Civil Engineering Dimension
https://ced.petra.ac.id

Axiomatization of Transit Flow Estimation

Teknomo, K.[1]
[1] Faculty of Industrial Technology, Petra Christian University
Jl. Siwalankerto 121-131, Surabaya 60236, INDONESIA

DOI: https://doi.org/10.9744/ced.27.1.95-112

Article Info:
Submitted: Feb 02, 2025
Reviewed: Mar 01, 2025
Accepted: Mar 18, 2025

Keywords:
ideal flow,
premagic,
middle link constraint,
left-right balance.

Corresponding Author:
Teknomo, K.
Faculty of Industrial Technology,
Petra Christian University
Jl. Siwalankerto 121-131, Surabaya
60236, INDONESIA
Email: kardi.teknomo@petra.ac.id

Abstract

Transit flows between stations are typically estimated indirectly using fare collecting data rather than through direct measurement. Traditional methods approximate transit flows by adapting Origin-Destination (OD) trip estimation techniques. However, these approaches have two significant limitations. First, transit link flows represent the number of passengers remaining within the transit vehicles between stations, while OD flows specifically represent passengers entering at one station and exiting at another station. Second, traditional methods rely on the assumption that a cost function is necessary without providing mathematical justification. Consequently, there lacks a robust theoretical foundation explicitly tailored for transit flow estimation. This paper addresses this gap by developing an axiomatic framework based on the Ideal Flow Network. Through systematic mathematical derivations, we identify key balance conditions and necessary constraints to achieve more accurate transit flow estimation.

This is an open access article under the CC BY license.

INTRODUCTION

Transit flow estimation between stations is crucial for transit operations management and planning of transit systems. Accurate knowledge of transit flows directly impacts service quality, passenger comfort, and overall system reliability. Unlike direct measurements, transit flows between stations are predominantly inferred from fare collection data. An inaccurate understanding of transit flows can lead to overcrowding, negatively affecting passenger satisfaction and potentially undermining the structural and operational integrity of transit systems. A practical illustration of such challenges occurred with Metro Manila's MRT3 system in 2017 [1], demonstrating the necessity for reliable transit flow estimations.

Transit flow estimation might initially seem analogous to Origin-Destination (OD) matrix estimation in trip distribution analysis. However, these two concepts differ fundamentally. The transit link flow refers to the passengers currently onboard as the train traverses between stations, whereas OD flow focuses on passengers entering at an origin station and alighting at a destination station.

Traditional OD estimation methods typically rely on cost functions and heuristic approaches [2], such as the gravity model [3] or linear programming techniques [4], to estimate passenger distributions across networks. These methods assume passenger distributions based on trip costs or other heuristics without mathematical validation specific to transit systems. Although widely adopted, these approaches lack conclusive mathematical justification specifically for transit flows. The direct applicability and conditions for validity of these assumptions within transit scenarios remain ambiguous. Consequently, there is a pressing need for a theoretically justified methodology explicitly addressing transit flow estimation.

Note : Discussion is expected before July, 1st 2025, and will be published in the "Civil Engineering Dimension", volume 27, number 2, September 2025.
ISSN : 1410-9530 print / 1979-570X online
Published by : **Petra Christian University**

In response to these limitations, this paper proposes a rigorous axiomatic theory for transit flow estimation specifically tailored to transit lines. The approach presented is based on an Ideal Flow Network (IFN) [5], which provides a mathematically justified framework for evaluating transit flows within single transit lines with two-way connections between consecutive stations. We shall bound our space to include a single transit line of any number of stations $N \geq 2$ with two-way links between any two consecutive stations.

The axiomatic approach is preferable due to an integrated assertiveness, which give us solid theoretical foundation. Through the series of deductions into propositions and theorem, we show the conditions and constraints of the nature. Through rigorous axiomatic development, this theory establishes clear conditions and constraints that guide practical transit flow estimation methods. Thus, our contribution fills a significant theoretical gap by offering clarity, rigor, and practical guidelines for transit flow estimation. The numerical examples are provided in the Appendix below.

Terminologies and General Definitions

In this paper, we focus on a single transit line operating in two directions with a minimum of two stations. To maintain clarity, several terms are explicitly defined:

Definition 1
Transit link flow f_{pq} refers to the number of passengers within trains traversing a link pq between two consecutive stations. The term *node* represents a *station* and *link* represents directional line segment connecting two consecutive stations. The *flow* is the number of passengers in a link. As our convention, the flow has range from zero or positive infinity. Negative flow is not allowed by this convention. *Inflow* denotes passengers entering a station. *Outflow* indicates passengers exiting the station. *Start station* is the first station in the transit line. Since we have two directions, we use the convention that the most left station would be the start station when the train move from left to right and the rightest station would be the start station when the train moves from right to left. *End station* is the final station in the transit line, depending on the direction of the train's movement. *Middle station*, if exist, is a station between the start and the end station. *Middle link* is a link between two consecutive middle stations.

External environment is represented by a "cloud node" or station z, which symbolizes everything external to the transit system. *Entry flow* of a station, denoted as g_k, is the number of passengers inflow to station k from the external environment (i.e. from outside of the station, excluding those who are already onboard the trains). *Exit flow* of a station h_k is the number of passengers leaving a station k toward the external environment (i.e. including those who came out of the trains).

In a network theory terms, a *sink node* only has inflows while a *source node* only emanates flow without receiving any. A network is *strongly connected* if there is a directed path between every pair of nodes. A *component* is a sub-network that is strongly connected within itself. Specific components such as *sink component* and *source component* indicate isolated part of network with restricted directional flows.

Through these definitions, we set the foundational terminology necessary for clear and precise mathematical analysis in subsequent sections.

Derivation of Transit Flow Estimation

In this section, we derive our main proposed method of estimation of transit flow based on Ideal Flow Network (IFN) [6].

<u>Axioms for Transit Flow Estimation</u>
To develop a foundation for transit flow estimation, we introduce three fundamental axioms:
Axiom 1: ***Flow conservation****: In each station, the total of inflow is equal to the total of outflow.*
Axiom 1 is the principle ensuring continuity and consistency in passenger flows within the network. This axiom is known as Kirchhoff Law, originated in electricity circuits [7].

Axiom 2: ***Single External Environment****: the external environment is represented by a single cloud node.*
The transit network interacts with a single external environment, represented by one cloud node. This axiom ensures network connectivity and irreducibility of the network matrix. The direct consequences of Axiom 2 is in Proposition 1 that the network is strongly connected and in Proposition 2 that the pattern matrix is irreducible.

Axiom 3: ***No Immediate passing-through passengers****: at any station, passenger who goes into the station will not immediately goes out of the same station.*
Passengers entering a station cannot exit immediately from the same station. This axiom provides realistic constraints by eliminating immediate passenger re-boarding at the same station, thus simplifying the estimation model. The immediate consequences of Axiom 3 is in Proposition 5 and Proposition 6 to give us start station constraint and end station constraint.

These three axioms form the foundation of our theoretical framework and lead directly to crucial constraints and conditions essential for accurate transit flow estimation.

Conditions
To further specify scenarios, certain practical conditions are introduced as optional assumptions that simplify analysis under specific circumstances. While the mathematical proof is important to establish the theory, the conditions give the interpretation of the formulas needed for the applications. The following conditions are useful to derive the special properties, which are not in the general theory of transit flow estimation. These conditions shall serve as optional postulates that should be hold true only in order to deduce those special properties in later section below.

