Assignment 7 Graph Theory

Ridhwan Firdaus bin Mohd Jata 16000250 EDB4623:Big Data Analytics

3/31/2021

Introduction

The two primary aspects of networks are a multitude of separate entities and the connections between them.The entities are referred to as nodes or vertices of a graph, while the connections are edges or links

Data used

The data sets that will be used relates to French trains; it contains aggregate daily total trips per connecting stations.

```
library(readr)

## Warning: package 'readr' was built under R version 3.6.3

url <-
"https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2019/2019-02-26/small_trains.csv"
small_trains <- read_csv(url)

##
## -- Column specification --------------------------------------------------
## cols(
##   year = col_double(),
##   month = col_double(),
##   service = col_character(),
##   departure_station = col_character(),
##   arrival_station = col_character(),
##   journey_time_avg = col_double(),
##   total_num_trips = col_double(),
##   avg_delay_all_departing = col_double(),
##   avg_delay_all_arriving = col_double(),
##   num_late_at_departure = col_double(),
##   num_arriving_late = col_double(),
##   delay_cause = col_character(),
##   delayed_number = col_double()
## )

head(small_trains)
```

```
## # A tibble: 6 x 13
##    year month service  departure_station arrival_station
journey_time_avg
##   <dbl> <dbl> <chr>    <chr>             <chr>
<dbl>
## 1  2017     9 National PARIS EST         METZ
85.1
## 2  2017     9 National REIMS             PARIS EST
47.1
## 3  2017     9 National PARIS EST         STRASBOURG
116.
## 4  2017     9 National PARIS LYON        AVIGNON TGV
161.
## 5  2017     9 National PARIS LYON        BELLEGARDE (AIN)
164.
## 6  2017     9 National PARIS LYON        BESANCON FRANCHE COMT~
129.
## # ... with 7 more variables: total_num_trips <dbl>,
## #   avg_delay_all_departing <dbl>, avg_delay_all_arriving <dbl>,
## #   num_late_at_departure <dbl>, num_arriving_late <dbl>, delay_cause
<chr>,
## #   delayed_number <dbl>
```

Data Cleaning and preparation

Data will be changed in order to create a new category called routes. It contains a single entry for each connected station. It also includes the average journey time it takes to go between stations.

```
library(dplyr)

## Warning: package 'dplyr' was built under R version 3.6.3

##
## 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

routes <- small_trains %>%  group_by(departure_station, arrival_station) %>%
summarise(journey_time = mean(journey_time_avg)) %>%  ungroup()%>%
mutate(from = departure_station,to = arrival_station) %>%  select(from, to,
journey_time)

## `summarise()` has grouped output by 'departure_station'. You can override
using the `.groups` argument.
```

```r
head(routes)
```

```
## # A tibble: 6 x 3
##   from                to                  journey_time
##   <chr>               <chr>                      <dbl>
## 1 AIX EN PROVENCE TGV PARIS LYON                 186.
## 2 ANGERS SAINT LAUD   PARIS MONTPARNASSE          97.5
## 3 ANGOULEME           PARIS MONTPARNASSE         146.
## 4 ANNECY              PARIS LYON                 225.
## 5 ARRAS               PARIS NORD                  52.8
## 6 AVIGNON TGV         PARIS LYON                 161.
```

