Estimation of flow trajectories in a multi-lines network

Guillaume Guex, Romain Loup and François Bavaud

Université de Lausanne, CH-1015, Lausanne, Suisse
guillaume.guex@unil.ch, romain.loup@unil.ch,
francois.bavaud@unil.ch

1 Introduction

2 Notations and formalism

2.1 Lines and junctions

Consider a transportation network made of bus lines numbered $\ell = 1,\ldots,q$, of respective lengths (number of stops) l_ℓ. Opposite lines, that is parallel lines running in the back and forth directions are considered as distinct.

The $l = \sum_{\ell=1}^{q} l_\ell$ bus stops constitute the nodes of the transportation network. Each stop $i = 1,\ldots,l$ belongs to a single bus line, and defines a unique next or forward stop $F(i)$ (unless i is the line terminus) and a unique backward stop $B(i)$ (unless i is the line start), both on the same line.

Junctions (super-stops), denoted $S = 1,\ldots,L$, are defined as cliques (equivalence classes) of stops which can be mutually reached from one another within walking distance. True junctions involve at least two lines, while pseudo-junctions comprise isolated stops i, possibly together with their opposite stop (noted i-) on the parallel line running in the opposite direction.

In summary, each stop i belongs to a single line $\ell(i)$ and a single junction $S(i)$. Lines and junctions define two partitions of the network nodes, with q, respectively L blocks.

2.2 Line edges, transfer edges and trips

Two sorts of oriented edges are involved in the transportation network:

- intra-line edges $(i,j) = (i,F(i))$ belonging to a single line $\ell(i) = \ell(j)$
- inter-line or transfer edges (i,j) connecting different lines $\ell(i) \neq \ell(j)$, taking place at some junction $S(i) = S(j)$.

A st-trip, noted $[s,t]$, consists of entering into the network at stop s, and leaving the network at t, by following the shortest-path (i.e. achieving the minimum distance, minimum time, or minimum cost), supposed unique, leading to s from t.

The succession of edges (ij) belonging to the st-trip, noted $(ij) \in [s,t]$, is unique. Define the edge-trip incidence matrix as

$$\chi_{ij}^{st} = \begin{cases} 1 & \text{if } (ij) \in [s,t], \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

A st-trip always starts with the edge $(s, F(s))$, and finishes with $(B(t), t)$. Transfers can occur in-between, but never at the beginning nor at the end of the trip.

Similarly, ST-trips between junctions, noted $[S, T]$, consist of the supposedly unique shortest st-trip $[s, t]$ among all $s \in S$ and $t \in T$, followed by a traveller entering the network as junction S, and leaving it at junction T. The corresponding edge-trip incidence matrix is

$$\tilde{\chi}_{ij}^{ST} = \begin{cases} 1 & \text{if } (ij) \in [S, T], \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Unicity of the shortest-path implies that a passenger wishing to travel from S to T chooses the unique stop $\sigma(S, T) \in S$ as initial node (on line $\ell(\sigma)$), and the unique node $\tau(S, T) \in T$ as final node (on line $\ell(\tau)$).

2.3 Transportation flows

Let x_{ij} count the number of travelers using edge (ij) in a given period, such as a given hour, day, week or year. The edge flow x_{ij} is denoted by y_{ij} for an intra-line edge (i, j), and by z_{ij} for a transfer edge (i, j). By construction, $x_{ij} = y_{ij} + z_{ij}$, where $y_{ij} z_{ij} = 0$.

Let a_i, respectively b_i, the number of passengers embarking, respectively disembarking at stop i. By construction,

$$\begin{cases} y_{i,F(i)} = a_i \text{ and } b_i = 0 & \text{if } i \text{ is a line start,} \\ y_{B(i),i} = b_i \text{ and } a_i = 0 & \text{if } i \text{ is a line terminus,} \\ y_{i,F(i)} = y_{B(i),i} + a_i - b_i & \text{otherwise.} \end{cases} \tag{3}$$

Also, $\mathbf{a}$ and $\mathbf{b}$ must be consistent, in the sense that $A_i \geq B_i$, where A_i (respectively B_i) is the cumulated number of embarked (resp. disembarked) passengers on the line under consideration, recursively defined as $A_{F(i)} = A_i + a_i$ (resp. $B_{F(i)} = B_i + b_i$). Moreover, $A_i = B_i$ at a terminal line stop i. This common value yields the total number of passengers transported by the line.