Condition 1: ***No turning-back flow****: if at any station, passenger who come from link ij will not immediately come back to link ji.*
Passengers entering a station via a particular link cannot immediately return on the same link. This avoids unrealistic immediate reversals of travel direction.

Condition 2: ***No passenger alighting until station $\boldsymbol{p}$****: if no passenger was alighting from start station up to this station $\boldsymbol{p}$.*
In certain scenarios, no passengers disembark from the start station up to a designated station p. This condition simplifies flow calculations for specific cases.

Condition 3: ***Complete Alighting up to station $\boldsymbol{p}$****: if all passengers who board the transit from previous stations up to this station alighting by reaching this station.*
All passengers who board trains from preceding stations alight by reaching a designated station q, providing another scenario-based simplification for specialized transit flow analysis.

These three conditions simplify and clarify specific scenarios, ensuring precise estimation in targeted cases without generalizing excessively.

Principles
The following general mathematical principles support our theoretical derivations:
Principle 1: ***Min-Max Exclusivity****: Suppose we are maximizing and minimizing only based on the two variables. If one of the variables is equal to the maximum value, then the other must be the minimum value. The maximum value is equal to the minimum value if and only if the two variables are equal.*

Proof:
Let $X = \max(A, B)$ and $P = \min(A, B)$.
Suppose $A \neq B$:

$$X = A \leftrightarrow P = B$$
$$X = B \leftrightarrow P = A \tag{1}$$

Suppose $A = B$:

$$A = B \leftrightarrow X = P \tag{2}$$

QED
When maximizing and minimizing two variables simultaneously, if one reaches its maximum, the other must reach its minimum, unless both variables are equal.

Principle 2: ***Min-Max Balance****: Suppose we are maximizing and minimizing only based on four variables in balance. The four variables represent two inflows and two outflows of a node. If the sum of one inflow and outflow is maximum, then the sum of the remaining inflow and outflow must be minimum to make it balance.*

$$X = \max(A, B),$$
$$Y = \max(C, D),$$
$$P = \min(A, B), \quad (3)$$
$$Q = \min(C, D),$$
$$B + C = A + D$$

Then the following relationship holds

$$X + Q = P + Y \quad (4)$$

Proof:
There are only five possible cases:
Case 0: $X = A = B$
Because $B + C = A + D$, then $X = A = B = P \leftrightarrow Y = C = D = Q$

$$X + Q = P + Y \quad (5)$$

Case 1: $X = A > B$ and $\mathrm{Y = C > D}$

$$X = A \leftrightarrow P = B$$
$$\mathrm{Y = C \leftrightarrow Q = D} \quad (6)$$

Because $B + C = A + D$, then

$$P + Y = X + Q \quad (7)$$

Case 2: $X = B > A$ and $\mathrm{Y = D > C}$

$$X = B \leftrightarrow P = A$$
$$\mathrm{Y = D \leftrightarrow Q = C} \quad (8)$$

Because $B + C = A + D$, then

$$X + Q = P + Y \quad (9)$$

Case 3: $X = A > B$ and $\mathrm{Y = D > C}$
We will prove that this case is invalid using contradiction.
Since $A > B$, we can write $A = B + e$, where $e > 0$.
Since $D > C$, we can write $D = C + r$, where $r > 0$.
The balance equation dictates that $B + C = A + D = (B + e) + (C + r) = B + C + (e + r)$.
However, this contradicts the previous statements that $e > 0$ and $r > 0$.
Thus, case 3 is invalid. The balance equation would only valid if $e = r = 0$, which the same is as case 0.

Case 4: $X = B > A$ and $\mathrm{Y = C > D}$
We will prove that this case is also invalid using contradiction.
Since $A < B$, we can write $A = B - e$, where $e > 0$.
Since $D < C$, we can write $D = C - r$, where $r > 0$.
The balance equation dictates that $B + C = A + D = (B - e) + (C - r) = B + C - (e + r)$.
However, this contradicts the previous statements that $e > 0$ and $r > 0$.
Thus, case 4 is invalid.
Out of all five possible cases, the three valid cases showed the given relationship.
QED.

If two inflows and two outflows at a node are balanced, maximizing one inflow-outflow pair necessitates minimizing the complementary pair. This balance is essential for maintaining flow equilibrium across nodes.

These three principles provide foundational logic essential for deriving constraints and interpreting flow conditions clearly and rigorously.

General Theory of Transit Flow Estimation

Networks can be classified based on connectivity: separate, weakly connected, or strongly connected. Our axiomatic framework ensures a strongly connected network by introducing the single external environment (cloud node). The following proposition guarantees that the network we have is strongly connected based on Axiom 2. Based on the following proposition, we can always convert a weakly connected network into a strongly connected network by adding a cloud node and a set of dummy links.

*Proposition 1: **Strongly Connected networks**: if a single cloud node is added to weakly connected network, then the network is irreducible.*

Proof:
A weakly connected network is exhaustively contained either a sink node, a source node, a sink component or a source component. The following technique can be used to make a weakly connected network into a strongly connected network:

- If a sink node exists, a dummy link with equal capacity of the sink node value would be directed from sink to a cloud node.
- If a source node exists, a dummy link with equal capacity of the source node value would be directed from a cloud node to the source node.
- If a sink component exists, dummy links from *all* nodes in the sink component would be directed to the cloud node. The capacity of the dummy link would be set to be equal to each node value it is connected to.
- If a source component exists, dummy links to *all* nodes in the sink component would be directed from the cloud node. The capacity of the dummy link would be set to be equal to each node value it is connected to.

Since we have addressed exhaustively all possible subsets of weakly connected component using a single cloud node, thus it proves the proposition. QED

Incorporating a single external environment transforms weakly connected networks into strongly connected networks, ensuring comprehensive interconnectivity.

*Definition 2: **Reducible and Irreducible**: A nonnegative matrix **A** is said to be reducible if there is a permutation matrix **P** such that it can be decomposed into submatrices with at least one zero submatrix in the off diagonal (block lower or upper triangular matrix).*

$$\mathbf{PAP}^T = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix} \tag{10}$$

*If no such permutation matrix exists, matrix **A** is called irreducible.*

*Definition 3: **Premagic:** A square nonnegative matrix $\boldsymbol{A} = [\boldsymbol{a_{ij}}]$ is called premagic if the sum of elements in each of its rows equals the sum of elements in the corresponding column.*

$$\sum_{j=1}^{n} a_{ij} = \sum_{j=1}^{n} a_{ji}\,, i = 1, \dots, n \tag{11}$$

In matrix notation,

$$\mathbf{Aj} = \mathbf{A}^T\mathbf{j} \tag{12}$$

*Proposition 2: **Irreducibility of Strongly Connected Networks**: Strongly connected network has irreducible adjacency matrix.*

Proof:
The mathematical proof that a matrix is irreducible if and only if its directed graph is connected can be found in [8]. QED

Strong connectivity guarantees that the network's adjacency matrix is irreducible, a key property in transit flow estimation.

Definition 4: ***Ideal flow matrix*** *is defined as non-negative irreducible premagic matrix [6]. The corresponding directed graph of the ideal flow matrix is called ideal flow network.*

The following proposition stated that the first two axioms lead to ideal flow network.

Proposition 3: ***Ideal flow network****: A network that satisfied Axiom 1 and Axiom 2 is an ideal flow network.*

Proof:
Based on Proposition 1, Axiom 2 leads to strongly connected network, which has irreducible adjacency matrix as stated in Proposition 2. A matrix that satisfied Axiom 1 is a premagic matrix. By Definition 4, we have an ideal flow network which adjacency matrix is irreducible premagic. QED

Figure 1 will be helpful in understanding the notations in the next propositions.

