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

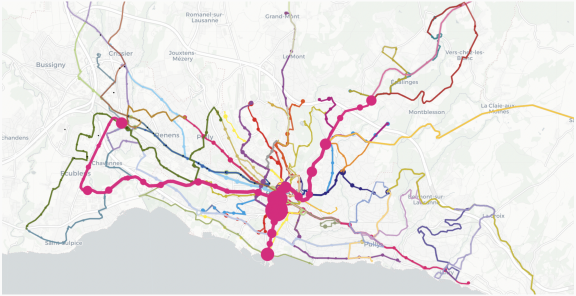
The **tl dataset**, used by Romain Loup for his PhD:

- 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
```



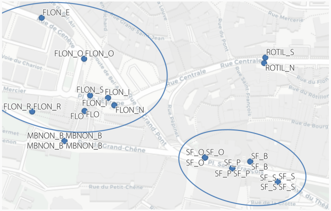
The multiple lines network

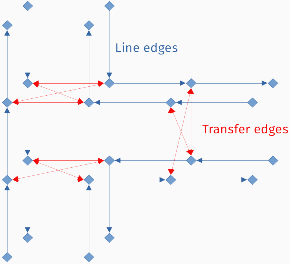
Having only lines data, the structure is a **disconnected oriented graph**.

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 ?*

The problematic

Short answer: **No**.

Thank you for your attention !

Questions ?

The problematic

Exact trajectories are impossible to know, but 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, indexed by line order. 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** $\mathbf{N} = (n_{st})$ 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:

$$\mathbf{N} = \begin{array}{c} \\ \rho_1^{\text{in}} \\ \rho_2^{\text{in}} \\ \vdots \\ \rho_{n-1}^{\text{in}} \\ \rho_n^{\text{in}} \end{array} \begin{array}{c} \begin{array}{ccccc} \rho_1^{\text{out}} & \rho_2^{\text{out}} & \cdots & \rho_{n-1}^{\text{out}} & \rho_n^{\text{out}} \end{array} \\ \begin{pmatrix} n_{11} & n_{12} & \cdots & n_{1,n-1} & n_{1n} \\ n_{21} & n_{22} & \cdots & n_{2,n-1} & n_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ n_{n-1,1} & n_{n-1,2} & \cdots & n_{n-1,n-1} & n_{n-1,n} \\ n_{n,1} & n_{n,2} & \cdots & n_{n,n-1} & n_{n,n} \end{pmatrix} \end{array}$$

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

$$\mathbf{N} = \begin{array}{c} \\ \rho_1^{\text{in}} \\ \rho_2^{\text{in}} \\ \vdots \\ \rho_{n-1}^{\text{in}} \\ \mathbf{0} \end{array} \begin{array}{c} \begin{array}{ccccc} \mathbf{0} & \rho_2^{\text{out}} & \cdots & \rho_{n-1}^{\text{out}} & \rho_n^{\text{out}} \end{array} \\ \begin{pmatrix} \mathbf{0} & n_{12} & \cdots & n_{1,n-1} & n_{1n} \\ \mathbf{0} & \mathbf{0} & n_{23} & \cdots & n_{2n} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & n_{n-1,n} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \end{pmatrix} \end{array}$$

Formal problem definition

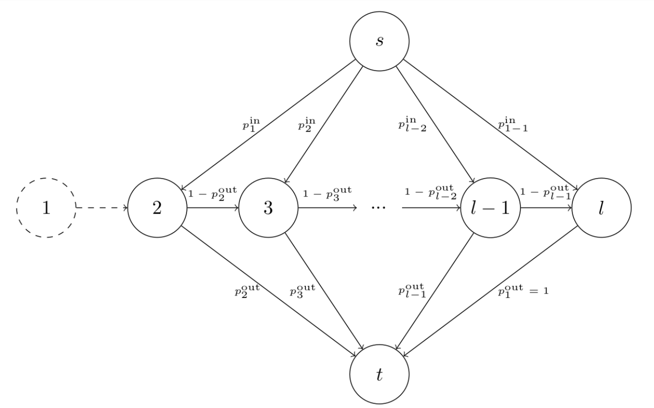
In this form, the problem is **ill posed**, because it has multiple solutions. An exemple of solution is to make passengers follow a **first in, first out (FIFO)** scheme.

A principle of mathematical modeling is to find the solution which makes the **least assumptions about passenger behavior**, in other words the **maximum entropy solution**.

In this case, it translates by supposing that there is the **same probability of leaving the line for every passenger which have traveled at least one stop**.