Let the transportation flow n_{st} denote the number of passengers following an st-trip, that is entering the network at s and leaving the network at t by using the shortest path.

Also, at the junction level, let $\tilde{n}_{ST}$ denote the number of passengers following an ST-trip. As implied by the end of section (2.2),

$$\tilde{n}_{ST} = \sum_{s \in S} \sum_{t \in T} n_{st} \qquad\qquad n_{st} = \sum_{ST} \tilde{n}_{ST}\, \delta_{s,\sigma(S,T)}\, \delta_{t,\tau(S,T)} \tag{4}$$

Definitions (1) and (2) yield

$$x_{ij} = \sum_{st} \chi_{ij}^{st}\, n_{st} = \sum_{ST} \tilde{\chi}_{ij}^{ST}\, \tilde{n}_{ST} \tag{5}$$

Among the passengers embarking in i, some transfer from another line, and some others enter into the network:

$$a_i = z_{\bullet i} + n_{i\bullet} \tag{6}$$

where "$\bullet$" denotes the summation over the replaced index, as in $n_{i\bullet} = \sum_{j=1}^{l} n_{ij}$. Similarly, among the passengers disembarking in i, some transfer to another line, and some others leave the network:

$$b_i = z_{i\bullet} + n_{\bullet i} \tag{7}$$

By construction

$$a_\bullet = b_\bullet = z_{\bullet\bullet} + n_{\bullet\bullet}$$

where $n_{\bullet\bullet}$ counts the number of passengers, and $z_{\bullet\bullet}$ counts the number of transfers. $z_{\bullet\bullet}/n_{\bullet\bullet}$ is the average number of transfers per passenger.

As explained in section 2.1, transfers can only occur at junctions, that is $z_{ij} > 0$ implies $S(i) = S(j)$. As $z_{ii} = 0$ (no traveller is supposed to disembark and re-embark later at the same stop), and $z_{ii\text{-}} = 0$ (where i- is the stop opposite to i, when existing), transfers can occur at true junctions only.

At the junction level, let

$$\tilde{a}_I = \sum_{i \in I} a_i \ , \ \tilde{b}_I = \sum_{i \in I} b_i \ , \ \tilde{z}_{\bullet I} = \sum_{i \in I} z_{\bullet i} \ , \ \tilde{z}_{I\bullet} = \sum_{i \in I} z_{i\bullet} \ , \ \tilde{n}_{I\bullet} = \sum_{i \in I} n_{i\bullet} \ , \ \tilde{n}_{\bullet I} = \sum_{i \in I} n_{\bullet i}$$

respectively denote the number of embarking (disembarking) passengers at junction I; of embarking passengers after a transfer and of disembarking passengers before a transfer at I; of passengers starting (ending) their trip at I. Then

$$\tilde{a}_I = \tilde{z}_{\bullet I} + \tilde{n}_{I\bullet} \qquad \tilde{b}_I = \tilde{z}_{I\bullet} + \tilde{n}_{\bullet I} \qquad \tilde{z}_{\bullet I} = \tilde{z}_{I\bullet} \le \min(\tilde{a}_I, \tilde{b}_I) \ . \tag{8}$$

2.4 Statement of the problem

Automatic passenger counters measure the number of passengers entering and leaving buses at each stop [Boyle, 1998], that is $\mathbf{a}$ and $\mathbf{b}$. Also, the geometry of the network permits to derive the edge-trip incidence matrix $\boldsymbol{\chi}$ (1), as well as the partition $\mathscr{S}$ of stops nested within junctions.

