A1

2024-03-19

```r
Connections = read.csv("/Users/apple/Desktop/Network Analysis/A1/Connections.csv", sep=",", skip=3)
attach(Connections)
```

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(stringr)
library(tidyr)
library(igraph)
```

```
## Warning: package 'igraph' was built under R version 4.3.2

##
## Attaching package: 'igraph'

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

## The following objects are masked from 'package:dplyr':
##
##     as_data_frame, groups, union

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

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

```r
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.2

##
## Attaching package: 'ggplot2'

## The following object is masked from 'Connections':
##
##     Position
```

```r
# Count of connections by company, sorted by count
number_of_connections <- Connections %>%
  group_by(Company) %>%
  summarise(Count = n()) %>%
  arrange(desc(Count))

# Print the number of connections by company
print(number_of_connections)
```

```
## # A tibble: 89 x 2
##    Company                                          Count
##    <chr>                                            <int>
##  1 "McGill University - Desautels Faculty of Management"  9
##  2 ""                                                   7
##  3 "McGill University"                                  7
##  4 "BRP"                                                4
##  5 "CN"                                                 4
##  6 "Air Transat"                                        3
##  7 "AtkinsRéalis"                                       3
##  8 "BOMBARDIER"                                         3
##  9 "CAE"                                                3
## 10 "CGI"                                                3
## # i 79 more rows
```

```r
# Total number of connections
total_num <- nrow(Connections)

# Print the total number of connections
print(total_num)
```

```
## [1] 142
```

```r
# Renaming columns to remove spaces
Connections <- Connections %>%
  rename(FirstName = `First.Name`, LastName = `Last.Name`)

library(purrr)
```

```
##
## Attaching package: 'purrr'
```

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

```r
Connections$Label <- with(Connections, paste(FirstName, substr(LastName, 1, 1)))

Connections <- Connections %>%
  mutate(ID = row_number())

# Nodes dataframe now uses ID, Label and Company
nodes <- Connections %>%
  distinct(ID, Label, Company)

linkedin_data_with_ids <- Connections %>%
  left_join(nodes, by = c("Label", "Company"))
```

```
## Warning in left_join(., nodes, by = c("Label", "Company")): Detected an unexpected many-to-many rela
## i Row 13 of `x` matches multiple rows in `y`.
## i Row 13 of `y` matches multiple rows in `x`.
## i If a many-to-many relationship is expected, set `relationship =
##   "many-to-many"` to silence this warning.
```

```r
# Create edges based on these IDs within the same company
edges <- nodes %>%
  group_by(Company) %>%
  filter(n() > 1) %>%
  summarise(Combo = list(combn(ID, 2, simplify = FALSE))) %>%
  unnest(Combo) %>%
  ungroup() %>%
  mutate(From = sapply(Combo, `[`, 1),
         To = sapply(Combo, `[`, 2)) %>%
  select(From, To)

# View the edges dataframe
print(edges)
```
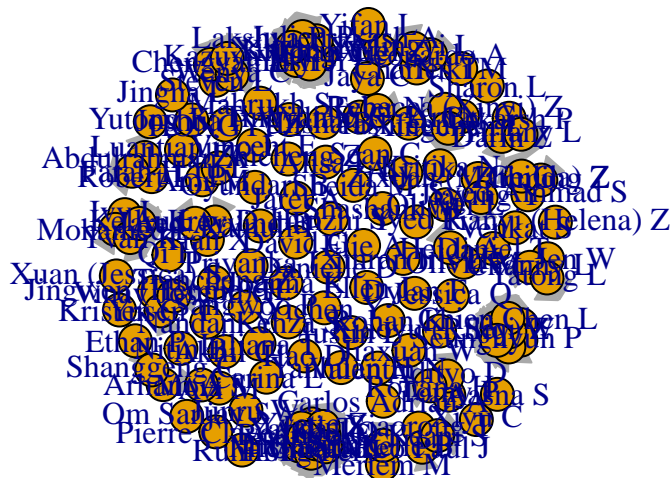
```
## # A tibble: 127 x 2
##      From    To
##     <int> <int>
## 1     13    27
## 2     13    40
## 3     13    47
## 4     13    55
## 5     13    68
## 6     13   121
## 7     27    40
## 8     27    47
## 9     27    55
## 10    27    68
## # i 117 more rows
```

```r
library(ggraph)
```

```
## Warning: package 'ggraph' was built under R version 4.3.2
```

```r
# Create graph from edges dataframe, using the updated nodes and labels
g <- graph_from_data_frame(d = edges, vertices = nodes, directed = TRUE)
# Plot the graph
plot(g, vertex.label = V(g)$Label)
```

```r
#Edges are based on your contacts being affiliated with Mcgill
# Create a new column 'McGill'
nodes <- nodes %>%
  mutate(McGill = ifelse(str_detect(Company, "McGill"), "McGill", "Other"))
nodes$ID <- as.character(nodes$ID)

# Generate layout
g_M <- graph_from_data_frame(d = edges, vertices = nodes, directed = FALSE)
layout <- as.data.frame(layout_with_fr(g_M))
names(layout) <- c("x", "y")
layout$ID <- V(g_M)$name

# Add McGill information to the layout
layout <- layout %>%
  left_join(nodes %>% select(ID, McGill), by = "ID")

edges$From <- as.character(edges$From)
edges$To <- as.character(edges$To)
# Join edge start positions
edges_coords <- edges %>%
  left_join(layout %>% select(ID, x_start = x, y_start = y), by = c("From" = "ID"))
# Join edge end positions
edges_coords <- edges_coords %>%
  left_join(layout %>% select(ID, x_end = x, y_end = y), by = c("To" = "ID"))

# Plotting
ggplot() +
  geom_segment(data = edges_coords, aes(x = x_start, y = y_start, xend = x_end, yend = y_end), color = 
  geom_point(data = layout, aes(x = x, y = y, color = McGill), size = 4) +
  scale_color_manual(values = c("McGill" = "red", "Other" = "blue")) +
  theme_void() +
  theme(legend.position = "right") +
  labs(color = "McGill")
```

McGill

- McGill
- Other