Sustainability Roles

Arjun Kumar

```r
#laod libraries
library(igraph)
library(tidyverse)
library(quanteda)
library(visNetwork)
```

```r
roles_df = read.csv('sustainability_roles.csv')
head(roles_df)
```

```
##                  name
## 1      Zhi Ling Wong
## 2         Rahul Prem
## 3  Alyssa Lin, CESGA
## 4 Faisal Abdul Karim
## 5      Karuna Bhatia
## 6        Wasif Ahmad
##
## 1                                                                          Head, Reputational & Susta
## 2                                                                                         Program Ma
## 3
## 4 Innovation leader in Banking | FinTech | Risk & Compliance | Digital Assets | Private Equity | Sus
## 5                                                               Head of Sustainability, Standard Ch
## 6                                                                                        Senior Mana
##              company
## 1 standardchartered
## 2 standardchartered
## 3 standardchartered
## 4 standardchartered
## 5 standardchartered
## 6 standardchartered
```

```r
names(roles_df)
```

```
## [1] "name"    "role"    "company"
```

```r
roles_corp = corpus(roles_df, text_field = 'role')
summary(roles_corp, 5)
```

```
## Corpus consisting of 1225 documents, showing 5 documents:
##
##   Text Types Tokens Sentences               name           company
##  text1    11     11         1      Zhi Ling Wong standardchartered
```

```
##  text2     7      8         1         Rahul Prem standardchartered
##  text3     7      7         1  Alyssa Lin, CESGA standardchartered
##  text4    22     29         1 Faisal Abdul Karim standardchartered
##  text5    13     15         1      Karuna Bhatia standardchartered
```

```
roles_toks = tokens(roles_corp,
                    remove_punct = TRUE,
                    remove_symbols = TRUE,
                    remove_numbers = TRUE,
                    remove_url = TRUE,
                    remove_separators = TRUE)

roles_toks %>% head()
```

```
## Tokens consisting of 6 documents and 2 docvars.
## text1 :
## [1] "Head"           "Reputational"   "Sustainability" "Risk"
## [5] "Frameworks"     "at"             "Standard"       "Chartered"
## [9] "Bank"
##
## text2 :
## [1] "Program"        "Manager"        "Financial"      "Markets"
## [5] "Sustainability" "Enthusiast"
##
## text3 :
## [1] "Make"   "the"    "world"  "a"      "better" "place"
##
## text4 :
##  [1] "Innovation"     "leader"         "in"             "Banking"
##  [5] "FinTech"        "Risk"           "Compliance"     "Digital"
##  [9] "Assets"         "Private"        "Equity"         "Sustainability"
## [ ... and 5 more ]
##
## text5 :
##  [1] "Head"           "of"             "Sustainability" "Standard"
##  [5] "Chartered"      "Bank"           "and"            "Member"
##  [9] "of"             "Board"          "United"         "Way"
## [ ... and 2 more ]
##
## text6 :
## [1] "Senior"         "Manager"        "Reputational"   "and"
## [5] "Sustainability" "Risk"           "Management"
```

```
roles_toks = roles_toks %>% tokens_remove(c(stopwords("english"), 'standard', "chartered", "bnp", "pari
                                            "CO", "w", "I", "morgan",
                                            "bnpp", "jp"))

roles_toks = roles_toks %>% tokens_replace("manager-", "manager")

roles_toks
```

```
## Tokens consisting of 1,225 documents and 2 docvars.
## text1 :
```

```
## [1] "Head"           "Reputational"   "Sustainability" "Risk"
## [5] "Frameworks"     "Bank"
##
## text2 :
## [1] "Program"        "Manager"        "Financial"      "Markets"
## [5] "Sustainability" "Enthusiast"
##
## text3 :
## [1] "Make"   "world"  "better" "place"
##
## text4 :
##  [1] "Innovation"     "leader"          "Banking"        "FinTech"
##  [5] "Risk"           "Compliance"      "Digital"        "Assets"
##  [9] "Private"        "Equity"          "Sustainability" "Consulting"
## [ ... and 4 more ]
##
## text5 :
## [1] "Head"           "Sustainability" "Bank"           "Member"
## [5] "Board"          "United"         "Way"            "Mumbai"
##
## text6 :
## [1] "Senior"         "Manager"        "Reputational"   "Sustainability"
## [5] "Risk"           "Management"
##
## [ reached max_ndoc ... 1,219 more documents ]
```

```
min_freq <- 5

# Create DTM, prune vocabulary and set binary values for presence/absence of types
roles_dfm <- roles_toks %>%
             dfm() %>%
             dfm_trim(min_docfreq = min_freq) %>%                                    dfm_weight("bo

roles_dfm
```

