Estimation of flow trajectories in a multiple lines network

Case studies with *transports publics de la région lausannoise* (tl) data

Guillaume Guex
Romain Loup
François Bavaud

University of Lausanne

Table of contents

Introduction

Context

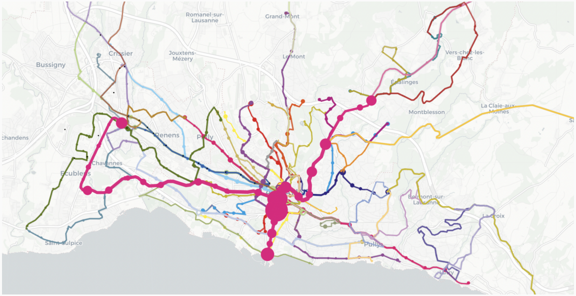
The **tl dataset**, used by Romain Loup for his PhD:

- 1 year of data (2019).
- 115 millions of passengers.
- 42 bus and subway lines.
- 1361 stops and 497 "superstops".
- Every journey data: traveling time, waiting time, embarking and disembarking passengers at each stops, etc.

Context

```
##   stop_id  stop_name line_id direction order embarkment disembarkment
## 1 MALAD_N  Maladière       1         A     1     164558             0
## 2 MTOIE_E    Montoie       1         A     2     136236         12705
## 3 BATEL_E  Batelière       1         A     3     203045         13409
## 4 RTCOU_E Riant-Cour       1         A     4     156015         24909

##    stop_id  stop_name line_id direction order embarkment disembarkment
## 42 RTCOU_O Riant-Cour       1         R    19      23634        132201
## 43 BATEL_O  Batelière       1         R    20      13707        168884
## 44 MTOIE_O    Montoie       1         R    21       4259        128255
## 45 MALAD_N  Maladière       1         R    22          0        146798
```



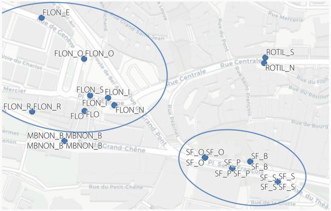
The multiple lines network

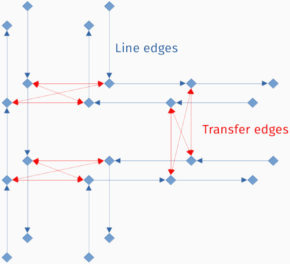
Having only lines data, the structure is a **disconnected oriented graph**.

In addition to **line edges**, it is possible to construct **transfer edges** to make the graph connected, by using, e.g.,

- Superstops names,
- Pedestrian time,
- Distance.

With transfer edges, we have a **unilaterally connected graph**.





The problematic

This dataset offers multiple axes of research. In this presentation, we will focus on one question:

> *Knowing* (1) the network structure *and* (2) the number of passengers embarking and disembarking at each stop, *can we deduce* trajectories of the passengers *in the network ?*

Short answer: **No**.

Thank you for your attention !

Questions ?

The problematic

Exact trajectories are impossible to know, but with additional hypotheses, we can **estimate** them.

We will divide this problematic into two parts:

- The estimation of trajectories on a **single line**.
- The estimation of trajectories on the **multiple lines network**.

The single line problem

Formal problem definition

Let a line (in one direction), which have n stops. Let $\boldsymbol{\rho}_{\text{in}} = (\rho_s^{\text{in}})$ and $\boldsymbol{\rho}_{\text{out}} = (\rho_t^{\text{out}})$ be two vectors representing, respectively, the **passengers entering and leaving lines at each stop**.

We search a $(n \times n)$ **origin-destination matrix** $\mathsf{N} = (n_{st})$ where components represents

$n_{st} =$ "**the number of passengers entering line at s and leaving at t**".

These components must verify

1. $n_{st} \geq 0$,
2. $n_{s\bullet} = \rho_s^{\text{in}}$,
3. $n_{\bullet t} = \rho_t^{\text{out}}$.

($\bullet$ indicates a sum on the replaced index)

Formal problem definition

It reads:

$$
N = \begin{array}{c c}
 & \begin{array}{ccccc} \sigma_1^{\text{out}} & \sigma_2^{\text{out}} & \cdots & \sigma_{n-1}^{\text{out}} & \sigma_n^{\text{out}} \end{array} \\
\begin{array}{c} \sigma_1^{\text{in}} \\ \sigma_2^{\text{in}} \\ \vdots \\ \sigma_{n-1}^{\text{in}} \\ \sigma_n^{\text{in}} \end{array} &
\begin{pmatrix}
n_{11} & n_{12} & \cdots & n_{1,n-1} & n_{1n} \\
n_{21} & n_{22} & \cdots & n_{2,n-1} & n_{2n} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
n_{n-1,1} & n_{n-1,2} & \cdots & n_{n-1,n-1} & n_{n-1,n} \\
n_{n,1} & n_{n,2} & \cdots & n_{n,n-1} & n_{n,n}
\end{pmatrix}
\end{array}
$$

In fact, we already know that some components are null:

$$
N = \begin{array}{c c}
 & \begin{array}{ccccc} \mathbf{0} & \sigma_2^{\text{out}} & \cdots & \sigma_{n-1}^{\text{out}} & \sigma_n^{\text{out}} \end{array} \\
\begin{array}{c} \sigma_1^{\text{in}} \\ \sigma_2^{\text{in}} \\ \vdots \\ \sigma_{n-1}^{\text{in}} \\ \mathbf{0} \end{array} &
\begin{pmatrix}
\mathbf{0} & n_{12} & \cdots & n_{1,n-1} & n_{1n} \\
\mathbf{0} & \mathbf{0} & n_{23} & \cdots & n_{2n} \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & n_{n-1,n} \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0}
\end{pmatrix}
\end{array}
$$

Formal problem definition

In this form, the problem is **ill posed**, because it has multiple solutions. An exemple of solution is to make passengers follow a **first in, first out (FIFO)** scheme.

In such a case, a principle of mathematical modeling is to find the solution which makes the **least assumptions about passenger behavior**.

References i