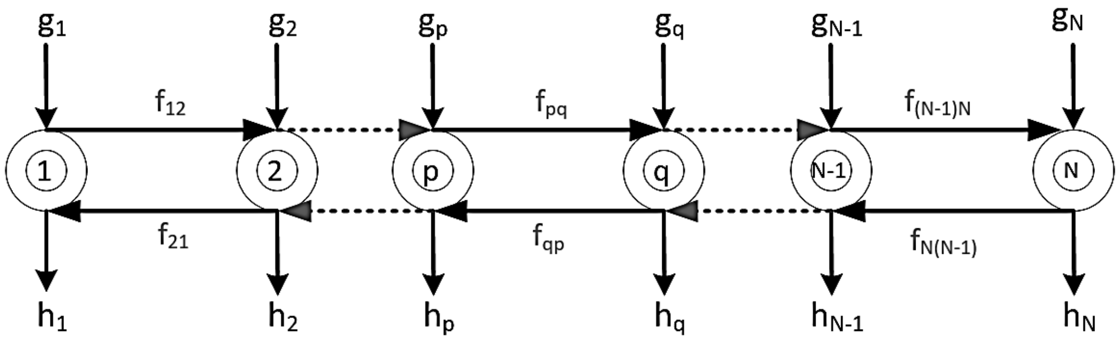


Figure 1. Transit Line of N Stations

Proposition 4: ***Conservation of Entry and Exit Flow****: The sum of all entry flows is equal to the sum of all exit flow.*

$$\sum_{k=1}^{N} g_k = \sum_{k=1}^{N} h_k \tag{13}$$

Proof.
Axiom 2 stated that we have only a single cloud node. The sum of all entry flows is equal the sum of outflow of the cloud node. The sum of all exit flow is equal to the sum of inflow of the cloud node. Since Axiom 1 said the flow in cloud node is also conversed, then the proposition is proved. QED.
Total entry flows across all stations equal total exit flows, maintaining flow conservation throughout the network.

Proposition 5: ***Starting Station Constraint****: the flow of between the start station and the immediate next station is equal to the entry flow of the start station.*

Proof:
Due to flow conservation of Axiom 1, from the start station 1, the entry flow g_1 can only be directed to either as exit flow $h_1$or link flow f_{12}. By Axiom 3, immediate passing through flow from the entry g_1 directly to the exit in the same station $h_1$is not allowed, thus, the only possibility is to make$f_{12} = g_1$. With similar argument for the start station N, we have $f_{N,N-1} = g_N$. QED

The flow from the starting station to the next immediate station equals the entry flow at the start station, directly from flow conservation and Axiom 3. Similarly, the flow into the final station from its preceding station equals the exit flow from the final station as stated in the following proposition.

Proposition 6: ***End Station Constraint****: the flow of between the end station and the immediate previous station is equal to the exit flow of the end station.*

Proof:
Due to flow conservation of Axiom 1, the only possible inflow to the end station 1 comes from either g_1 or link flow f_{21}. By Axiom 3, immediate passing through flow from the entry flow g_1 directly to the exit flow $h_1$is not allowed, thus, the only possibility is to make $f_{21} = h_1$. With similar argument for the end station N, we have $f_{N-1,N} = h_N$. QED

Proposition 7: ***Left and Right Balance****: Suppose we have link pq or link qp, the sum of all exit flow minus the sum of all entry flow on the left side of this link would be exactly equal to the sum of all entry flow minus the sum of all exit flow in all the stations of the right side of this middle link.*

$$\sum_{k=1}^{p} h_k - \sum_{k=1}^{p} g_k = \sum_{k=q}^{N} g_k - \sum_{k=q}^{N} h_k \tag{14}$$

Proof:
We can rearrange

$$\sum_{k=1}^{p} h_k + \sum_{k=q}^{N} h_k = \sum_{k=1}^{p} g_k + \sum_{k=q}^{N} g_k \tag{15}$$

$$\sum_{k=1}^{N} h_k = \sum_{k=1}^{N} g_k \tag{16}$$

QED
The total exit flow minus entry flow on one side of a middle link equals the entry flow minus exit flow on the opposite side, ensuring balance and continuity across middle links.

In the following propositions, it is useful to visualize the transit line as cluster of super nodes. Each super node is a component. To cluster into two components, we combine all the nodes on the left of link pq as a super node L and we combine all the nodes on the right of link pq as a super node R. Thus, we have a strongly connected network of three nodes including the cloud node z.

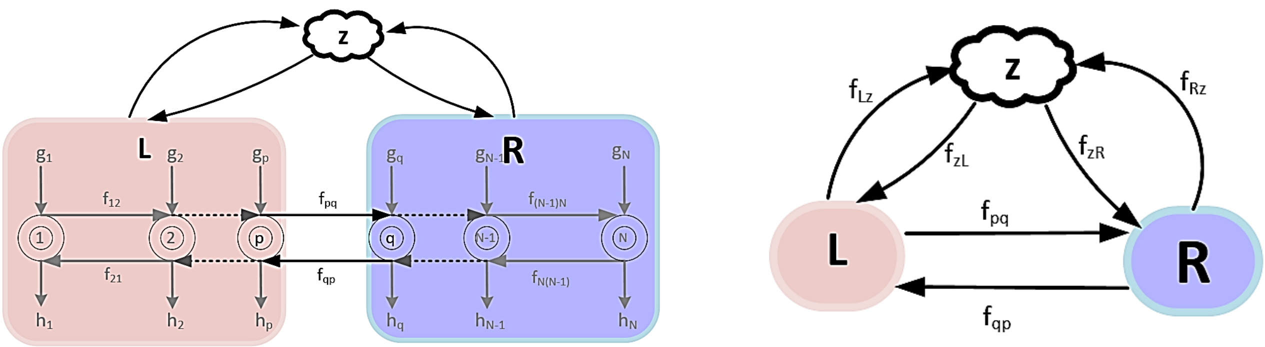


Figure 2. Transit Line as Two Super Nodes on The Left and Right of Middle Links

Definition 5

$$\begin{aligned} f_{Lz} &= \sum_{k=1}^{p} h_k \\ f_{zL} &= \sum_{k=1}^{p} g_k \\ f_{zR} &= \sum_{k=q}^{N} g_k \\ f_{Rz} &= \sum_{k=q}^{N} h_k \end{aligned} \tag{17}$$

The sum of exit flow minus the sum of entry flow on the left of a middle link pq is always equal to the sum of entry flow minus the sum of exit flow on the right of the same middle link pq. This phenomenon is stated in the following proposition.

Proposition 8: ***Left and Right Balance (shorter version)****:*

$$\Delta_{qp} = \Delta_{pq} \tag{18}$$

Proof:
Based on Proposition 7, we set

$$\Delta_{qp} = f_{Lz} - f_{zL}$$
$$\Delta_{pq} = f_{zR} - f_{Rz} \tag{19}$$

QED.

Theorem 1 below provides us with the constraint relationship between the two link flows of the same consecutive stations, f_{pq} and f_{qp}. If one flow is known then we can find the other flow. This theorem establishes a direct relationship between flows in consecutive middle links, providing necessary conditions to ensure consistent transit flow.

