Assignment 1 SNA

Marc Sparhuber

Table of contents

Task 1

> Write an R-script that reads your data. Transform your data to a data frame. Argue about the format you have chosen to analyze large network data.

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.4
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.4.4     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 stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

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

```
library(sna)
```

```
Loading required package: statnet.common
Attaching package: 'statnet.common'
The following objects are masked from 'package:base':
    attr, order
Loading required package: network
```

```
'network' 1.18.2 (2023-12-04), part of the Statnet Project
* 'news(package="network")' for changes since last version
* 'citation("network")' for citation information
* 'https://statnet.org' for help, support, and other information


Attaching package: 'network'

The following objects are masked from 'package:igraph':

    %c%, %s%, add.edges, add.vertices, delete.edges, delete.vertices,
    get.edge.attribute, get.edges, get.vertex.attribute, is.bipartite,
    is.directed, list.edge.attributes, list.vertex.attributes,
    set.edge.attribute, set.vertex.attribute

sna: Tools for Social Network Analysis
Version 2.7-2 created on 2023-12-05.
copyright (c) 2005, Carter T. Butts, University of California-Irvine
 For citation information, type citation("sna").
 Type help(package="sna") to get started.


Attaching package: 'sna'

The following objects are masked from 'package:igraph':

    betweenness, bonpow, closeness, components, degree, dyad.census,
    evcent, hierarchy, is.connected, neighborhood, triad.census
```

```
library(knitr)
library(kableExtra)
```

```
Attaching package: 'kableExtra'

The following object is masked from 'package:dplyr':

    group_rows
```

```r
# here I load in the data. To not have any issues with how the .txt file is structured I r

data <- as.data.frame(read_table(file = "Email-Enron.txt",
                    skip = 4,
                    col_names = c("FromNodeID","ToNodeID")))
```

```
-- Column specification --------------------------------------------------------
cols(
  FromNodeID = col_double(),
  ToNodeID = col_double()
)
```

Task 2

Make sure that before or while doing any calculations or plots that you handle nonresponse / missing data meaningfully if there is any.

```r
# check if there are any NAs

any(is.na(data))
```

```
[1] FALSE
```

```r
# this indicates that in neither of the two columns any NAs are present
```

Task 3

Justify what you do with isolates and multiple components if there are any.

```r
# first check how many isolates there are. There are also no multiple components.

isolates(data)
```

```
integer(0)
```

Task 4

Analyze the density of your network. Create a table that contains further descriptive network statistics for your network. Please include average degree (in-degree and outdegree), standard deviation of degree (in-degrees and out-degrees), reciprocity, and transitivity.

```
graph <- graph_from_data_frame(data, directed = FALSE)

# due to this being an undirected network, it makes no sense to caluclate in/outdegree. In
degrees <- igraph::degree(graph,mode='all')
mean_degrees <- mean(degrees)

# standard deviation of degrees
sd_degrees <- sd(degrees)

# is 1 because all ties a undirected and no ties go beyond the borders of the network [NOT
reciprocity <- reciprocity(graph)

# transitivity
transitivity <- transitivity(graph)

# combining these measures into a table
kable(tribble(~MeanDegrees, ~SDDegrees, ~Reciprocity, ~Transitivity,
        mean_degrees, sd_degrees, reciprocity, transitivity), caption = "Network Descripti
   add_footnote("Due to this network being undirected, in- and out-degree are not calculat
```

Table 1: Network Descriptives

MeanDegrees	SDDegrees	Reciprocity	Transitivity
20.04044	72.20099	1	0.0853108

[a] Due to this network being undirected, in- and out-degree are not calculated

```
# plot.igraph(graph,
#             vertex.label = "",
#             vertex.size = 2,
#             layout=layout_with_kk(graph),
#             main = "Enron Emails (internal)")

# -> this runs for more than 2 hours so I will let it run over night and subset the data w
```

```
# data |> group_by(FromNodeID) |> filter(n()>1) |> nrow()
# data |> group_by(FromNodeID) |> filter(n()>2) |> nrow()
# data |> group_by(FromNodeID) |> filter(n()>3) |> nrow()
# data |> group_by(FromNodeID) |> filter(n()>5) |> nrow()
# data |> group_by(FromNodeID) |> filter(n()>10) |> nrow()
# data |> group_by(FromNodeID) |> filter(n()>20) |> nrow()
# data |> group_by(FromNodeID) |> filter(n()>50) |> nrow()
```