Org_Net_Ex_1

2024-03-14

```r
options(warn = -1)

connections <- read.csv("E:/Users/pc/Downloads/Connections_Sohail.csv", skip = 2)

Connections_Sohail = connections

library(tidygraph)
```

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

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

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.5.0     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks tidygraph::filter(), stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
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 object is masked from 'package:tidygraph':
##
##     groups
##
## The following objects are masked from 'package:stats':
##
##     decompose, spectrum
##
## The following object is masked from 'package:base':
##
##     union
```

```r
library(ggplot2)
library(vroom)
```

```
##
## Attaching package: 'vroom'
##
## The following objects are masked from 'package:readr':
##
##     as.col_spec, col_character, col_date, col_datetime, col_double,
##     col_factor, col_guess, col_integer, col_logical, col_number,
##     col_skip, col_time, cols, cols_condense, cols_only, date_names,
##     date_names_lang, date_names_langs, default_locale, fwf_cols,
##     fwf_empty, fwf_positions, fwf_widths, locale, output_column,
##     problems, spec
```

```r
library(scales)
```

```
##
## Attaching package: 'scales'
##
## The following object is masked from 'package:vroom':
##
##     col_factor
##
## The following object is masked from 'package:purrr':
##
##     discard
##
## The following object is masked from 'package:readr':
```

```
##
##      col_factor
```

```
library(ggraph)
library(ggtext)
library(ggrepel)
library(ggforce)
library(ggthemes)
library(patchwork)
library(qualpalr)
library(viridis)
```

```
## Loading required package: viridisLite
##
## Attaching package: 'viridis'
##
## The following object is masked from 'package:scales':
##
##      viridis_pal
```

```
library(oaqc)

#Before Removing missing values

# Count the number of contacts by their current employer
counts_by_employer <- Connections_Sohail %>%
  group_by(Company) %>%
  summarise(count = n())

# Calculate the total count of contacts
total_count <- sum(counts_by_employer$count)

# Print the counts by employer
print(counts_by_employer)
```

```
## # A tibble: 1,135 x 2
##    Company                    count
##    <chr>                      <int>
##  1 ""                            73
##  2 " Warren Insurance Group "     1
##  3 "28 Global Pvt. Ltd"           1
##  4 "5Y Capital"                   1
##  5 "9 Mark Communications"        1
##  6 "9D Technologies"              1
##  7 "AEDesign (pvt) Ltd"           1
##  8 "AFRY"                         1
##  9 "AIESEC in Pakistan"           1
## 10 "AIN Engineering Services"     1
## # i 1,125 more rows
```

```
# Print the total count
print(paste("Total count of contacts:", total_count))
```

```
## [1] "Total count of contacts: 1813"
```

```r
Connections_Sohail <- select(Connections_Sohail, -'Email.Address')
```

```r
Connections_Sohail <- Connections_Sohail %>%
  mutate(
    label = str_c(
      First.Name,
      str_sub(Last.Name, 1, 1),
      as.character(row_number())
    )
  ) %>%
  select(label, Company) %>%
  filter(!is.na(Company) & Company != "")   # Filter out NA and empty strings
```

#After removing missing values

```r
# Count the number of contacts by their current employer
counts_by_employer <- Connections_Sohail %>%
  group_by(Company) %>%
  summarise(count = n())

# Calculate the total count of contacts
total_count <- sum(counts_by_employer$count)

# Print the counts by employer
print(counts_by_employer)
```

```
## # A tibble: 1,134 x 2
##    Company                   count
##    <chr>                     <int>
##  1 " Warren Insurance Group "     1
##  2 "28 Global Pvt. Ltd"           1
##  3 "5Y Capital"                   1
##  4 "9 Mark Communications"        1
##  5 "9D Technologies"              1
##  6 "AEDesign (pvt) Ltd"           1
##  7 "AFRY"                         1
##  8 "AIESEC in Pakistan"           1
##  9 "AIN Engineering Services"     1
## 10 "ALDO Group"                   5
## # i 1,124 more rows
```

```r
# Print the total count
print(paste("Total count of contacts:", total_count))
```