*Theorem 1: **Middle Link Constraint:** Suppose we have link* pq *is a middle link. Let*

$$\Delta_{qp} = \sum_{k=1}^{p} h_k - \sum_{k=1}^{p} g_k \tag{20}$$

Then

$$f_{pq} = f_{qp} - \Delta_{qp} = f_{qp} - \Delta_{pq} \tag{21}$$

Proof:
At super node L, the flow balance equation is

$$f_{zL} + f_{qp} = f_{Lz} + f_{pq}$$
$$f_{pq} = f_{qp} + f_{zL} - f_{Lz}$$
$$f_{pq} = f_{qp} - (f_{Lz} - f_{zL})$$
$$f_{pq} = f_{qp} - \Delta_{qp} \tag{22}$$

Note that

$$f_{zL} = \sum_{k=1}^{p} g_k$$
$$f_{Lz} = \sum_{k=1}^{p} h_k$$
$$\Delta_{qp} = f_{Lz} - f_{zL} = \sum_{k=1}^{p} h_k - \sum_{k=1}^{p} g_k \tag{23}$$

QED.

The middle link constraints (Theorem 1) give us clue on which is larger between f_{pq} and f_{qp} in the following corollary. Corollaries derived from these propositions introduce practical boundaries for middle link flows, assisting real-world estimations.

Corollary 1

$$\text{If } \Delta_{pq} = \Delta_{qp} > 0 \text{ then } f_{pq} < f_{qp}$$
$$\text{If } \Delta_{pq} = \Delta_{qp} < 0 \text{ then } f_{pq} > f_{qp}$$
$$\text{If } \Delta_{pq} = \Delta_{qp} = 0 \text{ then } f_{pq} = f_{qp} \tag{24}$$

The proof is straightforward consequence of the middle link constraint (Theorem 1) that make the flow non-negative.

Our goal is to provide transit flow estimation. Theorem 1 provides us with middle link constraint. Note, however, using this constraint alone, both link flow can go from zero to infinity. Thus, we need to set up the boundary of the estimation. Variable f_{pq}^{Δ} should be read as "f-pq-up" and f_{pq}^{∇} as "f-pq-down". Observe the Definition 6 mirrors Principle 2.

Definition 6

$$f_{pq}^{\Delta} = \max\{f_{zL}, f_{Rz}\}$$
$$f_{qp}^{\Delta} = \max\{f_{Lz}, f_{zR}\}$$
$$f_{pq}^{\nabla} = \min\{f_{zL}, f_{Rz}\}$$
$$f_{qp}^{\nabla} = \min\{f_{Lz}, f_{zR}\} \tag{25}$$

Corollary 2
Four Equivalences

$$f_{pq}^{\Delta} = f_{zL} \text{ iff } f_{pq}^{\nabla} = f_{Rz}$$
$$f_{pq}^{\Delta} = f_{Rz} \text{ iff } f_{pq}^{\nabla} = f_{zL}$$
$$f_{qp}^{\Delta} = f_{Lz} \text{ iff } f_{qp}^{\nabla} = f_{zR}$$
$$f_{qp}^{\Delta} = f_{zR} \text{ iff } f_{qp}^{\nabla} = f_{Lz} \quad (26)$$

Proof:
These are direct consequences of the Definition 6 and Principle 1 because two variables cannot become both maximum and minimum at the same time unless they are equal. QED.

Suppose a node has only two inflows and two outflows. If we maximize the sum of one inflow and one outflow, then the sum of the remaining inflow and outflow must be minimum to make it balance. The following lemma stated this principle in a formula.

Lemma 1*: Min-Max Balance*

$$f_{pq}^{\Delta} + f_{qp}^{\nabla} = f_{pq}^{\nabla} + f_{qp}^{\Delta} \quad (27)$$

Proof:
Consider inflows and outflows in the cloud node z:

$$\max\{f_{zL}, f_{Rz}\} + \min\{f_{Lz}, f_{zR}\} = \min\{f_{zL}, f_{Rz}\} + \max\{f_{Lz}, f_{zR}\} = \sum_{k=1}^{N} h_k = \sum_{k=1}^{N} g_k \quad (28)$$

In cloud node z, there are two inflows and two outflows. If the sum of one inflow and outflow is maximum, then the sum of the remaining inflow and outflow must be minimum to make it balance. If we use the notations in Principle 2, this is exactly the same as $X + Q = P + Y$.

Corollary 3

$$f_{pq}^{\Delta} - f_{pq}^{\nabla} = f_{qp}^{\Delta} - f_{qp}^{\nabla} \quad (29)$$

Proof:
Rearranging the result of Lemma *1*, we have

$$f_{qp}^{\nabla} - f_{pq}^{\nabla} = f_{qp}^{\Delta} - f_{pq}^{\Delta} \quad (30)$$

Multiply both side by -1 we have the result of Corollary 3
QED.

Corollary 4

$$f_{qp}^{\Delta} = f_{pq}^{\Delta} \text{ iff } f_{qp}^{\nabla} = f_{pq}^{\nabla} \quad (31)$$

Proof:
Based on the proof of Corollary 3, we have

$$f_{qp}^{\nabla} - f_{pq}^{\nabla} = f_{qp}^{\Delta} - f_{pq}^{\Delta} \quad (32)$$

$$\text{If } f_{qp}^{\Delta} = f_{pq}^{\Delta} \text{ then } f_{qp}^{\nabla} = f_{pq}^{\nabla} \quad (33)$$

QED

The following statement is a very special case where they are equal.

Corollary 5

$$f_{qp}^{\Delta} = f_{pq}^{\Delta} \leftrightarrow f_{qp}^{\nabla} = f_{pq}^{\nabla} \leftrightarrow f_{zL} = f_{Rz} = f_{Lz} = f_{zR} \quad (34)$$

Proof:
By definition, $f_{pq}^{\Delta} = \max\{f_{zL}, f_{Rz}\}$ and $f_{qp}^{\Delta} = \max\{f_{Lz}, f_{zR}\}$. Based on Principle 1 they cannot become maximum unless they are equal. Similar arguments go for the second equivalence. Another way to prove is to use of Case 0 of Principle 2 by setting $X = Y$.

QED.

Corollary 6

$$f_{qp}^{\nabla} - f_{pq}^{\nabla} = f_{qp}^{\Delta} - f_{pq}^{\Delta}$$
$$f_{pq}^{\Delta} - f_{pq}^{\nabla} = f_{qp}^{\Delta} - f_{qp}^{\nabla} \quad (35)$$

Proof:
These are simple rearrangements of Lemma *1*. QED

The following theorem shows that $\Delta_{qp} = \Delta_{pq}$ is the invariant.
Theorem 2

$$f_{qp}^{\nabla} - f_{pq}^{\nabla} = \Delta_{qp}$$
$$f_{qp}^{\Delta} - f_{pq}^{\Delta} = \Delta_{qp} \quad (36)$$

Proof:
Lemma *1* stated that $f_{pq}^{\Delta} + f_{qp}^{\nabla} = f_{pq}^{\nabla} + f_{qp}^{\Delta}$. We can rearranged it into Corollary 6 $f_{qp}^{\nabla} - f_{pq}^{\nabla} = f_{qp}^{\Delta} - f_{pq}^{\Delta}$, which is equivalent to left and right balance in Proposition 8. Thus, we only need to prove that $f_{qp}^{\nabla} - f_{pq}^{\nabla} = \Delta_{qp}$ to prove the second equation because the first equation is exactly equal. Based on the proof of Principle 2, we have only two valid cases for the inequality:
Case 1: $f_{pq}^{\Delta} = f_{zL} > f_{Rz}$ and $f_{qp}^{\Delta} = f_{Lz} > f_{zR}$

$$\Delta_{qp} = f_{Lz} - f_{zL} = f_{qp}^{\Delta} - f_{pq}^{\Delta} \quad (37)$$

Case 2: $f_{pq}^{\Delta} = f_{Rz} > f_{zL}$ and $f_{qp}^{\Delta} = f_{zR} > f_{Lz}$

$$\Delta_{pq} = f_{zR} - f_{Rz} = f_{qp}^{\Delta} - f_{pq}^{\Delta} \quad (38)$$

QED.

