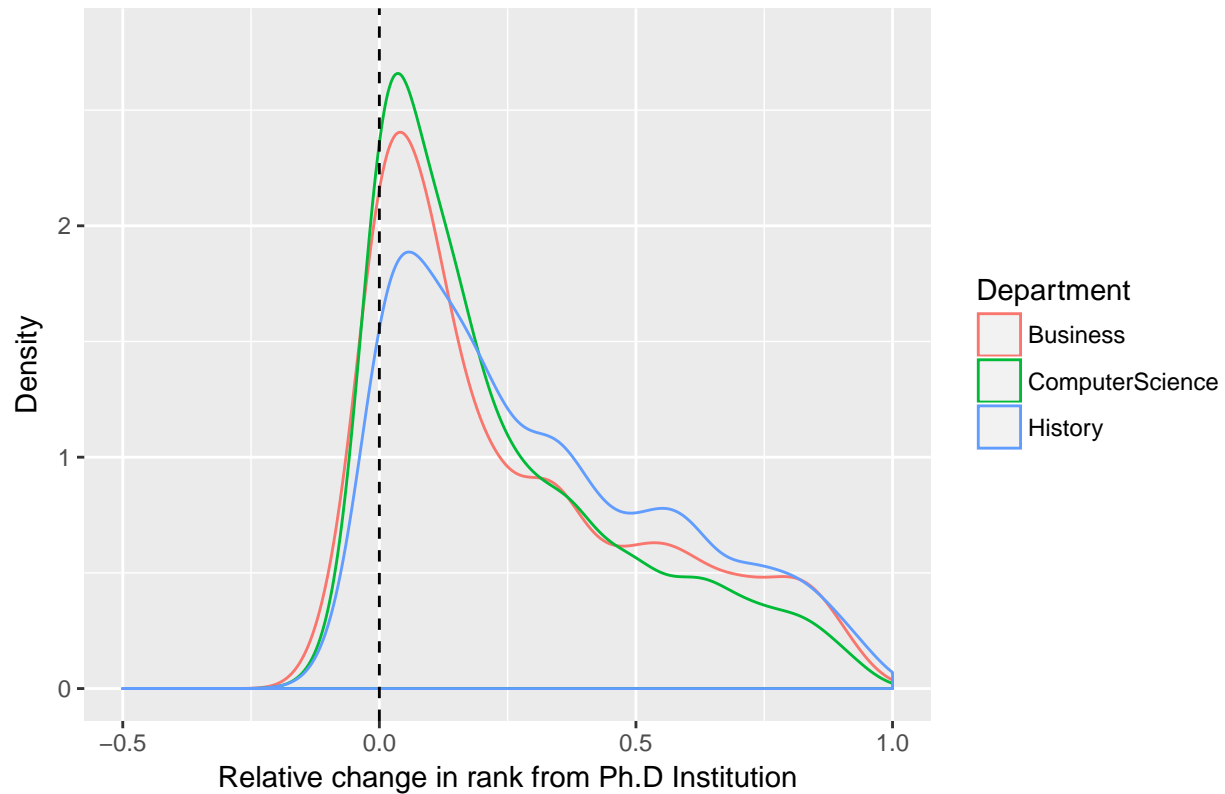
History Department

## History Department


```
#Making the density plots here

top15 %>%
  ggplot(aes(x = diff, color = dep) ) +
  geom_density() +
  geom_vline(xintercept = 0, linetype = 'dashed' ) +
  ylim(0, 2.8) +
  xlim(-.5, 1) +
  ggtitle("Top 15% of Institutions") +
  ylab('Density') +
  xlab('Relative change in rank from Ph.D Institution') +
  labs(color = 'Department')
```

Top 15% of Institutions


```
rest %>%
  ggplot(aes(x = diff, color = dep), ylim = c(0, 2.8)) +
  geom_density() +
  geom_vline(xintercept = 0, linetype = 'dashed' ) +
  ylim(0, 2.8) +
  xlim(-.5, 1) +
  ggtitle("All other Institutions") +
  ylab('Density') +
  xlab('Relative change in rank from Ph.D Institution') +
  labs(color = 'Department')
```

```
## Warning: Removed 36 rows containing non-finite values (stat_density).
```

