CIRRELT-2021-06

Public Transit Transfer Analysis from Smart Card Data

Benjamin Disson
Martin Trépanier
Catherine Morency

February 2021

Bureau de Montréal
Université de Montréal
C.P. 6 128, succ. Centre-Ville
Montréal (Québec) H3C 3J7
Tél : 1-514-343-7575
Télécopie : 1-514-343-7121

Bureau de Québec
Université Laval,
2325, rue de la Terrasse
Pavillon Palasis-Prince, local 2415
Québec (Québec) G1V 0A6
Tél : 1-418-656-2073
Télécopie : 1-418-656-2624

Public Transit Transfer Analysis from Smart Card Data

Benjamin Disson[1,2,*] , Martin Trépanier[1,2], Catherine Morency[1,3]

[1] Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT)

[2] Department of Mathematics and Industrial Enginnering, Polytechnique Montréal

[3] Department of civil, geological and mining engineering, Polytechnique Montréal

Abstract. Transit organizations want to improve their attractiveness by reducing the impact of transfers on users. To do so, it is vital to have better knowledge about passengers' behavior at transfer points as well as network utilization. Using smart card automated fare collection (SCAFC) systems is a way to get data from an entire network over a long period of time. However, because smart card data are made up of transactions, they endogenously lack transfer information. In this research, the objectives are to study both temporal transfer distributions and passengers' behavior with respect to transfers. To do so, an algorithm to transform a transactions database into a trips database is presented. It links a series of transactions together and estimates the transfers' count and location. Transfers are detected by analysing transactions sequences and by determining the alighting station of each metro transaction. The algorithm is then used over one-year of data from the Société de transport de Montréal (STM); it covers the 2016 and 10,965 OPUS cards. Results show that 97.15% of the 8,505,189 transaction records can be grouped into trips for a total of 4,696,729 trip records. The variation in the proportion of trips by transfer type and time of day is then studied, and a k-means segmentation is applied on the cards. Results suggest that passengers favor a metro transfer for constrained trips and direct trips otherwise. This study also shows six different passengers' patterns and demonstrates that users' behavior varies during the year.

Keywords: Public transit, smart card data, transfer, trip destination, passenger behavior.

Acknowledgements. The authors thank the editor and four anonymous referees whose comments have helped improve the paper. While working on the project, the first author was postdoctoral fellow at the School of Management Université du Québec à Montréal, while the second author was Adjunct Professor. We gratefully acknowledge the financial support provided by the Natural Sciences and Engineering Council of Canada (NSERC), through its Discovery and Acceleration Grant programs. We also gratefully acknowledge the support of Fonds de recherche du Québec through their infrastructure grants, and their new university researchers start up program.

---

* Corresponding author: benjamin.disson@polymtl.ca

Dépôt légal – Bibliothèque et Archives nationales du Québec
Bibliothèque et Archives Canada, 2021

INTRODUCTION

With the extension of transit networks, passengers often have to transfer to reach their destination. However, research demonstrates that transferring is an obstacle in the use of transit [1]. Reducing the transfer impact is an unavoidable objective for public transportation agencies if they want to improve their attractiveness.

Numerous surveys have been conducted to enumerate the attributes considered by a passenger when choosing his itinerary and to evaluate the influence on the transfer penalty. Another way to get information about transfers is to use smart card data. However, in some tap-in systems with no tap-out validation, there is no information about the alighting location and time. This lack of information creates an additional challenge for the analysis.

In this research, we present a methodology based on smart card data to estimate the count and the location of each transfer type in a transit network over a long period of time. The objective is to obtain several characteristics on passengers' behavior and network use regarding the type and the time of a transfer. This will bring a better understanding on the way people choose their route, including or not a transfer. To do so, we have created an algorithm to transform a transactions database into a trips database and we then applied a k-means method to cluster passengers based on their patterns. We also looked at the distributions of tranfer types during a week.