In practice, we suggest the following boundary.

Corollary 7
The minimum middle link flow is $\{f_{qp}^{\nabla}, f_{pq}^{\nabla}\}$ and the maximum middle link flow is $\{f_{qp}^{\Delta}, f_{pq}^{\Delta}\}$
Proof:
Theorem 2 give us guarantee that the minimum and the maximum middle link flow pair satisfied the middle link constraint.
QED.

Corollary 8 gives us a hint that we can test correctness of the computation by forming ideal flow matrix which is non-negative irreducible and premagic.

Corollary 8
Transit Flow estimation is valid if and only if it can form the ideal flow matrix.
Proof:
Forming ideal flow matrix from the estimation is done by adding one cloud node at the end. The flow to the cloud node would be equal to the exit flows and the entries from the cloud node would be equal to the entry flows. Since ideal flow is irreducible and premagic, the balance equation in each station and the cloud node is guaranteed.
QED.

Derivation of Special Properties

In this section, we derive special properties that only hold true under certain conditions. They give us hints for interpretation of the formulas and clues on why we select certain variables to be used in Definition 5 and Definition 6 that eventually lead to more general theory of transit flow estimation. Special properties provide additional insights under specific scenarios, clearly defined by conditions 1, 2, and 3 in earlier section. Condition 1 is used to give the interpretation of the maximum middle link flow that the maximum in here means without returning back to the previous link as derived in Proposition 9.

Proposition 9: ***Middle Link Maximum Flow****: the maximum flow of a middle link pq without turning back flow is equal to the sum of exit flow from q to the end station.*

$$f_{pq}^{\#} = f_{Rz} = \sum_{k=q}^{N} h_k$$
$$f_{qp}^{\#} = f_{Lz} = \sum_{k=1}^{p} h_k \tag{39}$$

Proof:
If $f_{pq} > \sum_{k=q}^{N} h_k$ it means there is an excess flow inside f_{pq} that will turn back as part of f_{qp}. However, Condition 1 states that turning back flow is not allowed and therefore $f_{pq} = \sum_{k=q}^{N} h_k$ is the maximum flow in link pq without turning back flow. With similar argument, we know that if $f_{qp} > \sum_{k=1}^{p} h_k$ there will be an excess flow inside f_{qp} that will turn back as part of f_{pq}. Since we do not allow these excess back flow, then $f_{qp} = \sum_{k=1}^{p} h_k$ is the maximum flow in link qp with no turn back flow. QED
Without immediate return flows, the maximum feasible flow on a middle link is the sum of exit flows from a specific station onward.

Proposition 10: ***No Alighting Passenger Scenario:*** *Assuming there is no passenger going down in any station from start up to station p and among those station there is no turning back passengers either, then the remaining total number of passengers is equal to f_{zL}. Similarly, the way back, if no passenger going down in any station from start up to station q and among those station there is no turning back passengers either, then the remaining total number of passengers is equal to f_{zR}.*

Proof:
Let all stations from the start station up to station p denoted by subscript zL, then the implication of Condition 1 and Condition 2 is

$$f_{pq}^{*} = f_{zL} = \sum_{k=1}^{p} g_k \tag{40}$$

$$f_{qp}^{*} = f_{zR} = \sum_{k=q}^{N} g_k \tag{41}$$

This is the accumulation of all entry flows from the start station up to middle station p or q, depending on the context. Since in any station there is no turning back passenger, then the above formulas are the only possibility.
QED.
If no passengers alight from the start station up to a particular station, the flow is a simple cumulative total of entry flows.

Proposition 11: ***Complete Alighting Scenario:*** *Assuming there is all passenger going down in any station from start up to a middle station and among those station there is no turning back passengers either, then the remaining total number of passengers is equal to*

$$f_{pq}^{b} = \sum_{k=1}^{p} g_k - \sum_{k=2}^{p} h_k \tag{42}$$

$$f_{qp}^{b} = \sum_{k=q}^{N} g_k - \sum_{k=q}^{N-1} h_k \tag{43}$$

Proof:
Axiom 3 prohibits the passing through passengers within one station, thus passengers come from start station cannot go out in the same station. When all passengers go out of the next stations as stated as Condition 3 and turning back passengers are prohibited by Condition 1, the only possibility for remaining passengers are given by the above formulas. QED

Conversely, if all passengers alight at intermediate stations without returning, the transit flow is the cumulative sum of exit flows. Note that f_{pq}^{b}, f_{qp}^{b} are undefined for the start station and f_{pq}^{b} can go beyond the boundary of $[f_{pq}^{\nabla}, f_{pq}^{\Delta}]$ and similarly f_{qp}^{b} can go beyond the boundary of $[f_{qp}^{\nabla}, f_{qp}^{\Delta}]$. The point that the flow can go beyond our expected boundary give us clue that the boundary $[f_{qp}^{\nabla}, f_{qp}^{\Delta}]$ is not the same as $[f_{qp}^{\min}, f_{qp}^{max}]$ because the theoretical boundary of $f_{qp}^{\min} = 0$ and $f_{qp}^{max} = \infty$.
Proposition 10 and Proposition 11 provide us with the extreme cases. Note that if we use these two extreme cases in our transit flow estimation, we should still make sure that the middle link constraint is (Theorem 1) still being hold. The following statement must be satisfied if we want to use other $f_{pq}^{max} \neq f_{pq}^{\Delta}$ and $f_{pq}^{min} \neq f_{pq}^{\nabla}$.

Corollary 9 and 10 establish constraints for realistic estimations between identified extremes, ensuring alignment with the middle link constraints.

Corollary 9

$$f_{pq}^{max} - f_{pq}^{min} \geq \Delta_{qp} = \Delta_{pq} \tag{44}$$

Proof:
This constraint must be satisfied such that the middle link constraint is (Theorem 1) still being hold. QED

In general, if we want to use other $f_{pq}^{max} \neq f_{pq}^{\Delta}$ and $f_{pq}^{min} \neq f_{pq}^{\nabla}$ then we can use the middle flow estimate between the given two extreme flows.

Corollary 10

$$f_{pq} = \tfrac{1}{2}\left(f_{pq}^{max} - f_{pq}^{min} + \Delta_{qp}\right) \rightarrow f_{qp} = \tfrac{1}{2}\left(f_{pq}^{max} - f_{pq}^{min} - \Delta_{qp}\right) \tag{45}$$

$$f_{pq} = \tfrac{1}{2}\left(f_{pq}^{max} - f_{pq}^{min} - \Delta_{qp}\right) \rightarrow f_{qp} = \tfrac{1}{2}\left(f_{pq}^{max} - f_{pq}^{min} + \Delta_{qp}\right) \tag{46}$$

Proof:
The middle flow estimate is equal to $\tfrac{1}{2}\left(f_{pq}^{max} - f_{pq}^{min}\right)$. To satisfy the middle link constraint (Theorem 1) we set $\tfrac{1}{2}\Delta_{qp}$ above and below the middle flow. QED

Error Distribution across Stations

When we use real world data (as shown in the appendix below), inconsistencies often arise due to temporal measurement discrepancies—passengers may still be traveling during data collection intervals. Thus, daily aggregate entry and exit flows often differ slightly. We face the problem that many of the data may not satisfy our assumption that the sum of entry flow must be exactly the same as the sum of exit flow over all stations as stated in Axiom 1. This kind of small error in measurement is common because of the travel time delay in transportation. The same passengers need some time to be transported into the destination while the data collection is based on the regular hourly interval. Thus, some of these passengers are still in the journey. The error of one-day flow must be much smaller than hourly flow because all passengers that enter must go out of the stations.