```
## Document-feature matrix of: 1,225 documents, 236 features (98.18% sparse) and 2 docvars.
##        features
## docs    head reputational sustainability risk bank program manager financial
##   text1    1            1              1    1    1       0       0         0
##   text2    0            0              1    0    0       1       1         1
##   text3    0            0              0    0    0       0       0         0
##   text4    0            0              1    1    0       0       0         0
##   text5    1            0              1    0    1       0       0         0
##   text6    0            1              1    1    0       0       1         0
##        features
## docs    markets enthusiast
##   text1       0          0
##   text2       1          1
##   text3       0          0
##   text4       0          0
##   text5       0          0
##   text6       0          0
## [ reached max_ndoc ... 1,219 more documents, reached max_nfeat ... 226 more features ]
```

```
roles_cc = t(roles_dfm) %*% roles_dfm
roles_cc[1:5, 1:5]
```

```
## 5 x 5 sparse Matrix of class "dgCMatrix"
##                head reputational sustainability risk bank
## head            194            2             50    9   42
## reputational      2           17             17   16   10
## sustainability   50           17            347   33   50
## risk              9           16             33  104   25
## bank             42           10             50   25  162
```

```
cooct <- "sustainability"
k <- nrow(roles_dfm)
ki <- sum(roles_dfm[, cooct])
kj <- colSums(roles_dfm)
names(kj) <- colnames(roles_dfm)
kij <- roles_cc[cooct, ]
```

```
# Read in the source code for the co-occurrence calculation
source("calcCoocStats.R")
ncoocs <- 13
cooct <- "sustainability"

coocs <- calcCoocStats(cooct, roles_dfm, measure="DICE")
# Display the ncoocs main terms
print(coocs[1:ncoocs])
```

```
##      manager   management         bank         head          esg         risk
##   0.22360248   0.22267206   0.19646365   0.18484288   0.16371681   0.14634146
##      officer     director    corporate     investor       global reputational
##   0.12626263   0.12500000   0.12121212   0.11859838   0.10185185   0.09340659
##   engagement
##   0.09207161
```

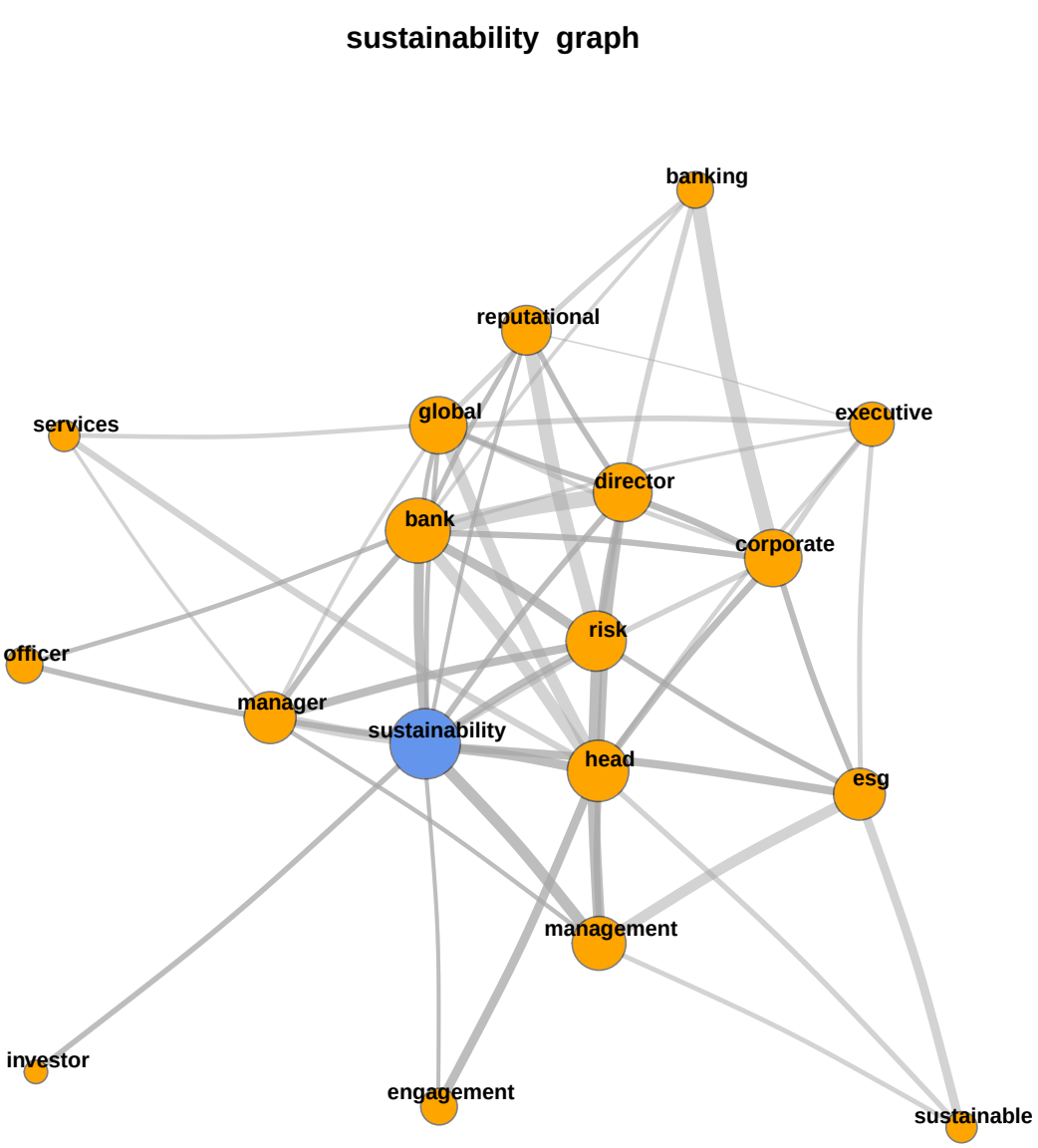
```
resultGraph <- data.frame(from = character(), to = character(), sig = numeric(0))

# The structure of the temporary graph object is equal to that of the resultGraph
tmpGraph <- data.frame(from = character(), to = character(), sig = numeric(0))

# Fill the data.frame to produce the correct number of lines
tmpGraph[1:ncoocs, 3] <- coocs[1:ncoocs]
# Entry of the search word into the first column in all lines
tmpGraph[, 1] <- cooct
