Assignment1

Load packages

```
library("tidyverse")
```

```
## Warning: package 'tidyverse' was built under R version 4.1.3
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1
```

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

```
library("igraph")
```

```
## Warning: package 'igraph' was built under R version 4.1.3
```

```
##
## Attaching package: 'igraph'
```

```
## 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
```

```
library("tidygraph")
```

```
## Warning: package 'tidygraph' was built under R version 4.1.3
```

```
##
## Attaching package: 'tidygraph'
```

```
## The following object is masked from 'package:igraph':
##
##     groups
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
library("ggraph")
```

```
## Warning: package 'ggraph' was built under R version 4.1.3
```

Upload dataset

```
network <- read_csv("C:/Users/0/OneDrive/McGill - Summer 2022/ORGB 672 - Org Network Analysis/Data/Conne
```

```
## Rows: 512 Columns: 6
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (6): First Name, Last Name, Email Address, Company, Position, Connected On
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
network %>% head(5)
```

```
## # A tibble: 5 x 6
##   `First Name` `Last Name`  `Email Address`  Company   Position   `Connected On`
##   <chr>        <chr>        <chr>            <chr>     <chr>      <chr>
## 1 Ayman        Mahin Gostar 19aymanpm@gmail~ EDF Rene~ Renewable~ 26 Apr 2022
## 2 Fahid        Hasin        <NA>             Standard~ Payment &~ 22 Apr 2022
## 3 Franck       Benichou, M~ <NA>             Intact    Data Scie~ 22 Apr 2022
## 4 Uzair        Ahmad        <NA>             Intellig~ Researcher 22 Apr 2022
## 5 Kristen      Chen         <NA>             McGill U~ Undergrad~ 19 Apr 2022
```

Get the total count of employers

```
network %>%
  count(Company) %>%
  arrange(-n)
```

```
## # A tibble: 392 x 2
##    Company                                                 n
##    <chr>                                               <int>
##  1 McGill University - Desautels Faculty of Management    18
##  2 <NA>                                                   18
##  3 Global Affairs Canada | Affaires mondiales Canada       9
##  4 Rogers Communications                                   6
##  5 Scotiabank                                              6
##  6 University of Waterloo                                  6
##  7 Air Transat                                             5
##  8 Novartis                                                5
##  9 Sia Partners                                            5
## 10 TJX Canada/Winners Merchants International L.P.         5
## # ... with 382 more rows
```

Nodes dataframe

```
# remove extra characters from last name and create a unique label
network$last_initial <- substr(network$`Last Name`, 1, 1)
network$label <- paste(network$`First Name`, network$last_initial)

# filter for only the unique node labels
nodes <- network %>% distinct(label)
nodes <- nodes %>% rowid_to_column('Main_ID')
nodes %>% head(5)
```

```
## # A tibble: 5 x 2
##   Main_ID label
##     <int> <chr>
## 1       1 Ayman M
## 2       2 Fahid H
## 3       3 Franck B
## 4       4 Uzair A
## 5       5 Kristen C
```

Edges Dataframe

```
# create a duplicate dataframe and merge with original
duplicate <- network
colnames(duplicate) <- paste(colnames(duplicate), "2", sep="")
join <- tidyr::crossing(network, duplicate, .name_repair="minimal")

# remove contacts connected to themselves and join on unique companies
edges <- filter(join, join$Company == join$Company2 & join$label != join$label2)
edges <- edges %>% select(label, Company, label2, Company2)
edges <- edges %>%
  left_join(nodes, by = c("label" = "label")) %>%
  rename(Node_1 = Main_ID)
```

```
edges <- edges %>%
  left_join(nodes, by = c("label2" = "label")) %>%
  rename(Node_2 = Main_ID)

# final edges dataframe
edges <- select(edges, Node_1, Node_2)
edges %>% head(5)
```

```
## # A tibble: 5 x 2
##   Node_1 Node_2
##    <int>  <int>
## 1    433    269
## 2    433    324
## 3     19    307
## 4     19     89
## 5    318    304
```

Fit the model

```
final_network <- tbl_graph(nodes=nodes, edges=edges, directed=FALSE)
final_network
```

```
## # A tbl_graph: 506 nodes and 670 edges
## #
## # An undirected multigraph with 403 components
## #
## # Node Data: 506 x 2 (active)
##   Main_ID label
##     <int> <chr>
## 1       1 Ayman M
## 2       2 Fahid H
## 3       3 Franck B
## 4       4 Uzair A
## 5       5 Kristen C
## 6       6 Sheiva A
## # ... with 500 more rows
## #
## # Edge Data: 670 x 2
##    from    to
##   <int> <int>
## 1   269   433
## 2   324   433
## 3    19   307
## # ... with 667 more rows
```

Plot the final graph

```
ggraph(final_network) + geom_edge_link() + geom_node_point() + theme_graph()
```

```
## Using ‘stress‘ as default layout
```