To handle this practical issue, we propose a method to distribute errors proportionally across entry and exit flows, ensuring compliance with the fundamental flow conservation axiom. Formally, given observed entry and exit flows at each station, errors are calculated as deviations from their averages. These deviations are then proportionally adjusted, ensuring that adjusted entry and exit flows satisfy the balance conditions precisely. In this section, we will explain how the error is distributed over the entry and exit data to make the data satisfy Axiom 1.

Let subscript i represents entry-flow and subscript j represents exit-flow. Suppose x_i, x_j are the entry-flow and exit-flow respectively at station i, j. (i.e. $x_i = g_i, x_j = h_j$) The average is given as

$$\bar{x} = \frac{1}{2}\left(\sum_i x_i + \sum_j x_j\right) \tag{47}$$

The error is computed as

$$e_i = \sum_i x_i - \bar{x} = \frac{1}{2}\left(\sum_i x_i + \sum_j x_j\right) \tag{48}$$

Percentage of error is defined as

$$e_i\% = \frac{100\, e_i}{\bar{x}} \tag{49}$$

The change of distribution becomes

$$\Delta x_i = \frac{e_i x_i}{\sum_i x_i} = \frac{x_i}{2}\left(1 + \frac{\sum_j x_j}{\sum_i x_i}\right) \tag{50}$$

The estimated values of exit flow and entry flow are

$$\tilde{x}_i = x_i - \Delta x_i = \frac{x_i}{2}\left(1 - \frac{\sum_j x_j}{\sum_i x_i}\right) \tag{51}$$

With the above formulations, we can provide the estimate of the real-world data that satisfy Axiom 1.

CONCLUSIONS

The following conclusions can be deduced from the results and analysis. This paper established a rigorous axiomatic theory tailored explicitly for transit flow estimation using the Ideal Flow Network approach. The primary outcomes of this theoretical framework include clearly defined constraints, such as the starting and ending station constraints, middle link constraints, and precise definitions of minimum and maximum middle link flows. These findings ensure accurate and theoretically justified transit flow estimation, overcoming significant limitations inherent in traditional Origin-Destination methods. Based on ideal flow network on a single two-way transit line, we have the following important properties.

1. The flow of between the start station and the immediate next station is equal to the entry flow of the start station because the outflow of the start station is equal to the entry flow of the start station (start station constraint).
2. The flow of between the end station and the immediate previous station is equal to the exit flow of the end station because the inflow of the end station is equal to the exit flow of the end station (end station constraint)
3. The sum of exit flow minus the sum of entry flow on the left of a middle link is always equal to the sum of entry flow minus the sum of exit flow on the right of the same middle link (left and right balance). If the sum of entry and exit passengers on the left of a middle link is called Δ_{qp} and the sum of exit and entry passengers on the right of the same middle link is called Δ_{pq}, then $\Delta_{qp} = \Delta_{pq}$.
4. Furthermore, for any middle link pq and qp, $f_{pq} = f_{qp} - \Delta_{qp} = f_{qp} - \Delta_{pq}$ (middle link constraint).
5. The maximum flow of a middle link pq without turning back flow is equal to the sum of exit flow from q to the end station.
6. The suggested reasonable low and high estimates of the middle link transit flows are $\{f_{pq}^{\nabla}, f_{qp}^{\nabla}\}, \{f_{pq}^{\Delta}, f_{qp}^{\Delta}\}$, respectively.
7. The correctness of the transit flow estimation can be tested by checking if the matrix form ideal flow matrix, which is non-negative, irreducible and premagic.

Future research could address several open problems. First, expanding the theory to consider multiple interacting transit lines would increase practical applicability. Second, developing robust methods for handling dynamic transit conditions, such as varying passenger behaviour patterns and irregular train operations, would further enhance applicability. Third, investigating methods to incorporate passenger behaviour models into the axiomatic framework could provide more realistic flow estimations.

Practically, transportation agencies are recommended to implement these axiomatic principles to enhance the accuracy of transit flow data, which can significantly improve operational decisions and service reliability. Continuous calibration with real-time passenger data and further validation through extensive empirical studies are recommended to refine and validate this theoretical approach.

REFERENCES

1. Marchadesch, B., *The MRT3's Troubles in 2017*, 2017. [Online]. Available: http://www.gmanetwork.com/news/news/specialreports/635778/the-mrt3-s-troubles-in-2017/story/. [Accessed 15 5 2018].
2. Ortuzar, J.D.D. and Willumsen, L.G., *Modelling Transport*, Essex: John Wiley & Sons, 1990.
3. Evans, S.P., A Relationship between the Gravity Model for Trip Distribution and the Transportation Problem in Linear Programming, *Transportation Research, 7*(1), 1957, pp. 39-61.
4. Wilson, A.G., *Entropy in Urban and Regional Modelling*, London: Pion, 1970.
5. Teknomo, K., Ideal Relative Flow Distribution on Directed Network, *Journal of the Eastern Asia Society for Transportation Studies, 12*, 2017, pp. 939-958.
6. Teknomo, K., Ideal Flow of Markov Chain, *Discrete Mathematics, Algorithms and Applications, 10*(06), 2018.
7. Kirchhoff, G., On the Motion of Electricity in Wires, *Philosophical Magazine, 13,* 1857, p. 393–412.
8. Axelsson, O., *Iterative Solution Methods*, New York: Cambridge University Press, 1996.
9. Department of Transportation and Communications, *MRT3 Daily Passenger Traffic per Hour per Station*, 2014. [Online]. Available: http://www.dotr.gov.ph/images/front/Data_Sets/MRT3-DailyPassengerTrafficperHourperStationforCY2014.xlsx. [Accessed 10 May 2018].

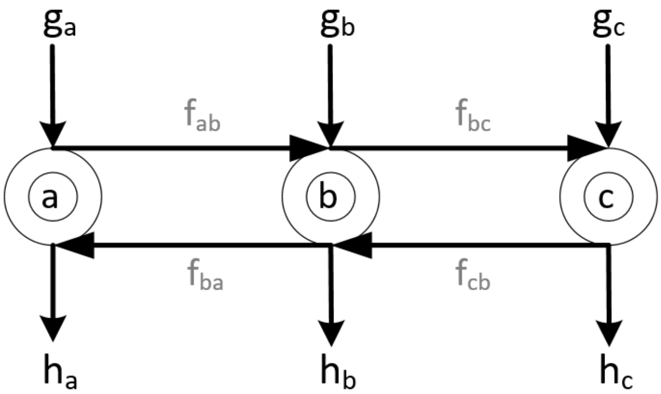
APPENDIX

Illustrative Examples

These three sections of appendices clearly illustrate the practicality and mathematical rigor of the theory, making it applicable for transit system management and providing guidance for future empirical validation.