The paper is structured as follows. The background section exposes the potentialities of smart card data and presents the existing studies on transfers. Then, the methodology section describes the data collection and explains the developed algorithm. The results section presents the experiment based on one-year of data from the *Société de transport de Montréal* (STM) and shows observations about transfer variability and passengers' behavior. Finally, a conclusion is made to expose the prospects of the algorithm and the contributions of this research.

BACKGROUND

Smart Card Data Use for Public Transit Planning

Smart card is a mid-seventies technology that is now more and more used by public transportation agencies to improve transaction times and to offer a better pricing system. It makes multimodal transportation possible because it can be used in several modes in a city and sometimes by various agencies [2]. When a user enters a vehicle, he validates his card on a reader. If the card contains a valid fare, the system records the transaction with data from the card (personal and purchase information) and data from the vehicle (mode, line, direction, time, location…) [3]. Smart cards thus provide planners with a large amount of data, with continuous, temporal and spatial information.

Numerous studies now use smart card automated fare collection (SCAFC) data to better understand passengers' behavior and network use [3]. As an example, SCAFC can be used to classify users with data mining methods [4]; to identify the purpose of a trip [2]; to evaluate the use variability of the public transit lines [5]; to assess transit reliability [6]; to model the impact of weather on public transportation user behavior [7]; and others.

Connection Analysis

Passengers often need to transfer when they use transit. It means that they have to change from a vehicle to another (sometime between two different modes) which involves walking and waiting [8]. However, it has been demonstrated that transferring is an obstacle in the use of public transportation services [1]. To reduce the transfer penalty, many researches tried to understand the influence of each attribute taken into consideration by a passenger when choosing an itinerary [9]. Several data types are useful to model trip transfers.

Survey Data

Surveys are used to obtain information about the attributes that influence the passengers' choices. First, Iseki and Taylor [9] have shown that for transfers, three sectors of improvement exist: fares, operational aspects and physical attributes. Guo and Wilson [10] wanted to determine where and how to invest money and showed that cost comes from transfer walking distance, transfer waiting time and transfer environment. Some studies have listed transfer waiting and walking time, information availability, alternative frequencies and reliability, safety, comfort, transfer station furniture and design as the most important attributes [1], [11]–[14]. Another survey objective is to measure the value of time, because it will modify the passenger's willingness to make a transfer in his transit trip. Wardman [15] demonstrated that the value of time varies depending on the user and the mode. Later, Iseki and Taylor [16] and Ceder and al. [17] have shown that the "out-of-vehicle" time is more burdening than the "in-vehicle" time and that users often choose the less risky alternative. Moreover, a transfer is privileged provided it saves 33% of travel time or 16% of travel cost [17]. According to De Keizer and al. [11], the ideal transfer time is 4 minutes. Finally, Guo [18] have shown that a transfer is equivalent to a 10 minutes walk.

Smart Card Data

Survey data are limited by their low users' coverage, their small spatial coverage and their short time period. SCAFC data can be a solution as they allow studying an entire network for a long period of time [19].

Smart cart data are often used to model network traffic at a transfer node. Nishiuchi and al. [20] used the Data Envelop Analysis approach to measure the transfer stations' efficiency. They showed that the performance depends on the time of the day, the location, the passengers crowd and the available alternatives. To go further, Sun and Schonfeld [21] took an interest in the 'fail-to-board' problem. When too many passengers are using the same station to transfer from a vehicle to another, there is a risk that some of them miss their first available connection and must wait for the next one. They used this model to identify network bottlenecks.

In the same way, SCAFC data can be used to model passengers' behavior during the transfer. In 2015, Sun and al. [22] tried to evaluate the impact of several attributes on a transfer walking time, in a tap-in tap-out network, thanks to a linear regression model. They concluded that passengers are faster in the morning peak due to work hours. They showed that crowding has an impact only if the station's capacity is reached. In the same way, they explained that passengers adapt their walking speed depending on the available information. In smart card analysis, passengers are often considered homogeneous. To solve this problem, Li and al. [23] used smart card data to segment users into several groups, depending on the selected routes. They showed that the passengers' choices depend on timetables and time of the day. In the morning peak, users do not always take the fastest itinerary but rather choose the most reliable. Conversely, in the evening peak, passengers are less sensitive to trip time, they choose longer itineraries to avoid transfers and to have better comfort.