Creating the adjency matrix

```r
adjmatrix <- subset(routes, select = -c(journey_time))
head(adjmatrix)
```

```
## # A tibble: 6 x 2
##   from                to
##   <chr>               <chr>
## 1 AIX EN PROVENCE TGV PARIS LYON
## 2 ANGERS SAINT LAUD   PARIS MONTPARNASSE
## 3 ANGOULEME           PARIS MONTPARNASSE
## 4 ANNECY              PARIS LYON
## 5 ARRAS               PARIS NORD
## 6 AVIGNON TGV         PARIS LYON
```

```r
edges_mat = matrix(nrow = length(adjmatrix$from), ncol =
length(adjmatrix$to), data=0, dimnames = list(adjmatrix$from, adjmatrix$to))
edges_mat[as.matrix(adjmatrix)] = 1
head(edges_mat)
```

```
##                     PARIS LYON PARIS MONTPARNASSE PARIS MONTPARNASSE PARIS
LYON
## AIX EN PROVENCE TGV          1                  0                  0
0
## ANGERS SAINT LAUD            0                  1                  0
0
## ANGOULEME                    0                  1                  0
0
## ANNECY                       1                  0                  0
0
## ARRAS                        0                  0                  0
0
## AVIGNON TGV                  1                  0                  0
0
##                     PARIS NORD PARIS LYON PARIS LYON PARIS LYON PARIS LYON
## AIX EN PROVENCE TGV          0          0          0          0          0
## ANGERS SAINT LAUD            0          0          0          0          0
## ANGOULEME                    0          0          0          0          0
## ANNECY                       0          0          0          0          0
```

```
## ARRAS                          1          0          0          0          0
## AVIGNON TGV                    0          0          0          0          0
##                      PARIS MONTPARNASSE PARIS VAUGIRARD TOURCOING
## AIX EN PROVENCE TGV                  0               0         0
## ANGERS SAINT LAUD                    0               0         0
## ANGOULEME                            0               0         0
## ANNECY                               0               0         0
## ARRAS                                0               0         0
## AVIGNON TGV                          0               0         0
##                      PARIS MONTPARNASSE PARIS LYON PARIS LYON PARIS NORD
## AIX EN PROVENCE TGV                  0          0          0          0
## ANGERS SAINT LAUD                    0          0          0          0
## ANGOULEME                            0          0          0          0
## ANNECY                               0          0          0          0
## ARRAS                                0          0          0          0
## AVIGNON TGV                          0          0          0          0
##                      PARIS NORD PARIS EST PARIS LYON PARIS LYON PARIS LYON
## AIX EN PROVENCE TGV          0         0          0          0          0
## ANGERS SAINT LAUD            0         0          0          0          0
## ANGOULEME                    0         0          0          0          0
## ANNECY                       0         0          0          0          0
## ARRAS                        0         0          0          0          0
## AVIGNON TGV                  0         0          0          0          0
##                      PARIS MONTPARNASSE PARIS LYON PARIS MONTPARNASSE PARIS
LYON
## AIX EN PROVENCE TGV                  0          0                 0
0
## ANGERS SAINT LAUD                    0          0                 0
0
## ANGOULEME                            0          0                 0
0
## ANNECY                               0          0                 0
0
## ARRAS                                0          0                 0
0
## AVIGNON TGV                          0          0                 0
0
##                      PARIS MONTPARNASSE LYON PART DIEU MARSEILLE ST CHARLES
## AIX EN PROVENCE TGV                  0              0                    0
## ANGERS SAINT LAUD                    0              0                    0
## ANGOULEME                            0              0                    0
## ANNECY                               0              0                    0
## ARRAS                                0              0                    0
## AVIGNON TGV                          0              0                    0
##                      PARIS NORD LILLE MARNE LA VALLEE MARSEILLE ST CHARLES
## AIX EN PROVENCE TGV          0     0               0                     0
## ANGERS SAINT LAUD            0     0               0                     0
## ANGOULEME                    0     0               0                     0
## ANNECY                       0     0               0                     0
## ARRAS                        0     0               0                     0
```

```
## AVIGNON TGV                  0     0               0                    0
##                       MONTPELLIER PARIS LYON RENNES PARIS LYON
## AIX EN PROVENCE TGV           0          0      0          0
## ANGERS SAINT LAUD             0          0      0          0
## ANGOULEME                     0          0      0          0
## ANNECY                        0          0      0          0
## ARRAS                         0          0      0          0
## AVIGNON TGV                   0          0      0          0
##                       MARSEILLE ST CHARLES LYON PART DIEU MARSEILLE ST
CHARLES
## AIX EN PROVENCE TGV                    0              0
0