Three Stations Computational Problem

To illustrate the complexity and challenges of transit flow estimation, consider a simple scenario involving three stations connected by a transit line. The given data includes passenger entries and exits at each station. The objective is to determine the internal link flows accurately. We have only one cloud node z but we do not draw the cloud node to avoid clutter in drawing.



$$\mathbf{F} = \begin{array}{c|cccc|c} \nearrow & a & b & c & z & sum \\ \hline a & 0 & f_{ab} & 0 & g_a & \pi_a \\ b & f_{ba} & 0 & f_{bc} & g_b & \pi_b \\ c & 0 & f_{cb} & 0 & g_c & \pi_c \\ z & h_a & h_b & h_c & 0 & \pi_z \\ \hline sum & \pi_a & \pi_b & \pi_c & \pi_z & \kappa \end{array}$$

Given passenger counts at each station, the transit flow matrix must satisfy balance equations as established by the axioms. We will show using a counter example that can be easily derived from Axiom 3. Suppose direct passing-through is prohibited, it would reduce the problem's complexity significantly. The equations for the three-station problem clearly outline eight unknown variables matched with eight linear equations, illustrating the initially unsolvable nature of the problem without applying the constraints. To resolve this, the entry and exit flows at start and end stations are explicitly matched with their adjacent link flows. This step significantly reduces complexity, enabling direct resolution of the transit flow problem. Specifically, we are given the following numbers from the station data: $g_a, g_b, g_c, h_a, h_b, h_c$

What we really want to know are the following $f_{ab}, f_{ba}, f_{bc}, f_{cb}$. All the rests $\pi_i = \{\pi_a, \pi_b, \pi_c, \pi_z\}$ are useful for the computation and should be derived from the computed values. Because of Axiom 1, we can be sure that the flow matrix is premagic and the irreducibility is automatically satisfied when we added the cloud node. In this problem, we have 8 unknown $(f_{ab}, f_{ba}, f_{bc}, f_{cb}, \pi_a, \pi_b, \pi_c, \pi_z)$ and 8 equations:

$$\begin{gathered} f_{ab} + g_a = \pi_a \\ f_{ba} + f_{bc} + g_b = \pi_b \\ f_{cb} + g_c = \pi_c \\ h_a + h_b + h_c = \pi_z \\ f_{ba} + h_a = \pi_a \\ f_{ab} + f_{cb} + h_b = \pi_b \\ f_{bc} + h_c = \pi_c \\ g_a + g_b + g_c = \pi_z \end{gathered} \tag{52}$$

Note that the only requirement from the given data is total passengers out of all stations must be exactly the same as the sum of passengers in to all stations. While this assumption is incorrect dynamically in short term, since Ideal Flow Network is about equilibrium in the long run, the average of boarding is exactly equal to alighting in each station is probably sill acceptable. This will lead to

$$g_a + g_b + g_c = \pi_z = h_a + h_b + h_c \tag{53}$$

Now we rearrange the 8 equations above such that the unknown variables are on the left-hand side and the known values are on the right-hand side.

$$\begin{gathered} f_{ab} - \pi_a = -g_a \\ f_{ba} + f_{bc} - \pi_b = -g_b \\ f_{cb} - \pi_c = -g_c \\ f_{ba} - \pi_a = -h_a \\ f_{ab} + f_{cb} - \pi_b = -h_b \\ f_{bc} - \pi_c = -h_c \\ \pi_z = h_a + h_b + h_c \\ \pi_z = g_a + g_b + g_c \end{gathered} \tag{54}$$

The resulting solution is clearly expressed in a structured matrix form. Explicit solutions are provided based on clearly defined balance equations. Each variable is distinctly associated with entry and exit flows at each station. Each row in matrix **B** represents each equation, each column is to indicate the unknown variable. The right-hand side vector consists of constants from the given data.

$$\begin{bmatrix} 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f_{ab} \\ f_{ba} \\ f_{bc} \\ f_{cb} \\ \pi_a \\ \pi_b \\ \pi_c \\ \pi_z \end{bmatrix} = \begin{bmatrix} -g_a \\ -g_b \\ -g_c \\ -h_a \\ -h_b \\ -h_c \\ h_a + h_b + h_c \\ g_a + g_b + g_c \end{bmatrix} \tag{55}$$

In matrix form

$$\mathbf{Bx} = \mathbf{c} \tag{56}$$

Unfortunately, the rank of matrix **B** is 6, not full rank. Thus, matrix **B** has no inverse. Both left inverse and right inverse also do not exist. Thus, we have proven by counter example that without Axiom 3 and Condition 1, this problem is unsolvable.

Solution of the Three Station Problem

To give the solution of three station problem, we can use Axiom 3. Using Axiom 3, additional constraints simplify the mathematical model:

- Outflow from the start station precisely equals its entry flow.
- Inflow to the end station matches its exit flow.

This simplification allows a direct, systematic solution to the three-station problem. The provided equations neatly align with the axiomatic constraints, clearly demonstrating their necessity for achieving a solvable and practical transit flow estimation. Because we do not allow passing-through within the station, any outflow from the start station would be practically determined by the entry of the start station. Similarly, any inflow to the end station would be fully determined by the exit flow of the end station. This means Axiom 3 provided us with additional two constraints:

Start station constraints:

$$\begin{gathered} f_{ab} = f_{za} = h_a \\ f_{cb} = f_{zc} = h_c \end{gathered} \tag{57}$$

End station constraints:

$$\begin{gathered} f_{ba} = f_{az} = g_a \\ f_{bc} = f_{cz} = g_c \end{gathered} \tag{58}$$

The solution is given in the following flow matrix

$$\mathbf{F} = \begin{array}{c} \nearrow \\ a \\ b \\ c \\ z \\ sum \end{array} \begin{array}{c} \begin{array}{cccc} a & b & c & z \end{array} \\ \begin{bmatrix} 0 & h_a & 0 & g_a \\ g_a & 0 & g_c & g_b \\ 0 & h_c & 0 & g_c \\ h_a & h_b & h_c & 0 \end{bmatrix} \\ \begin{array}{cccc} \pi_a & \pi_b & \pi_c & \pi_z \end{array} \end{array} \begin{array}{c} sum \\ \pi_a \\ \pi_b \\ \pi_c \\ \pi_z \\ \kappa \end{array} \tag{59}$$

Where,

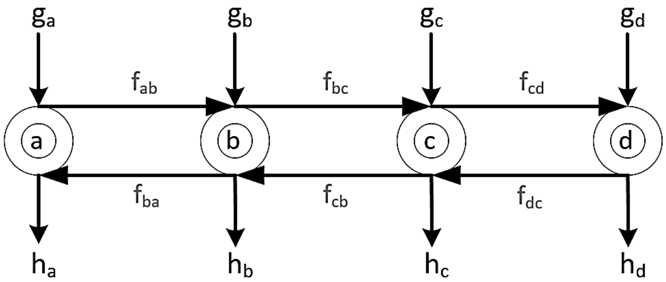
$$\kappa = 2h_a + h_b + 2h_c + 2g_a + g_b + 2g_c$$
$$f_{ab} + f_{ba} + f_{bc} + f_{cb} = h_a + h_c + g_a + g_c \tag{60}$$

While these three stations scenarios can be easily solved based on the three axioms alone, larger number of stations requires just one more additional Condition 1.

Four Stations Problem

Extending to four stations further illustrates the theory's practical application. Here, the balance equations become slightly more complex but remain manageable due to clearly defined constraints and conditions. Through systematic simplifications using the established axioms and conditions (particularly Condition 1, prohibiting immediate return flows), the four-station problem is resolved similarly. Explicitly defined constraints reduce the complexity of computations, providing a clear, rigorous solution.