```
## [1] "Total count of contacts: 1740"
```

#Summarising

4

```r
companies_with_more_than_5 <- Connections_Sohail %>%
  group_by(Company) %>%
  count(sort = TRUE) %>%
  filter(n > 5)

companies_with_more_than_5
```

```
## # A tibble: 37 x 2
## # Groups:   Company [37]
##    Company                                           n
##    <chr>                                         <int>
##  1 McGill University - Desautels Faculty of Management    38
##  2 Desautels Capital Management                     30
##  3 McGill University                                26
##  4 National University of Sciences and Technology (NUST)    19
##  5 BAT                                              17
##  6 CureMD                                           15
##  7 TD                                               12
##  8 S&P Global                                       11
##  9 Jazz                                             10
## 10 RBC                                              10
## # i 27 more rows
```

#Histogram

```r
ggplot(companies_with_more_than_5, aes(x = reorder(Company, n), y = n, fill = Company)) +
  geom_bar(stat = "identity",width=0.8) +
  coord_flip() +
  labs(
    x = "Count of Contacts",
    y = "Company",
    title = "Companies with More Than 5 Contacts",
    subtitle = "Horizontal bar chart showing companies with more than 5 contacts"
  ) +
  theme_minimal() +
  scale_fill_viridis_d() +
  theme(
    legend.position = "none",
    axis.text.y = element_text(size = 8),  # Adjust font size of y-axis labels
    plot.subtitle = element_text(size = 12)
  )
```

Companies with More Than 5 Contacts

Horizontal bar chart showing companies with
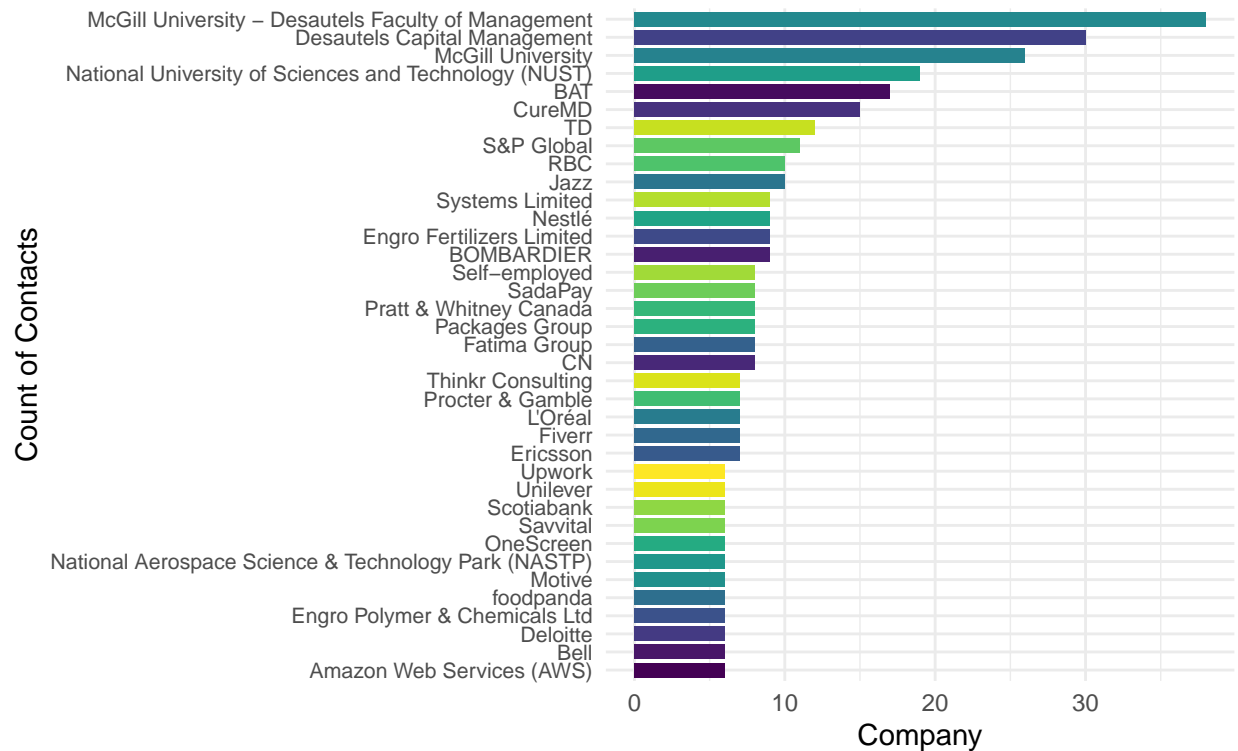