Trip Destination

In most networks with SCAFC system, users only have to validate their card when boarding a bus or entering a metro station. It is called a tap-in system. Consequently, there is no information about the time and the alighting locations. Trépanier and al. [24] proposed a four-step algorithm to estimate the alighting time and stop from a tap-in record. It is assumed that a user's alighting stop is the nearest one to the next boarding stop recorded. In the same way, the last validation of a day is compared with the first validation of this day. For unlinked validation in a day (when the alighting station can not be deduced by the algorithm), He and Trépanier [25] used regular trips analysis. This part is essential to model the distribution of transfers in a network using smart card data.

Literature on smart card use and transfer analysis shows a lack of studies about transfer distribution over a year and across a network. Most studies about transit transfers are focused on one short period of time and often only on some stations. The aim of this paper is to propose a methodology based on smart card data to obtain knowledge about transfers (whatever the time of the year) over an entire network.

METHODOLOGY

After the characterization of the dataset used for this study, this section presents the model to locate and count the transfers between two points and the algorithm to transform a transactions database into a trips database.

Data Collection

The research is conducted in collaboration with the *Société de transport de Montréal* (STM), which is the principal transit authority in Montreal, operating both buses and subway. In 2016, STM operated 236 bus lines with 4,370 stops and 4 metro lines with 68 stations [26]. STM uses the OPUS card, introduced in 2008, as SCAFC system. This card allows access to both buses and metro (subway) and the system records all transactions with operational information.

The study leans on transactions from smart cards. Transactions were selected from the thousand most frequent cards in addition to 9,965 randomly chosen cards with at least 50 records over the year for a total of 10,965 cards. Over 366 days (from January 1 to December 31, 2016), this sample gathers 8,505,189 transactions.

The STM network is a tap-in system: users do not need to validate their card when alighting from a vehicle or a station. As a result, the only relevant information in a transaction record is the transaction number; day and hour; card number; card type and mode. If it is a bus, there is the line and the direction. If it is a metro, there is the metro station. In all the cases, there is a statement code. The statement code is used to link two transactions together. If a transaction code is 'D1' then it means this transaction has no link with the previous one. If a transaction code is 'DC' then it means this transaction belongs to the same trip as the previous one.

The statement code is defined by the STM's business rules. It specifies that a transaction code is 'D1' if it is the first card transaction in a day; if the previous transaction was more than 120 minutes earlier; if it is the same line than the previous transaction; or if both the transactions are in a subway station. In all other cases, the transaction code is 'DC' [27].

Connections' Counts and Location Estimation

Two modes are considered in the research, bus and metro. Consequently, four types of transfer exist: bus to bus (BB), bus to metro (BM), metro to bus (MB) and metro to metro (MM). The objectives are to detect, locate and count all the transfers using the sequence of transactions made by a card during a day.

A BB transfer is detected directly, because a user validates his card in each bus. This corresponds to two bus transactions, the second having the code 'DC'. In this research, the location of a BB connection is not investigated. BM and MB transfers are easily detected too, but the location must be deduced. A BM transfer is located in the metro station, because the station location is known. But an MB transfer is not located because the alighting metro station is unknown. It is considered that the MB transfer location is in the metro station which is the nearest from the bus line of the next transaction, as found by a Geographic Information System (GIS) program.

For each metro transaction, it is possible that there exist some MM transfers. To count and locate these transfers, the alighting station has to be deduced. This is made by an adaptation of the algorithm created by He and al. [25]. The pair of boarding and alighting stations is then compared with a 'metro transfer dataset' that indicates the number and the location of transfers that must be done to connect each pair of stations. In Montreal, there is a maximum of two transfers to connect any two metro stations.