The final computed matrix clearly presents the transit flows, showcasing the practical utility and mathematical robustness of the axiomatic framework.



$$\mathbf{F} = \begin{array}{c} \nearrow \\ a \\ b \\ c \\ d \\ z \\ sum \end{array} \begin{array}{c} \begin{array}{ccccc} a & b & c & d & z \end{array} \\ \begin{bmatrix} 0 & f_{ab} & 0 & 0 & g_a \\ f_{ba} & 0 & f_{bc} & 0 & g_b \\ 0 & f_{cb} & 0 & f_{cd} & g_c \\ 0 & 0 & f_{dc} & 0 & g_d \\ h_a & h_b & h_c & h_d & 0 \end{bmatrix} \\ \begin{array}{ccccc} \pi_a & \pi_b & \pi_c & \pi_d & \pi_z \end{array} \end{array} \begin{array}{c} sum \\ \pi_a \\ \pi_b \\ \pi_c \\ \pi_d \\ \pi_z \\ \kappa \end{array}$$

The node equations are

$$f_{ab} + g_a = \pi_a = f_{ba} + h_a$$
$$f_{ba} + f_{bc} + g_b = \pi_b = f_{ab} + f_{cb} + h_b$$
$$f_{cb} + f_{cd} + g_c = \pi_c = f_{bc} + h_c$$
$$f_{dc} + g_d = \pi_d = f_{cd} + h_d \tag{61}$$

Start station constraints:

$$f_{ab} = f_{za} = h_a$$
$$f_{dc} = f_{zd} = h_d \tag{62}$$

End station constraints:

$$f_{ba} = f_{az} = g_a$$
$$f_{cd} = f_{dz} = g_d \tag{63}$$

The flow matrix now has only two unknowns: f_{bc}, f_{cb}

$$\mathbf{F} = \begin{array}{c} \nearrow \\ a \\ b \\ c \\ d \\ z \\ sum \end{array} \begin{array}{c} \begin{array}{ccccc} a & b & c & d & z \end{array} \\ \begin{bmatrix} 0 & h_a & 0 & 0 & g_a \\ g_a & 0 & f_{bc} & 0 & g_b \\ 0 & f_{cb} & 0 & g_d & g_c \\ 0 & 0 & h_d & 0 & g_d \\ h_a & h_b & h_c & h_d & 0 \end{bmatrix} \\ \begin{array}{ccccc} \pi_a & \pi_b & \pi_c & \pi_d & \pi_z \end{array} \end{array} \begin{array}{c} sum \\ \pi_a \\ \pi_b \\ \pi_c \\ \pi_d \\ \pi_z \\ \kappa \end{array} \tag{64}$$

Now we put the start station constraints and end station constraints to the node equations

$$h_a + g_a = \pi_a = g_a + h_a$$
$$g_a + f_{bc} + g_b = \pi_b = h_a + f_{cb} + h_b$$
$$f_{cb} + g_d + g_c = \pi_c = f_{bc} + h_d + h_c$$
$$h_d + g_d = \pi_d = g_d + h_d \tag{65}$$

Thus, we have two equations and two unknowns

$$f_{bc} = f_{cb} - g_a - g_b + h_a + h_b$$
$$f_{bc} = f_{cb} - h_c - h_d + g_d + g_c \quad (66)$$

This gives us very interesting constraint: the sum of entry and exit passengers on the left is exactly equal to the sum of entry and exit passengers on the right of the link $b \leftrightarrow c$. Actually, this middle link constraint happens in any middle stations. Now we will use Condition 1 that there is no flow turning back (such as from b->c->b again) in any station (or we shall assume maximum of turning back flow in any station) to solve the transit flow estimation problem for any number of station. Without this assumption, the number of link flow can have range from zero up to infinity. Based on Condition 1, we have the maximum flow in the middle link would be

$$f_{bc} = h_c + h_d$$
$$f_{cb} = h_a + h_b \quad (67)$$

The middle link flow may be lower than that the above value (up to zero) but any excess of that maximum value would mean there is turning back flow. Now the final flow matrix becomes

$$\mathbf{F} = \begin{matrix} \nearrow \\ a \\ b \\ c \\ d \\ z \\ sum \end{matrix} \begin{matrix} a & b & c & d & z & sum \\ \left[\begin{matrix} 0 & h_a & 0 & 0 & g_a \\ g_a & 0 & h_c + h_d & 0 & g_b \\ 0 & h_a + h_b & 0 & g_d & g_c \\ 0 & 0 & h_d & 0 & g_d \\ h_a & h_b & h_c & h_d & 0 \end{matrix}\right] & \begin{matrix} \pi_a \\ \pi_b \\ \pi_c \\ \pi_d \\ \pi_z \end{matrix} \\ \pi_a \quad \pi_b \quad \pi_c \quad \pi_d \quad \pi_z & \kappa \end{matrix} \quad (68)$$

It should be noted that the flow matrix is not the same as an origin-destination (OD) matrix. To access the OD flow, we must get the flow (using k-path premagic matrix, which we shall discuss in another paper) from $z \to O \to path \to D \to z$. As numerical example of four stations problem, we are given the entry flow and exit flow of each station. Then the link flow can be computed as the equations above.

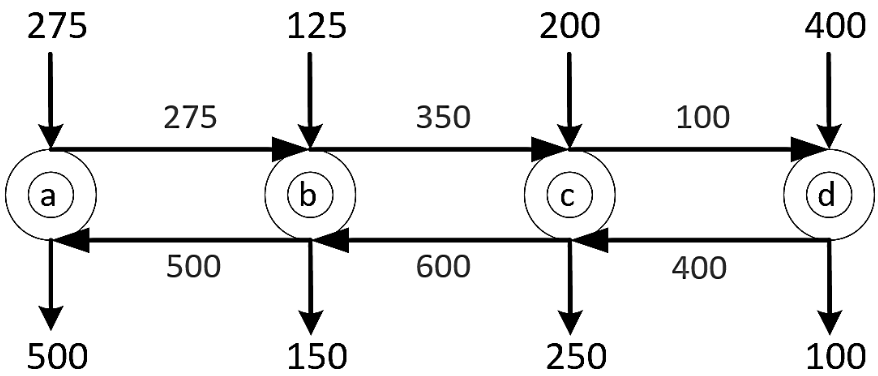
Note that

$$f_{bc} = f_{cb} - g_a - g_b + h_a + h_b = f_{cb} - 275 - 125 + 500 + 150 = f_{cb} + 250$$
$$f_{bc} = f_{cb} - h_c - h_d + g_d + g_c = f_{cb} - 250 - 100 + 400 + 200 = f_{cb} + 250 \quad (69)$$

Observe the middle link constraint

$$\Delta_{cb} = -g_a - g_b + h_a + h_b = -h_c - h_d + g_d + g_c \quad (70)$$

$$f_{bc} = f_{cb} - \Delta_{cb} \quad (71)$$



If $f_{bc} > h_c + h_d$ it means there is an excess flow inside f_{bc} that will turn back as part of f_{cb}. However, Condition 1 states that turning back flow is not allowed and therefore $f_{bc}^{\Delta} = h_c + h_d$ is the maximum flow in link bc without turning back flow. With similar argument, we know that if $f_{cb} > h_a + h_b$ there will be an excess flow inside f_{cb} that will turn back as part of f_{bc}. Since we do not allow these excess back flows, then $f_{cb}^{\Delta} = h_a + h_b$ is the maximum flow in link cb.

The middle link constraint gives us

$$f_{bc} = f_{cb} - \Delta_{cb} = f_{cb} - 250 \quad (72)$$

We set $f_{bc} = f_{bc}^{\nabla} = 350$, it would give us $f_{cb} = f_{cb}^{\nabla} = 600$.