```r
# Preparing a dataframe for graph creation
df_connections <- Connections_Sohail %>%
  filter(Company %in% companies_with_more_than_5$Company) %>%
  group_by(Company) %>%
  summarise(label_combinations = list(combn(label, 2, simplify = FALSE))) %>%
  unnest(label_combinations) %>%
  transmute(
    from = map_chr(label_combinations, 1),
    to = map_chr(label_combinations, 2),
    company = Company
  )

# View the resulting dataframe
print(df_connections)
```

```
## # A tibble: 2,683 x 3
##    from          to            company
##    <chr>         <chr>         <chr>
##  1 SeungyeonH310 SebastienP313 Amazon Web Services (AWS)
##  2 SeungyeonH310 SungbaeP318   Amazon Web Services (AWS)
##  3 SeungyeonH310 NikosV325     Amazon Web Services (AWS)
##  4 SeungyeonH310 VincentR332   Amazon Web Services (AWS)
##  5 SeungyeonH310 AnamZ1514     Amazon Web Services (AWS)
##  6 SebastienP313 SungbaeP318   Amazon Web Services (AWS)
##  7 SebastienP313 NikosV325     Amazon Web Services (AWS)
##  8 SebastienP313 VincentR332   Amazon Web Services (AWS)
```

```
##  9 SebastienP313 AnamZ1514      Amazon Web Services (AWS)
## 10 SungbaeP318   NikosV325      Amazon Web Services (AWS)
## # i 2,673 more rows
```

```r
# Setting seed for reproducibility
set.seed(180)

# Sampling 30% of the connections and creating a graph
graph_data <- df_connections %>%
  slice_sample(prop = 0.30) %>%
  as_tbl_graph(directed = FALSE)

# Print the graph object to view its summary
print(graph_data)
```

```
## # A tbl_graph: 357 nodes and 804 edges
## #
## # An undirected simple graph with 42 components
## #
## # A tibble: 357 x 1
##   name
##   <chr>
## 1 RoushniA834
## 2 MaazZ1417
## 3 AyeshaS792
## 4 Syed AliM832
## 5 Sania KhanT16
## 6 BrittannyJ161
## # i 351 more rows
## #
## # A tibble: 804 x 3
##    from    to company
##   <int> <int> <chr>
## 1     1   257 Upwork
## 2     2   258 SadaPay
## 3     3   214 Fatima Group
## # i 801 more rows
```

