Untitled

Smitesh Patil

2023-02-01

1. Basic measures

```
library(igraph)
```

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

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

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

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

```
library(kableExtra)
```

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

```
dib_graph<-read.graph("dib2.graphml",format="graphml")
```

1.1 Give the number of nodes and edges

```
cat("num vertices:", vcount(dib_graph), "\n")
```

```
## num vertices: 8969
```

```
cat("num edges :", ecount(dib_graph), "\n")
```

```
## num edges : 46750
```

1.2 Is the network strongly or weakly connected. If neither, what is the distribution of component sizes.

1. Strongly connected components

```
strong_component = as.data.frame(table(factor(components(dib_graph, mode="strong")$csize)))
names(strong_component)[1] = "Component Size"
kable(strong_component)
```

Component Size	Freq
1	3024
2	180
3	25
4	5
5	1
6	1
5479	1

Answer: The table above shows strongly connected components in the directed network. There is one component with 5479, 6 and 5 nodes and a distribution of component with sizes varying from 1 to 4.

2. Weakly connected components

```
weak_component = as.data.frame(table(factor(components(dib_graph, mode="weak")$csize)))
names(weak_component)[1] = "Component Size"
kable(weak_component)
```

Component Size	Freq
2	30
3	11
4	1
8872	1

Answer: The table above shows weakly connected components in the directed network. There is one component with 8872 and 4 nodes with 30 weakly connected components of size 2 and 11 components with size 3.

1.3 What is the diameter of the network ?

```
cat("The diameter of the network is : ", diameter(dib_graph, directed = T, unconnected = TRUE, weights=N
```

```
## The diameter of the network is :  18
```

1.4 What is the average path length of the network ?

```
cat("The average path length of the network :", mean_distance(dib_graph, directed = T), "\n")
```

```
## The average path length of the network : 6.017593
```

1.5 What is the clustering coefficient of the network ?

```
cat("The clustering coeff of the graph is :", transitivity(dib_graph, type="localaverage"), "\n")
```

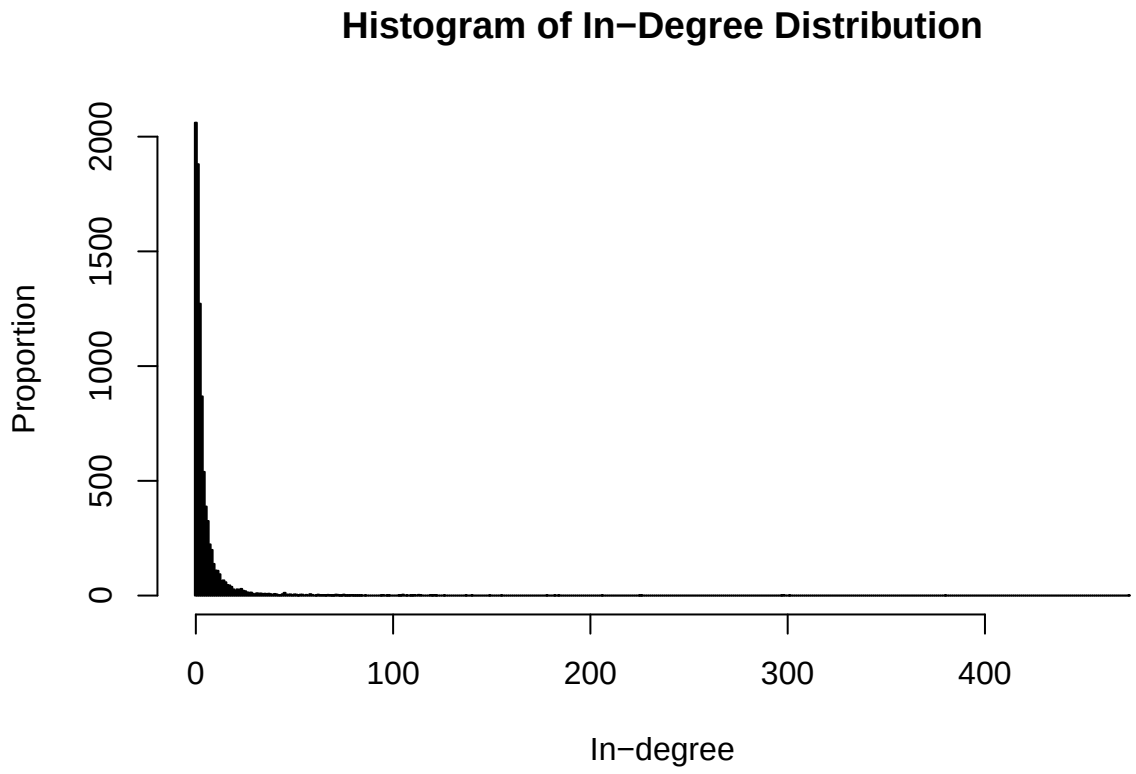
```
## The clustering coeff of the graph is : 0.2300017
```