All other Institutions
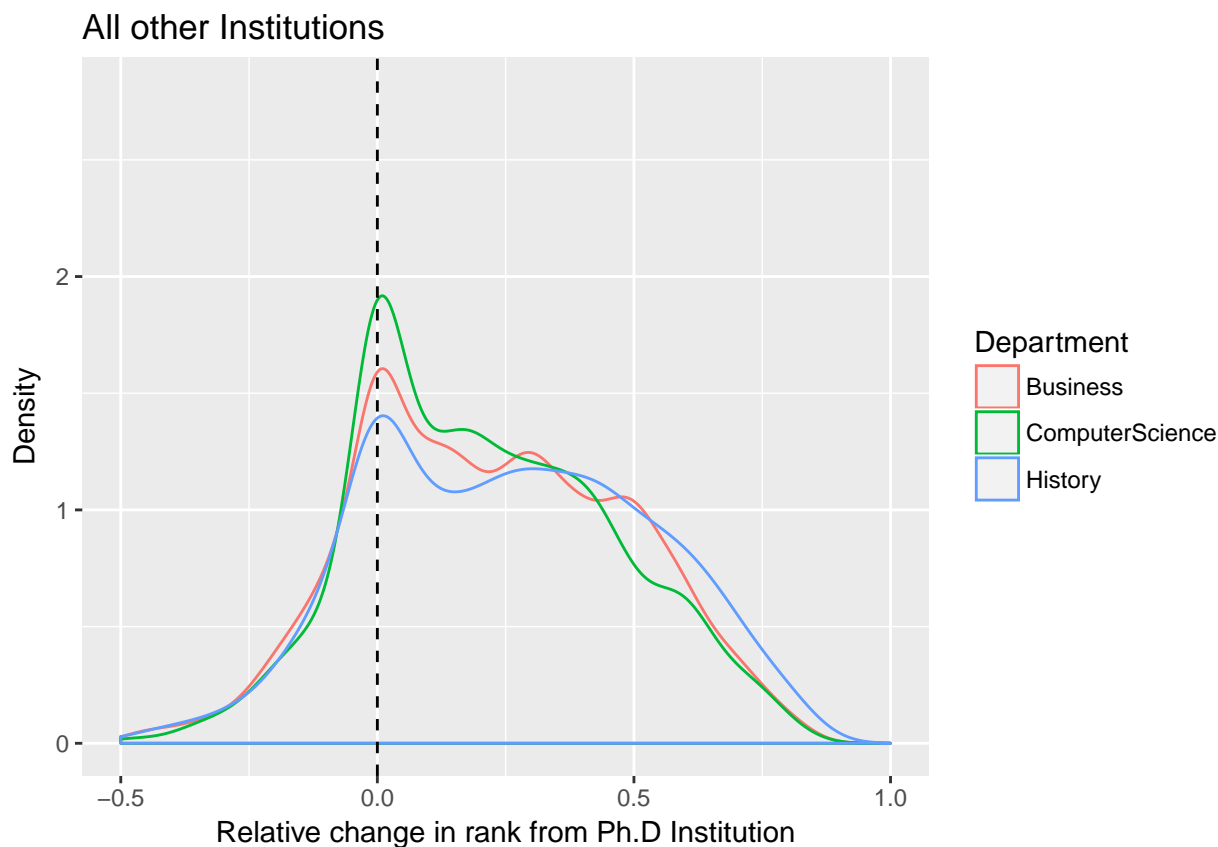
The following is a list of all packages used to generate these results. (Leave at very end of file.)

```
sessionInfo()
```

```
## R version 3.4.3 (2017-11-30)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS Sierra 10.12.6
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/3.4/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.4/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] igraph_1.2.4    modelr_0.1.4    forcats_0.3.0   stringr_1.3.1
##  [5] dplyr_0.8.0.1   purrr_0.3.0     readr_1.1.1     tidyr_0.8.1
##  [9] tibble_2.0.1    ggplot2_2.2.1   tidyverse_1.2.1
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.0       cellranger_1.1.0 pillar_1.3.1     compiler_3.4.3
##  [5] plyr_1.8.4       tools_3.4.3      digest_0.6.15    lubridate_1.7.4
##  [9] jsonlite_1.5     evaluate_0.10.1  nlme_3.1-131     gtable_0.2.0
## [13] lattice_0.20-35  pkgconfig_2.0.2  rlang_0.3.1      cli_1.0.1
```

```
## [17] rstudioapi_0.7    yaml_2.1.18      haven_1.1.2    xml2_1.2.0
## [21] httr_1.3.1        knitr_1.20       hms_0.4.2      rprojroot_1.3-2
## [25] grid_3.4.3        tidyselect_0.2.5 glue_1.3.0     R6_2.2.2
## [29] readxl_1.1.0      rmarkdown_1.10   magrittr_1.5   backports_1.1.2
## [33] scales_0.5.0      htmltools_0.3.6  rvest_0.3.2    assertthat_0.2.0
## [37] colorspace_1.3-2 labeling_0.3     stringi_1.2.2  lazyeval_0.2.1
## [41] munsell_0.4.3     broom_0.5.0      crayon_1.3.4
```