## ANGERS SAINT LAUD                      0              0
0
## ANGOULEME                              0              0
0
## ANNECY                                 0              0
0
## ARRAS                                  0              0
0
## AVIGNON TGV                            0              0
0
##                       LILLE LYON PART DIEU MADRID MARNE LA VALLEE PARIS LYON
## AIX EN PROVENCE TGV     0              0      0               0          0
## ANGERS SAINT LAUD       0              0      0               0          0
## ANGOULEME               0              0      0               0          0
## ANNECY                  0              0      0               0          0
## ARRAS                   0              0      0               0          0
## AVIGNON TGV             0              0      0               0          0
##                       TOURCOING PARIS EST LYON PART DIEU PARIS LYON PARIS
LYON
## AIX EN PROVENCE TGV         0         0              0          0
0
## ANGERS SAINT LAUD           0         0              0          0
0
## ANGOULEME                   0         0              0          0
0
## ANNECY                      0         0              0          0
0
## ARRAS                       0         0              0          0
0
## AVIGNON TGV                 0         0              0          0
0
##                       PARIS EST PARIS MONTPARNASSE PARIS VAUGIRARD
STRASBOURG
## AIX EN PROVENCE TGV         0                  0              0
0
## ANGERS SAINT LAUD           0                  0              0
0
## ANGOULEME                   0                  0              0
```

```
0
## ANNECY                       0                  0              0
0
## ARRAS                        0                  0              0
0
## AVIGNON TGV                  0                  0              0
0
##                      PARIS LYON PARIS LYON FRANCFORT METZ NANCY REIMS
STRASBOURG
## AIX EN PROVENCE TGV          0          0         0    0     0     0
0
## ANGERS SAINT LAUD            0          0         0    0     0     0
0
## ANGOULEME                    0          0         0    0     0     0
0
## ANNECY                       0          0         0    0     0     0
0
## ARRAS                        0          0         0    0     0     0
0
## AVIGNON TGV                  0          0         0    0     0     0
0
##                      STUTTGART AIX EN PROVENCE TGV ANNECY AVIGNON TGV
BARCELONA
## AIX EN PROVENCE TGV          0                   0      0           0
0
## ANGERS SAINT LAUD            0                   0      0           0
0
## ANGOULEME                    0                   0      0           0
0
## ANNECY                       0                   0      0           0
0
## ARRAS                        0                   0      0           0
0
## AVIGNON TGV                  0                   0      0           0
0
##                      BELLEGARDE (AIN) BESANCON FRANCHE COMTE TGV
## AIX EN PROVENCE TGV                0                         0
## ANGERS SAINT LAUD                  0                         0
## ANGOULEME                          0                         0
## ANNECY                             0                         0
## ARRAS                              0                         0
## AVIGNON TGV                        0                         0
##                      CHAMBERY CHALLES LES EAUX DIJON VILLE GENEVE GRENOBLE
## AIX EN PROVENCE TGV                         0           0      0        0
## ANGERS SAINT LAUD                           0           0      0        0
## ANGOULEME                                   0           0      0        0
## ANNECY                                      0           0      0        0
## ARRAS                                       0           0      0        0
## AVIGNON TGV                                 0           0      0        0
##                      ITALIE LAUSANNE LE CREUSOT MONTCEAU MONTCHANIN
```

```
## AIX EN PROVENCE TGV      0        0                              0
## ANGERS SAINT LAUD        0        0                              0
## ANGOULEME                0        0                              0
## ANNECY                   0        0                              0
## ARRAS                    0        0                              0
## AVIGNON TGV              0        0                              0
##                      LYON PART DIEU MACON LOCHE MARSEILLE ST CHARLES
MONTPELLIER
## AIX EN PROVENCE TGV              0           0                    0
0
## ANGERS SAINT LAUD                0           0                    0
0
## ANGOULEME                        0           0                    0
0
## ANNECY                           0           0                    0
0
## ARRAS                            0           0                    0
0
## AVIGNON TGV                      0           0                    0
0
##                      MULHOUSE VILLE NICE VILLE NIMES PERPIGNAN
## AIX EN PROVENCE TGV              0          0     0         0
## ANGERS SAINT LAUD                0          0     0         0
## ANGOULEME                        0          0     0         0
## ANNECY                           0          0     0         0
## ARRAS                            0          0     0         0
## AVIGNON TGV                      0          0     0         0
##                      SAINT ETIENNE CHATEAUCREUX TOULON VALENCE ALIXAN TGV
ZURICH