1.6 What is the in- and out-degree distribution ?

```
deg<-degree(dib_graph, mode = "in")
cat("The in-degree distribution of the graph varies from ", min(deg), "to ", max(deg))
```

```
## The in-degree distribution of the graph varies from  0 to  473
```

```
hist(deg,
     breaks=(min(deg)-1):(max(deg))+0.5,
     xlab = "In-degree",
     ylab = "Proportion",
     main = "Histogram of In-Degree Distribution",
     border="black",
     col="white",
     )
```

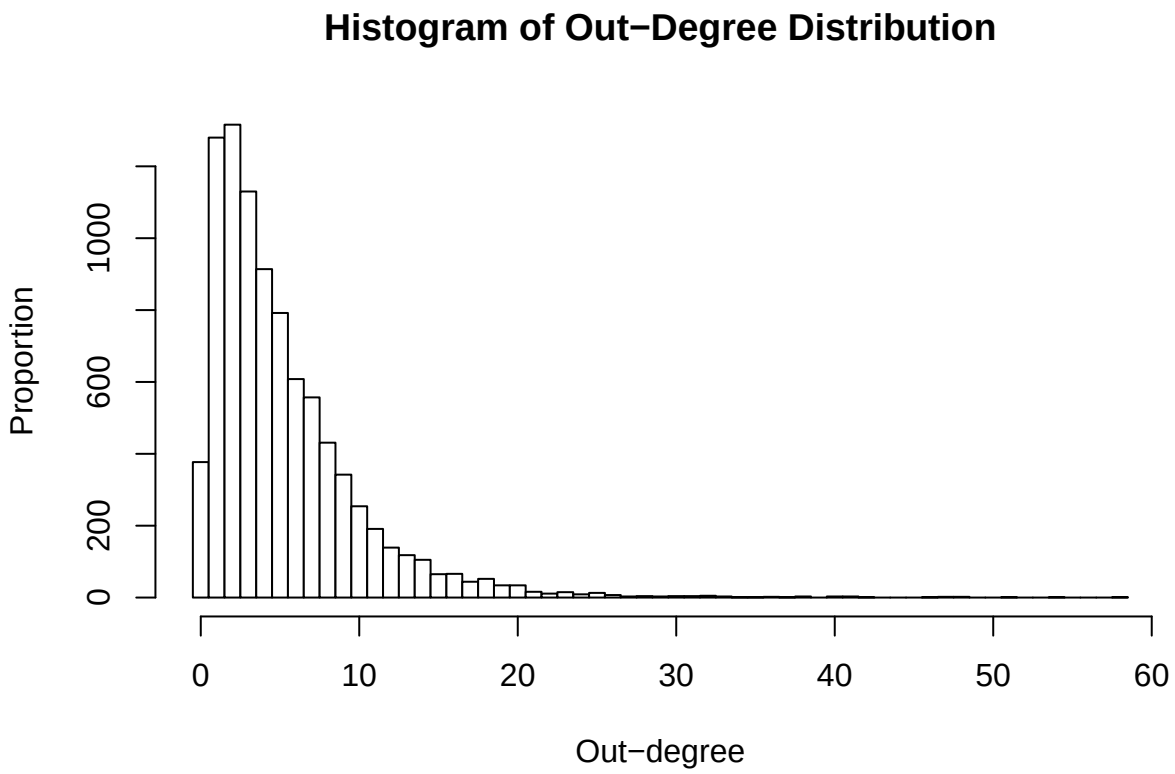



```
deg<-degree(dib_graph, mode = "out")
cat("The in-degree distribution of the graph varies from ", min(deg), "to ", max(deg))
```

```
## The in-degree distribution of the graph varies from  0 to  58
```

```
hist(deg,
     breaks=(min(deg)-1):(max(deg))+0.5,
     xlab = "Out-degree",
     ylab = "Proportion",
```

```
     main = "Histogram of Out-Degree Distribution",
     border="black",
     col="white")
```




```
clu<-components(dib_graph,mode = "strong")

scc_index <- which.max(clu$csize)

# These are are the vertices in the largest SCC
scc<- V(dib_graph)[which(clu$membership==scc_index)]$name

scc_career<- V(dib_graph)[which(clu$membership==scc_index)]$career
scc_century<- V(dib_graph)[which(clu$membership==scc_index)]$century

IN_component=c()
IN_component_career = c()
IN_component_century = c()
vertices<-V(dib_graph)

non_SCC <-vertices[!(vertices$name %in% scc)]

for (v in non_SCC){

  dist<- bfs(dib_graph, root=v, mode="out",unreachable =F, dist=T)$dist
```

```
  connected_to_SCC<-!is.nan(dist[scc])

  if(TRUE %in% connected_to_SCC){
    IN_component<-c(IN_component,V(dib_graph)[v]$name)
    IN_component_career <- c(IN_component_career, V(dib_graph)[v]$career)
    IN_component_century <- c(IN_component_century, V(dib_graph)[v]$century)
  }
}


OUT_component = c()
OUT_component_career = c()
OUT_component_century = c()

non_SCC2 <- vertices[!(vertices$name %in% c(scc, IN_component))]

for (v in non_SCC2){
  dist<- bfs(dib_graph, root=v, mode="in",unreachable =F, dist=T)$dist
  connected_to_non_SCC<-!is.nan(dist[scc])
  if(TRUE %in% connected_to_non_SCC){
    OUT_component<-c(OUT_component,V(dib_graph)[v]$name)
    OUT_component_career <- c(OUT_component_career, V(dib_graph)[v]$career)
    OUT_component_century <- c(OUT_component_century, V(dib_graph)[v]$century)
  }
}

tube = c()
tube_career = c()
tube_century = c()
nodes_minus_SCC <- vertices[!(vertices$name %in% scc)]
g_minus_SCC<-induced_subgraph(dib_graph ,nodes_minus_SCC)
out = c()

for (v in V(g_minus_SCC)[IN_component]){
  paths<-all_simple_paths(g_minus_SCC, from=v,to = V(g_minus_SCC)[OUT_component], mode = "out", cutoff =
  tube<-c(tube, names(unlist(paths)))
}

tube<-unique(tube)

tube<-tube[!(tube %in% c(IN_component,OUT_component))]


tube_data = c()
tube_data <- V(dib_graph)[V(dib_graph)$name %in% tube]
tube_data_career <- tube_data$career
tube_data_century <- tube_data$century

length(OUT_component)
```

```
## [1] 223
```

```
library(dplyr)
```