```r
# Generating a color palette
color_palette <- qualpal(
  n = graph_data %>%
    activate(edges) %>%
    pull(company) %>%
    unique() %>%
    length(), # Number of unique companies in the graph's edges
  colorspace = "pretty" # Specifies the color space for the palette
)
```
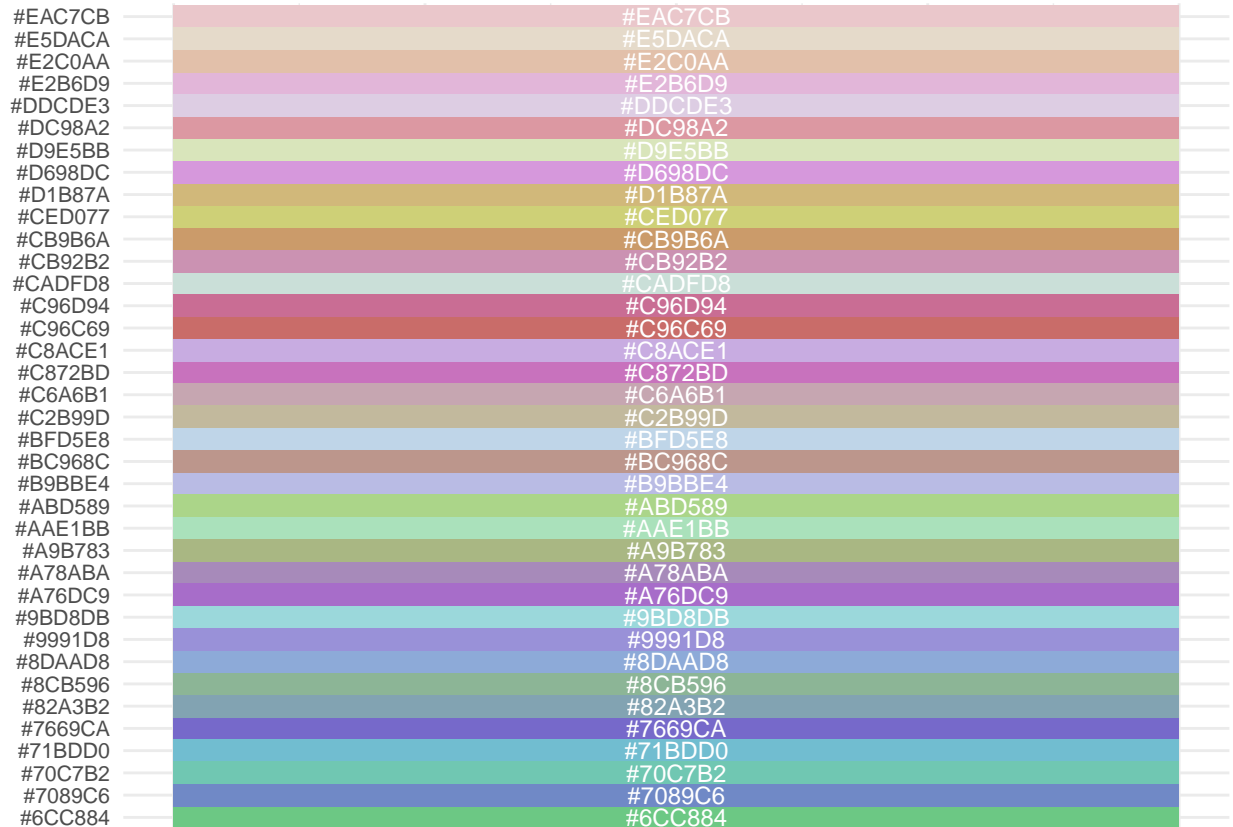
```r
# Create a data frame for plotting
palette_df <- data.frame(colors = color_palette$hex)

# Create a plot of the color palette with custom text sizes
ggplot(palette_df, aes(x = 1, y = colors, fill = colors, label = colors)) +
```

```r
  geom_tile() +
  geom_text(color = "white", size = 3) + # Adjust text size here
  scale_fill_identity() +
  theme_minimal() +
  theme(axis.text.x = element_blank(),
        axis.text.y = element_text(size = 8), # Adjust axis text size here
        axis.ticks = element_blank(),
        axis.title = element_blank(),
        legend.position = "none")
```

#EAC7CB	#EAC7CB
#E5DACA	#E5DACA
#E2C0AA	#E2C0AA
#E2B6D9	#E2B6D9
#DDCDE3	#DDCDE3
#DC98A2	#DC98A2
#D9E5BB	#D9E5BB
#D698DC	#D698DC
#D1B87A	#D1B87A
#CED077	#CED077
#CB9B6A	#CB9B6A
#CB92B2	#CB92B2
#CADFD8	#CADFD8
#C96D94	#C96D94
#C96C69	#C96C69
#C8ACE1	#C8ACE1
#C872BD	#C872BD
#C6A6B1	#C6A6B1
#C2B99D	#C2B99D
#BFD5E8	#BFD5E8
#BC968C	#BC968C
#B9BBE4	#B9BBE4
#ABD589	#ABD589
#AAE1BB	#AAE1BB
#A9B783	#A9B783
#A78ABA	#A78ABA
#A76DC9	#A76DC9
#9BD8DB	#9BD8DB
#9991D8	#9991D8
#8DAAD8	#8DAAD8
#8CB596	#8CB596
#82A3B2	#82A3B2
#7669CA	#7669CA
#71BDD0	#71BDD0
#70C7B2	#70C7B2
#7089C6	#7089C6
#6CC884	#6CC884