## AIX EN PROVENCE TGV                           0      0                 0
0
## ANGERS SAINT LAUD                             0      0                 0
0
## ANGOULEME                                     0      0                 0
0
## ANNECY                                        0      0                 0
0
## ARRAS                                         0      0                 0
0
## AVIGNON TGV                                   0      0                 0
0
##                      ANGERS SAINT LAUD ANGOULEME BORDEAUX ST JEAN BREST
## AIX EN PROVENCE TGV                 0         0                0     0
## ANGERS SAINT LAUD                   0         0                0     0
## ANGOULEME                           0         0                0     0
## ANNECY                              0         0                0     0
## ARRAS                               0         0                0     0
## AVIGNON TGV                         0         0                0     0
##                      LA ROCHELLE VILLE LAVAL LE MANS NANTES POITIERS
QUIMPER
```

```
## AIX EN PROVENCE TGV                0     0       0      0        0
0
## ANGERS SAINT LAUD                  0     0       0      0        0
0
## ANGOULEME                          0     0       0      0        0
0
## ANNECY                             0     0       0      0        0
0
## ARRAS                              0     0       0      0        0
0
## AVIGNON TGV                        0     0       0      0        0
0
##                     RENNES ST MALO ST PIERRE DES CORPS TOULOUSE MATABIAU
TOURS
## AIX EN PROVENCE TGV      0       0                   0                 0
0
## ANGERS SAINT LAUD        0       0                   0                 0
0
## ANGOULEME                0       0                   0                 0
0
## ANNECY                   0       0                   0                 0
0
## ARRAS                    0       0                   0                 0
0
## AVIGNON TGV              0       0                   0                 0
0
##                     VANNES ARRAS DOUAI DUNKERQUE LILLE BORDEAUX ST JEAN
NANTES
## AIX EN PROVENCE TGV      0     0     0         0     0                0
0
## ANGERS SAINT LAUD        0     0     0         0     0                0
0
## ANGOULEME                0     0     0         0     0                0
0
## ANNECY                   0     0     0         0     0                0
0
## ARRAS                    0     0     0         0     0                0
0
## AVIGNON TGV              0     0     0         0     0                0
0
##                     RENNES PARIS LYON PARIS MONTPARNASSE PARIS
MONTPARNASSE
## AIX EN PROVENCE TGV      0          0                 0
0
## ANGERS SAINT LAUD        0          0                 0
0
## ANGOULEME                0          0                 0
0
## ANNECY                   0          0                 0
0
```

```
## ARRAS                     0          0                 0
0
## AVIGNON TGV               0          0                 0
0
##                      PARIS EST LYON PART DIEU PARIS MONTPARNASSE PARIS
VAUGIRARD
## AIX EN PROVENCE TGV         0              0                  0
0
## ANGERS SAINT LAUD           0              0                  0
0
## ANGOULEME                   0              0                  0
0
## ANNECY                      0              0                  0
0
## ARRAS                       0              0                  0
0
## AVIGNON TGV                 0              0                  0
0
##                      PARIS LYON PARIS MONTPARNASSE PARIS MONTPARNASSE
NANTES
## AIX EN PROVENCE TGV          0                  0                  0
0
## ANGERS SAINT LAUD            0                  0                  0
0
## ANGOULEME                    0                  0                  0
0
## ANNECY                       0                  0                  0
0
## ARRAS                        0                  0                  0
0
## AVIGNON TGV                  0                  0                  0
0
##                      PARIS EST PARIS EST PARIS LYON PARIS MONTPARNASSE
## AIX EN PROVENCE TGV         0         0          0                  0
## ANGERS SAINT LAUD           0         0          0                  0
## ANGOULEME                   0         0          0                  0
## ANNECY                      0         0          0                  0
## ARRAS                       0         0          0                  0
## AVIGNON TGV                 0         0          0                  0
##                      BORDEAUX ST JEAN MARSEILLE ST CHARLES PARIS
MONTPARNASSE
## AIX EN PROVENCE TGV               0                   0
0
## ANGERS SAINT LAUD                 0                   0
0
## ANGOULEME                         0                   0
0
## ANNECY                            0                   0
0
## ARRAS                             0                   0
```

```
0
## AVIGNON TGV                         0                    0
0
##                      PARIS LYON PARIS MONTPARNASSE PARIS LYON
## AIX EN PROVENCE TGV          0                    0          0
## ANGERS SAINT LAUD            0                    0          0
## ANGOULEME                    0                    0          0
## ANNECY                       0                    0          0
## ARRAS                        0                    0          0
## AVIGNON TGV                  0                    0          0
```