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

## The following object is masked from 'package:kableExtra':
##
##     group_rows

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

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

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

```
components_prop <- c(length(scc), length(IN_component), length(OUT_component), length(tube))

prop_table <- prop.table(components_prop) *100

names(prop_table) <- c("SCC", "IN Component", "OUT component", "Tube")

kable(prop_table, col.names = "Percentage")
```

	Percentage
SCC	64.6718602
IN Component	32.5542965
OUT component	2.6322002
Tube	0.1416431

```
library(dplyr)

scc_career <- trimws(unlist(strsplit(scc_career, split = ",")))
scc_career <- as_tibble(table(scc_career)) %>% arrange(desc(n)) %>% head()


IN_component_career <- trimws(unlist(strsplit(IN_component_career, split = ",")))
IN_component_career <- as_tibble(table(IN_component_career)) %>% arrange(desc(n)) %>% head()

OUT_component_career <- trimws(unlist(strsplit(OUT_component_career, split = ",")))
OUT_component_career <- as_tibble(table(OUT_component_career)) %>% arrange(desc(n)) %>% head()


tube_component_career <- trimws(unlist(strsplit(tube_data_career, split = ",")))
tube_component_career <- as_tibble(table(tube_component_career)) %>% arrange(desc(n)) %>% head()

knitr::kable(list(scc_career, IN_component_career, OUT_component_career,tube_component_career))
```

scc_career	n
Politics	1909
Religion	1004
Literature	587
Military	502
Gentry and Aristocracy	486
Administration and Diplomacy	443