Intra-line edge flows $\mathbf{Y} = (y_{ij})$ can be determined by (3), but transfer edge flows $\mathbf{Z} = (z_{ij})$ are, here and typically, unknown. The objective is to estimate the $l \times l$ transportation flow $\mathbf{N} = (n_{st})$ at the stop level, or equivalently the $L \times L$ transportation flow $\tilde{\mathbf{N}} = (\tilde{n}_{ST})$ at the junction level. Many consistent solutions coexist in general, even for a single line with no transferts (section 3).

This issue of incompletely observed data can be tackled by the maximum entropy formalism: let $f_{st} = n_{st}/n_{\bullet\bullet}$ be the proportion of st-trips (empirical distribution) and let g_{st} be some prior guess on its shape (theoretical distribution). Let $\mathscr{F}$ consist of the empirical distributions $\mathbf{f}$ satisfying R linear constraints of the form $\sum_{st} f_{st}\, o^r_{st} = \bar{o}^r$ for $r = 1, \ldots, R$. Then the minimum Kullback-Leibler divergence

$$\min_{\mathbf{f} \in \mathscr{F}} \sum_{st} f_{st} \log \frac{f_{st}}{g_{st}}$$

is reached for the maximum entropy solution

$$f^0_{st} = \frac{1}{Z(\boldsymbol{\lambda})}\, g_{st} \exp(-\sum_{r=1}^{R} \lambda_r o^r_{st}) \qquad Z(\boldsymbol{\lambda}) = \sum_{st} g_{st} \exp(-\sum_{r=1}^{R} \lambda_r o^r_{st}) \tag{9}$$

where the Lagrange multipliers $\boldsymbol{\lambda}$ are determined so as to satisfy the R linear constraints (**see e.g. Bavaud and *** and references therein).

3 Single line

Let $i = 1,\ldots,l$ enumerate the bus stops in increasing order, i.e. $F(i) = i+1$. Define $g_{st} = \frac{1}{(l-1)(l-2)}\chi(s<t)$ (where $\chi(.)$ denotes the 0/1 indicator function) as the maximally uniform prior, reflecting only the unidirectional nature of trips. Constraints $n_{r\bullet} = a_r$ (resp. $n_{\bullet r} = b_r$), correspond to $o^r_{st} = \delta_{sr}$ (resp. $o^r_{st} = \delta_{tr}$) in (9), which finally yields, after reparametrization, the maximum entropy flow reads

$$n^0_{st} = I(s<t)\,c_s\,d_t \qquad \text{where} \quad \sum_{s<t} c_s\,d_t = n_{\bullet\bullet} = a_\bullet = b_\bullet \tag{10}$$

In addition, the (dis-)embarking constraints yield

$$a_i = c_i \sum_{t>i} d_t = c_i D_i \qquad\qquad b_j = d_j \sum_{s<j} c_s = d_j C_j \tag{11}$$

where $D_i = \sum_{t>i} d_t$ and $C_j = \sum_{s<j} c_s$. This maximum entropy solution constitutes one possible consistent transportation flow among many others, such as the "first in, first out" (FIFO) flow. Interestingly enough, (10) is reminiscent of the so-called gravity flows of quantitative Geography (**ref**) $n_{st} = h_{st}\,c_s\,d_t$, but with a purely asymmetric "distance deterrence function" $h_{st} = I(s<t)$. Also, (10) shows that the conditional probability to exit on t, given an entrance on s, is $w_{st} = n^0_{st}/n^0_{s\bullet} = I(s<t)\,d_t/D_s$.

*** ici la (les) figure de l'exemple "starting from Maladière, Riant-Cour, Dapples"... ? ***

*** ici l'équivalence avec l'approche chaîne de Markov de Guillaume ? ***

4 Multiple lines

*** $a_\bullet = b_\bullet$ (where "$\bullet$" denotes summation over the replaced index).

Previous submission to complex networks 2022

Statement of the problem

Automatic passenger counters measure the number of passengers entering and leaving buses at each stop [Boyle, 1998]. Given this information, can we estimate the complete trajectories of passengers within the entire multi-line network? This communication attempts to propose an estimation of all passenger trajectories in the multi-line network with an algorithm based on iterative proportional fitting (IPF) [Bishop et al., 2007].