Transform data into graph table

In order to prepare routes for this transformation, it has to contain two variables specifically named from and to.The member of the group is called a node (or vertex) in the graph, and a relationship between nodes is called an edge.

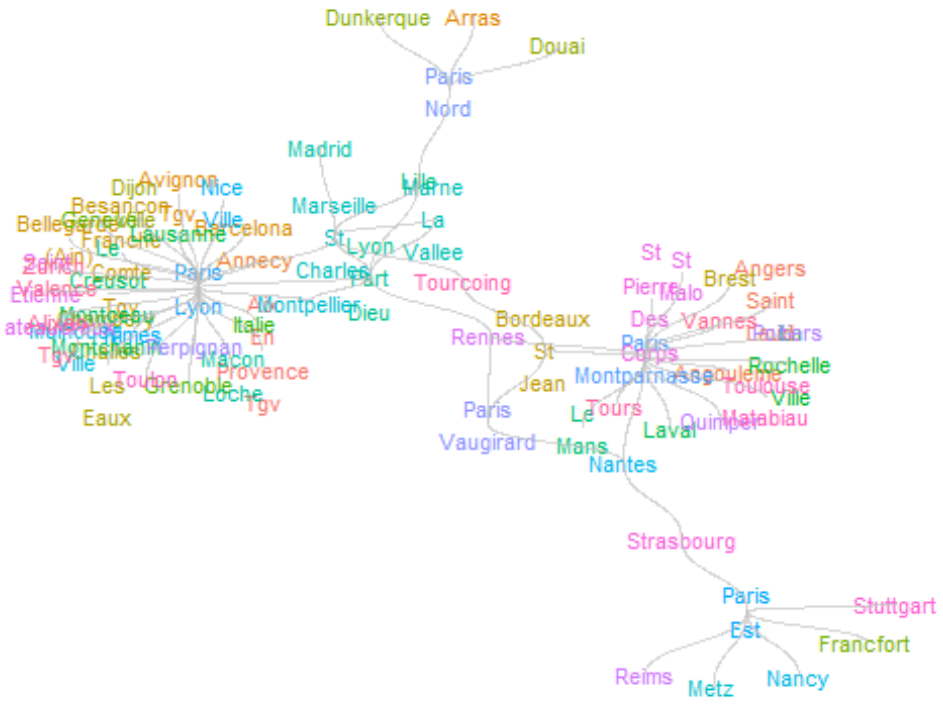
```
library(tidygraph)

## Warning: package 'tidygraph' was built under R version 3.6.3

##
## Attaching package: 'tidygraph'

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

graph_routes <- as_tbl_graph(routes)

graph_routes

## # A tbl_graph: 59 nodes and 130 edges
## #
## # A directed simple graph with 1 component
## #
## # Node Data: 59 x 1 (active)
##   name
##   <chr>
## 1 AIX EN PROVENCE TGV
## 2 ANGERS SAINT LAUD
## 3 ANGOULEME
## 4 ANNECY
## 5 ARRAS
## 6 AVIGNON TGV
## # ... with 53 more rows
## #
## # Edge Data: 130 x 3
##    from    to journey_time
##   <int> <int>        <dbl>
## 1     1    39        186.
```

```
## 2     2    40         97.5
## 3     3    40        146.
## # ... with 127 more rows

library(stringr)

graph_routes <- graph_routes %>%  activate(nodes) %>%  mutate(    title =
str_to_title(name),    label = str_replace_all(title, " ", "\n")    )
stations <- graph_routes %>%  activate(nodes) %>%  pull(title)
```

Visualizing

In this section we will visualize all the nodes and edges from the dataset

```
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.6.3

library(ggraph)

## Warning: package 'ggraph' was built under R version 3.6.3

thm <- theme_minimal() +
  theme(
    legend.position = "none",
     axis.title = element_blank(),
     axis.text = element_blank(),
     panel.grid = element_blank(),
     panel.grid.major = element_blank(),
  )

theme_set(thm)
graph_routes %>%  ggraph(layout = "kk") + geom_node_text(aes(label = label,
color = name), size = 3) +geom_edge_diagonal(color = "gray", alpha = 0.4)
```



Finding the shortest distance In this section we will identify the shortest route from station Arras to station Nancy