IN_component_career	n
Politics	591
Religion	487
Literature	312
Business and Finance	240
Journalism and Broadcasting	224
Administration and Diplomacy	209

OUT_component_career	n
Religion	37
Science and Technology	26
Politics	25
Sport	20
Administration and Diplomacy	18
Business and Finance	18

tube_component_career	n
Military	3
Science and Technology	2
The Sea	2
Travel and Exploration	2
Administration and Diplomacy	1
Archaeology and Antiquarianism	1

scc_century	n
19	1871
18	1133
20	597
17	595
16	474
15	118

IN_component_century	n
19	1090
20	789
18	419
17	170
16	120
13	36

OUT_component_century	n
19	112
20	52
18	39
17	10
16	4
13	3

tube_data_century	n
20	8
19	3
18	1

```
library(dplyr)


scc_century <- as_tibble(table(scc_century)) %>% arrange(desc(n)) %>% head()

IN_component_century <- as_tibble(table(IN_component_century)) %>% arrange(desc(n)) %>% head()

OUT_component_century <- as_tibble(table(OUT_component_century)) %>% arrange(desc(n)) %>% head()

tube_component_century <- as_tibble(table(tube_data_century)) %>% arrange(desc(n)) %>% head()

knitr::kable(list(scc_century, IN_component_century, OUT_component_century,tube_component_century))
```

```
library(sjmisc)
```

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

```
## Learn more about sjmisc with 'browseVignettes("sjmisc")'.
```

```
century_data <- list(induced.subgraph(dib_graph, V(dib_graph)[V(dib_graph)$century == 15]),
                     induced.subgraph(dib_graph, V(dib_graph)[V(dib_graph)$century == 16]),
                     induced.subgraph(dib_graph, V(dib_graph)[V(dib_graph)$century == 17]),
                     induced.subgraph(dib_graph, V(dib_graph)[V(dib_graph)$century == 18]),
```

```
                    induced.subgraph(dib_graph, V(dib_graph)[V(dib_graph)$century == 19]),
                    induced.subgraph(dib_graph, V(dib_graph)[V(dib_graph)$century == 20]))


page_rank <- vector(mode = "list", length = 6)
katz_centrality <- vector(mode = "list", length = 6)
eigen_centrality <- vector(mode = "list", length = 6)
between <- vector(mode = "list", length = 6)
close <- vector(mode = "list", length = 6)

names(century_data) <- c("15th century",
                         "16th century",
                         "17th century",
                         "18th century",
                         "19th century",
                         "20th century")


for(i in 1:length(century_data)){
  page_rank[[i]] <- page_rank(century_data[[i]],  directed = TRUE, damping = 0.85)$vector
  eigen_centrality[[i]] <- eigen_centrality(century_data[[i]], weights=NA)$vector
  page_rank[[i]] <- page_rank[[i]] %>% sort(decreasing = TRUE) %>% head(1)
  eigen_centrality[[i]] <- eigen_centrality[[i]] %>% sort(decreasing = TRUE) %>% head(1)
  between[[i]] <- betweenness(century_data[[i]], weights=NA)
  between[[i]] <- between[[i]] %>% sort(decreasing = TRUE) %>% head(1)
}

out <- tibble(page_rank = names(unlist(page_rank)),
              eigen_centrality = names(unlist(eigen_centrality)),
              between = names(unlist(between)))
out <- rotate_df(out)

colnames(out) <-  c("15th century",
                         "16th century",
                         "17th century",
                         "18th century",
                         "19th century",
                         "20th century")
kable(out)
```

	15th century	16th century	17th century	18th century	19th century
page_rank	Gerald FitzGerald	Hugh O'Neill	James Butler	Daniel O'Connell	Charles Stewart Parnell
eigen_centrality	Gerald FitzGerald	Hugh O'Neill	James II and VII	Wolfe Tone	Ã‰amon De Valera
between	Gerald FitzGerald	Hugh O'Neill	Richard Talbot	Wolfe Tone	Ã‰amon De Valera