The exploited dataset is provided by the Lausanne Transportation Agency (tl) in Switzerland. The dataset includes 42 lines of buses (or subways) and more than 1361 stops, including 497 clusters of stops (Fig. 1), carrying around 115 million passengers in 2019. Each stop refers to a single directed line, and return lines are considered as distinct. In addition to *line edges*, it is possible to construct pedestrian *transfer edges* (Fig. 2) to make the graph unilaterally connected by considering, e.g., clusters of stops connected.

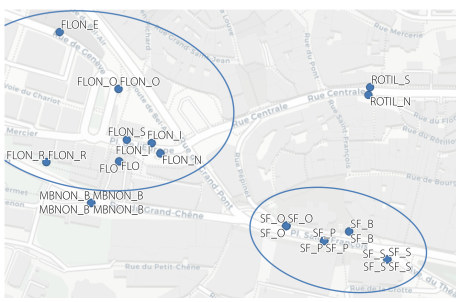

Fig. 1: Cluster of stops

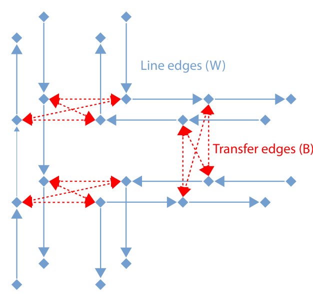

Fig. 2: The multi-line network structure

Knowing only the network structure and the number of passengers embarking and disembarking at each stop, how can we infer the most probable passenger trajectories in the network? Before examining the general multi-line network problem, we address the estimation of trajectories on a single line.

5 Single line

Consider a one-directional line with n stops indexed regarding the order found in the line. Let $\mathbf{a}_{\text{in}} = (a_s^{\text{in}})$ and $\mathbf{a}_{\text{out}} = (a_t^{\text{out}})$ be two vectors representing, respectively, the passengers entering and leaving the line at each stop. The goal is to estimate the $(n \times n)$ origin-destination matrix $\mathbf{N} = (n_{st})$, where n_{st} represents the number of passengers entering the line at s and leaving at t, subject to constraints $n_{s\bullet} = a_s^{\text{in}}$ and $n_{\bullet t} = a_t^{\text{out}}$. Among many feasible solutions, arguably the most elegant one is the maximum entropy solution, which can be derived from different means: (i) passenger flows can be modelled using a Markovian assumption, which translates by assuming that every passenger has the same probability to continue the trip after having travelled at last one stop ; (ii) an iterative proportional fitting algorithm can be performed, starting with an initial origin-destination affinity matrix $\mathbf{S} = (s_{st})$, defined as the upper triangular $n \times n$ matrix filled with 1, and then iterated to satisfy the margin constraints given by $\mathbf{a}_{\text{in}}$ and $\mathbf{a}_{\text{out}}$. Both approaches give the same solution, but only the latter remains pertinent in the multi-line problem.

6 Multi-lines

In the multi-line problem, a passenger can transfer from a line to another. The problem cannot be tackled with Markov chain modelling anymore, which generate unrealistic random trajectories. Instead, we will assume that passengers follow shortest paths. Starting from an origin-destination matrix $\mathbf{N} = (n_{st})$, where s denotes the stop at which a passenger *enters into the network* (and not simply enters a particular line), and t denotes the stop where the passenger *leaves definitively the network*, this shortest paths assumption allows us to compute the flow matrix on edges $\mathbf{X} = (x_{ij})$. The latter decomposes into the within-line flow and the transfer flow, i.e., $\mathbf{X} = \mathbf{X}_{\text{W}} + \mathbf{X}_{\text{B}}$. Moreover, in the multi-line problem, we also have to distinguish between:

- passengers who enter and leave bus lines at each stop, represented by vectors $\mathbf{a}_{\text{in}}$ and $\mathbf{a}_{\text{out}}$, which are *measured*,
- and passengers who enter and leave the network at each stop i, represented by the *unknown* quantities $n_{\bullet i}$ and $n_{i \bullet}$.