```r
#Example seeing connections in McGill University - Desautels Faculty of Management and Desauetls Capita
# Extract names associated with McGill University
mcgill_names <- df_connections %>%
  filter(company %in% c(
    "McGill University - Desautels Faculty of Management",
    "Desautels Capital Management"
  )) %>%
  select(from, to) %>%
  pivot_longer(cols = c(from, to)) %>%
  distinct() %>%
  pull(value)
```

```r
# Create a graph layout
graph_layout <- create_layout(graph_data, layout = "backbone", keep = 0.7)
```

```r
# Visualize the graph
graph_vis <- ggraph(graph_layout) +
  geom_node_point(
    size = 4,
    color = ifelse(
      graph_data %>%
        activate(nodes) %>%
        pull(name) %in% mcgill_names,
      "brown",
      "black"
    )
  ) +
  geom_node_text(
    aes(label = name),
    repel = TRUE,
    max.overlaps = 5,
    check_overlap = TRUE,
    size = 2 # Adjust text size here if needed
  ) +
  geom_edge_link0(
    aes(color = company),
    show.legend = TRUE,
    width = 1
  ) +
  scale_edge_color_manual(values = color_palette$hex) +
  theme_void() +
  theme(
    legend.position = "bottom",
    panel.border = element_rect(fill = NA)
  ) +
  labs(
    title = "LinkedIn Connections Network",
    subtitle = "Edges are colored based on the organization (sample of 30% of connections, organization
    caption = "Linkedin Connections Data by Ammad Sohail"
  )

# Preview the graph in the R plotting window
print(graph_vis)
```

LinkedIn Connections Network

Edges are colored based on the organization (sample of 30% of connections, organizations with
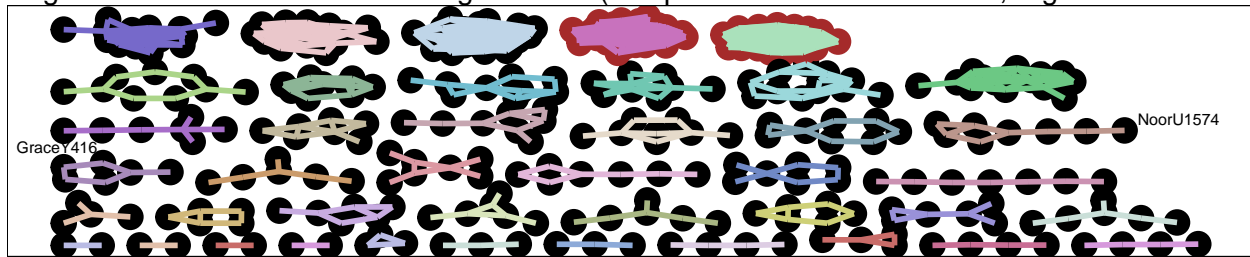
S)	— Engro Fertilizers Limited	— McGill University	— Pr
	— Engro Polymer & Chemicals Ltd	— McGill University – Desautels Faculty of Management	— Pr
	— Ericsson	— Motive	— RB
	— Fatima Group	— National Aerospace Science & Technology Park (NASTP)	— S&
	— Fiverr	— National University of Sciences and Technology (NUST)	— Sa
	— foodpanda	— Nestlé	— Sa
	— Jazz	— OneScreen	— Sc
nent	— L'Oréal	— Packages Group	— Se

Linkedin Connections Data by Ammad Sohail

```r
# Saving the graph to a PNG file
ggsave("E:/Linkedin_Connections_Networks_Graph.png",
  graph_vis,
  width = 20,    # Set the width of the saved image
  height = 15,   # Set the height of the saved image
  dpi = 300,     # Set the resolution in dots per inch
  device = "jpg" # Specify the type of graphics device
)

options(warn = 0)  # Reset to default behavior
```