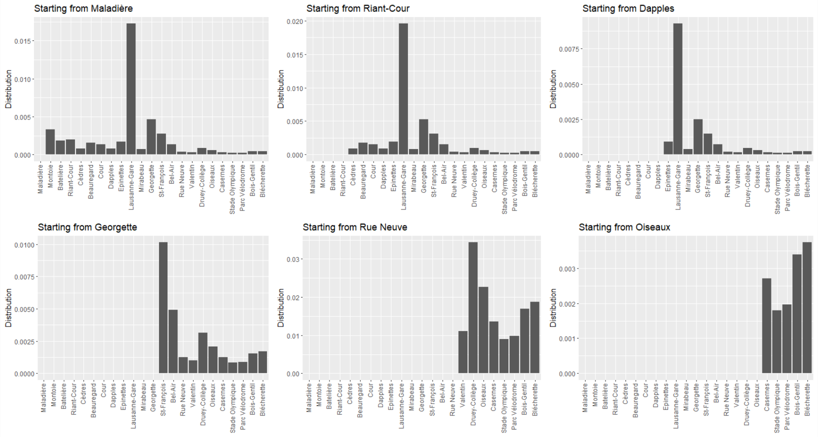
Solution with Markov chain modeling

We can then model passenger flow with a **Markov chain**:

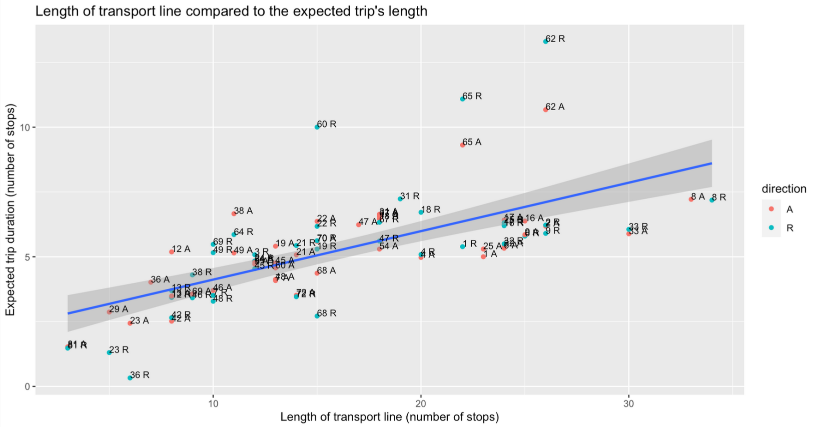


with $p_i^{\text{in}} = \frac{\rho_i^{\text{in}}}{\rho_\bullet^{\text{in}}}$ and $p_i^{\text{out}} = \frac{\rho_i^{\text{out}}}{\sum_{1 \leq k \leq (i-1)} (\rho_k^{\text{in}} - \rho_k^{\text{out}})}$.

Solution with Markov chain modeling



Solution with Markov chain modeling



Iterative proportional fitting

Iterative proportional fitting

The same solution can be obtained with the **iterative proportional fitting (IPF)** algorithm. Let

1. $\mathbf{P} = (p_{ij})$ a $(n \times m)$ matrix,
2. $\mathbf{u} = (u_i)$ a n-length vector, and
3. $\mathbf{v} = (v_i)$ a m-length vector,

all of them with strictly positive components. We can find two vectors $\mathbf{a} = (a_i)$ and $\mathbf{b} = (b_i)$ such that the matrix $\mathbf{Q} = (q_{ij})$, defined with

$$q_{ij} = a_i b_j p_{ij},$$

verifies

- $q_{i\bullet} = u_i,$
- $q_{\bullet j} = v_j,$
- $K(\mathbf{Q}|\mathbf{P}) := \sum_{ij} \frac{q_{ij}}{q_{\bullet\bullet}} \log\left(\frac{q_{ij}/q_{\bullet\bullet}}{p_{ij}/p_{\bullet\bullet}}\right)$ is minimum.

Solution with iterative proportional fitting

In our context, it means that if we define an **origin-destination affinity matrix** $\mathbf{S} = (s_{st})$ with

$$\mathbf{S} = \begin{pmatrix} 0 & 1 & \cdots & 1 & 1 \\ 0 & 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}$$

we can find the **maximum entropy solution with iterative proportional fitting**, i.e we find $\mathbf{a} = (a_s)$ and $\mathbf{b} = (b_t)$, such that $n_{ij} = a_s b_t s_{st}$ verify:

1. $n_{s\bullet} = \rho_s^{\text{in}}$,
2. $n_{\bullet t} = \rho_t^{\text{out}}$,
3. $K(\mathbf{N}|\mathbf{S})$ is minimum.

(a small number ϵ has to be added on null components).

Solution with iterative proportional fitting

By **decreasing (resp. increasing)** s_{st}, we **reduce (resp. expand)** the resulting number of passengers going from s to t obtained with IPF.

Thus, this approach is more **flexible**, because we could give a **specific affinity matrix** $\mathbf{S} = (s_{st})$, based on other data (additional assumptions).

The relationship $n_{ij} = a_s b_t s_{st}$ can be seen as a **gravity model** (still to investigate).

The multiple lines problem

References i