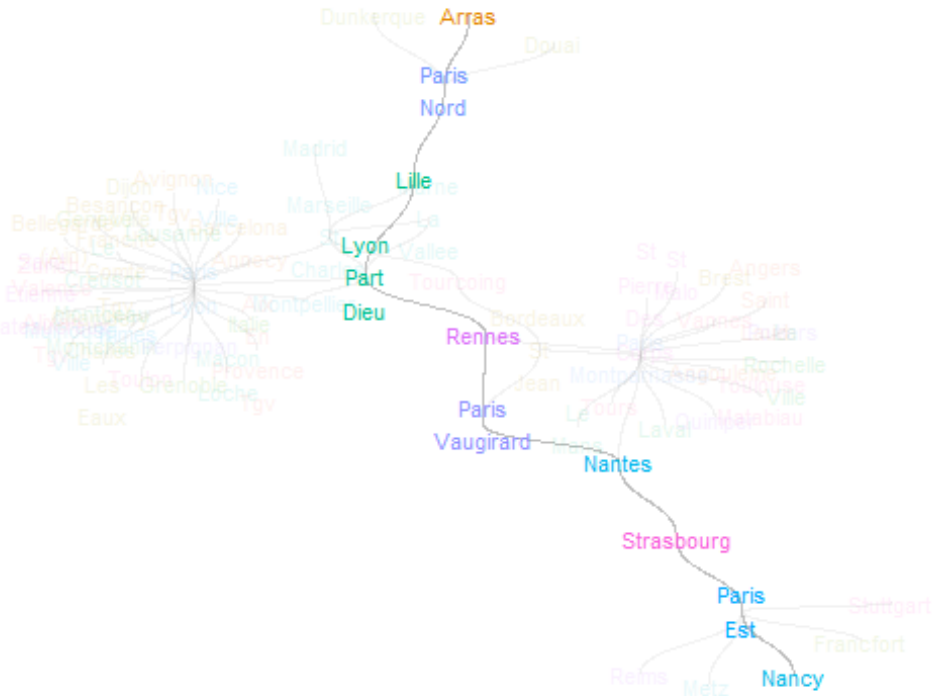
```
from <- which(stations == "Arras")
to <-  which(stations == "Nancy")

shortest <- graph_routes %>%  morph(to_shortest_path, from, to, weights =
journey_time)
shortest %>%  mutate(selected_node = TRUE) %>%  unmorph()

## # A tbl_graph: 59 nodes and 130 edges
## #
## # A directed simple graph with 1 component
## #
## # Node Data: 59 x 4 (active)
##   name                title               label
selected_node
##   <chr>               <chr>               <chr>                      <lgl>
## 1 AIX EN PROVENCE TGV Aix En Provence Tgv "Aix\nEn\nProvence\nTgv" NA
## 2 ANGERS SAINT LAUD   Angers Saint Laud   "Angers\nSaint\nLaud"    NA
## 3 ANGOULEME           Angouleme           "Angouleme"                NA
## 4 ANNECY              Annecy              "Annecy"                   NA
## 5 ARRAS               Arras               "Arras"                    TRUE
## 6 AVIGNON TGV         Avignon Tgv         "Avignon\nTgv"             NA
## # ... with 53 more rows
## #
## # Edge Data: 130 x 3
```

```
##    from    to journey_time
##   <int> <int>        <dbl>
## 1     1    39        186.
## 2     2    40         97.5
## 3     3    40        146.
## # ... with 127 more rows
```

```
shortest <- shortest %>%   mutate(selected_node = TRUE) %>%   activate(edges)
%>%   mutate(selected_edge = TRUE) %>%   unmorph()
shortest <- shortest %>%  activate(nodes) %>%  mutate(selected_node =
ifelse(is.na(selected_node), 1, 2)) %>%  activate(edges) %>%
mutate(selected_edge = ifelse(is.na(selected_edge), 1, 2)) %>%
arrange(selected_edge)
```

Visualizing the shortest distance

```
shortest %>%
  ggraph(layout = "kk") +
    geom_edge_diagonal(aes(alpha = selected_edge), color = "gray") +
    geom_node_text(aes(label = label, color =name, alpha = selected_node ),
size = 3)
```



Summarizing the graph table

```
shortest %>%
  activate(edges) %>%
  filter(selected_edge == 2) %>%
  as_tibble() %>%
```

```
  summarise(
    total_stops = n() - 1,
    total_time = round(sum(journey_time) / 60)
    )
```

```
## # A tibble: 1 x 2
##   total_stops total_time
##         <dbl>      <dbl>
## 1           8         23
```

As we can see there will be a total of 8 stops between Arras and Nancy and the total time is 23 minutes