# Entry of the co-occurrences into the second column of the respective line
tmpGraph[, 2] <- names(coocs)[1:ncoocs]
# Set the significances
tmpGraph[, 3] <- coocs[1:ncoocs]

# Attach the triples to resultGraph
resultGraph <- rbind(resultGraph, tmpGraph)

# Iteration over the most significant ncoocs co-occurrences of the search term
```

```r
for (i in 1:ncoocs){

  # Calling up the co-occurrence calculation for term i from the search words co-occurrences
  newcooct <- names(coocs)[i]
  coocs2 <- calcCoocStats(newcooct, roles_dfm, measure="DICE")

  #print the co-occurrences
  coocs2[1:10]

  # Structure of the temporary graph object
  tmpGraph <- data.frame(from = character(), to = character(), sig = numeric(0))
  tmpGraph[1:ncoocs, 3] <- coocs2[1:ncoocs]
  tmpGraph[, 1] <- newcooct
  tmpGraph[, 2] <- names(coocs2)[1:ncoocs]
  tmpGraph[, 3] <- coocs2[1:ncoocs]

  #Append the result to the result graph
  resultGraph <- rbind(resultGraph, tmpGraph[2:length(tmpGraph[, 1]), ])
}
```

```r
#resultGraph %>% filter(to == "sustainability")
```

```r
# set seed for graph plot
set.seed(1)

# Create the graph object as undirected graph
graphNetwork <- graph.data.frame(resultGraph, directed = F)

# Identification of all nodes with less than 2 edges
verticesToRemove <- V(graphNetwork)[degree(graphNetwork) < 3]
# These edges are removed from the graph
graphNetwork <- delete.vertices(graphNetwork, verticesToRemove)

# Assign colors to nodes (search term blue, others orange)
V(graphNetwork)$color <- ifelse(V(graphNetwork)$name == cooct, 'cornflowerblue', 'orange')

# Set edge colors
E(graphNetwork)$color <- adjustcolor("DarkGray", alpha.f = .5)
# scale significance between 1 and 10 for edge width
E(graphNetwork)$width <- scales::rescale(E(graphNetwork)$sig, to = c(1, 10))

# Set edges with radius
E(graphNetwork)$curved <- 0.05
# Size the nodes by their degree of networking (scaled between 5 and 15)
V(graphNetwork)$size <- scales::rescale(log(degree(graphNetwork)), to = c(5, 15))

# Define the frame and spacing for the plot
par(mai=c(0,0,1,0))

# Final Plot
plot(
  graphNetwork,
  layout = layout.fruchterman.reingold, # Force Directed Layout
```

```
  main = paste(cooct, ' graph'),
  vertex.label.family = "sans",
  vertex.label.cex = 0.8,
  vertex.shape = "circle",
  vertex.label.dist = 0.5,          # Labels of the nodes moved slightly
  vertex.frame.color = adjustcolor("black", alpha.f = .5),
  vertex.label.color = 'black',     # Color of node names
  vertex.label.font = 2,            # Font of node names
  vertex.label = V(graphNetwork)$name,      # node names
  vertex.label.cex = 1 # font size of node names
)
```



```
#visIgraph(graphNetwork)
```