By construction,

$$a_i^{\text{in}} = n_{i\bullet} + x_{i\bullet}^{\text{B}}, \qquad\qquad a_i^{\text{out}} = n_{\bullet i} + x_{\bullet i}^{\text{B}}, \tag{12}$$

Using these two constraints, along with the shortest paths assumption and iterative proportional fitting, we propose the following iterative algorithm in order to find $\mathbf{N}$ from measured $\mathbf{a}_{\text{in}}$ and $\mathbf{a}_{\text{out}}$.

Initialisation: $\mathbf{S}^{(0)}$ is filled with 1 excepted for aberrant origin-destination pairs (such as t beeing a previous stop of the same line as s). The margins of $\mathbf{N}$ are fixed as $\mathbf{n}_{\text{in}}^{(0)} = \mathbf{a}_{\text{in}}$ and $\mathbf{n}_{\text{out}}^{(0)} = \mathbf{a}_{\text{out}}$.
Step 1, Iterative proportional fitting: We use IPF to compute $\mathbf{N}^{(r)}$ starting from $\mathbf{S}^{(r)}$, such that margin constraints, defined by $\mathbf{n}_{\text{in}}^{(r)}$ and $\mathbf{n}_{\text{out}}^{(r)}$ are satisfied.
Step 2, Shortest paths flow: Using shortest paths information, we compute $\mathbf{X}_{\text{B}}^{(r)}$ from $\mathbf{N}^{(r)}$.
Step 3, Affinity and margin update: $\mathbf{S}^{(r+1)}$, $\mathbf{n}_{\text{in}}^{(r+1)}$ and $\mathbf{n}_{\text{out}}^{(r+1)}$ are updated in order to respect constraints defined by (1).

Step 1, 2, and 3 are iterated until convergence, giving an admissible solution to the problem.

7 A small example

As an illustration, an estimated solution proposed by the algorithm on a restricted network made of four lines only is depicted on Fig. 3. A total of $n_{\bullet\bullet} = 16,837,494$ passengers using this network is estimated by the algorithm. The red circle on the bottom left represents the start of the trip s and the size of the circles at stops t represents the estimated number of passengers terminating their trip at t. In this example, the majority of passengers exit the network on the same initial embarkment line. A small fraction of them takes another line.

Table 1 represents the estimated ten most frequented transfer edges. The code of the stop represents the number of the line, the direction and a condensed name of its stop cluster. The third column gives the number (in thousands) of passengers transferring through this edge.

Fig. 3: Example map

From stop	To stop	Count
S7_A_SF_O	S9_A_SF_O	192k
S9_R_CH_E	S6_A_CH_E	187k
S7_A_SF_O	S8_R_SF_S	135k
S6_R_CH_O	S9_A_CH_O	135k
S8_A_GTE_N	S9_R_GTE_E	103k
S9_R_B-AIR_C	S8_A_B-AIR_N	99k
S9_A_SF_O	S7_A_SF_O	88k
S8_R_B-AIR_D	S6_A_B-AIR_C	87k
S6_R_SF_O	S8_A_SF_O	86k
S9_A_GTE_O	S8_R_GTE_S	84k

Table 1: List of the ten most frequented transfer edges

The current work performs computer-intensive simulations of flow over the entire network (1361 stops), permitting to extract usual network indices (centrality, betweenness...) characterizing both the stops *and* the lines. In parallel, the computational effects of various fine tuning calibration parameters used in the algorithm are investigated.

References

[Bishop et al., 2007] Bishop, Y. M., Fienberg, S. E., and Holland, P. W. (2007). *Discrete Multivariate Analysis: Theory and Practice*. Springer, New York.

[Boyle, 1998] Boyle, Douglas, M. C. D. (1998). *Passenger counting technologies and procedures*. National Academy Press, Washington, DC. OCLC: 632725908.