Figure 1 presents a series of four transactions from a card (transactions A, B, C and D) as an example. These four transactions illustrate all the possible connections' types. The transaction successions A – B, B – C and C – D are used to find the connections.

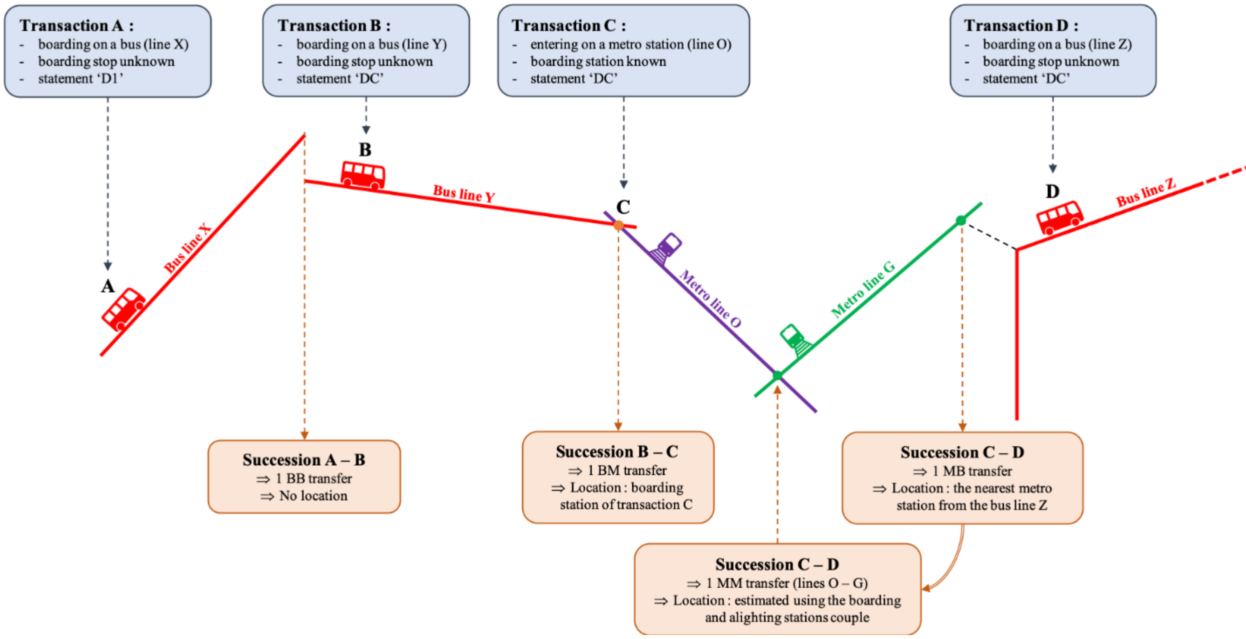

Figure 1. Transfer estimation methodology

Transactions A and B are in a bus and transaction B has a 'DC' statement. In this way, there is a BB transfer with no location. Transaction C has a 'DC' statement too and is in metro. So, there is a BM transfer and because the entry station is known, it is the transfer location. Transaction D has 'DC' too and is in a bus, so there is a MB transfer. This transfer location is assumed to be the metro station which is the nearest from the bus line of transaction D (line Z). Last, because the metro alighting station is not on the same line as the boarding station, there is a MM transfer. To locate this transfer, the 'metro transfer dataset' is used. Because the transfer detection must be done for each transaction, this part is integrated with the transactions-to-trip algorithm, as illustrated by the fourth stage in Figure 2.

Transactions-to-Trip Algorithm

SCAFC data provides a large database filled with transactions. But because of the transfers, a transaction cannot always be considered as a trip: each time there is a transfer, the trip is composed of several transactions. The objective of the proposed algorithm is to transform the transactions database into a trips database. To do so, there are six necessary stages, shown schematically in Figure 2 and described in the following paragraphs.

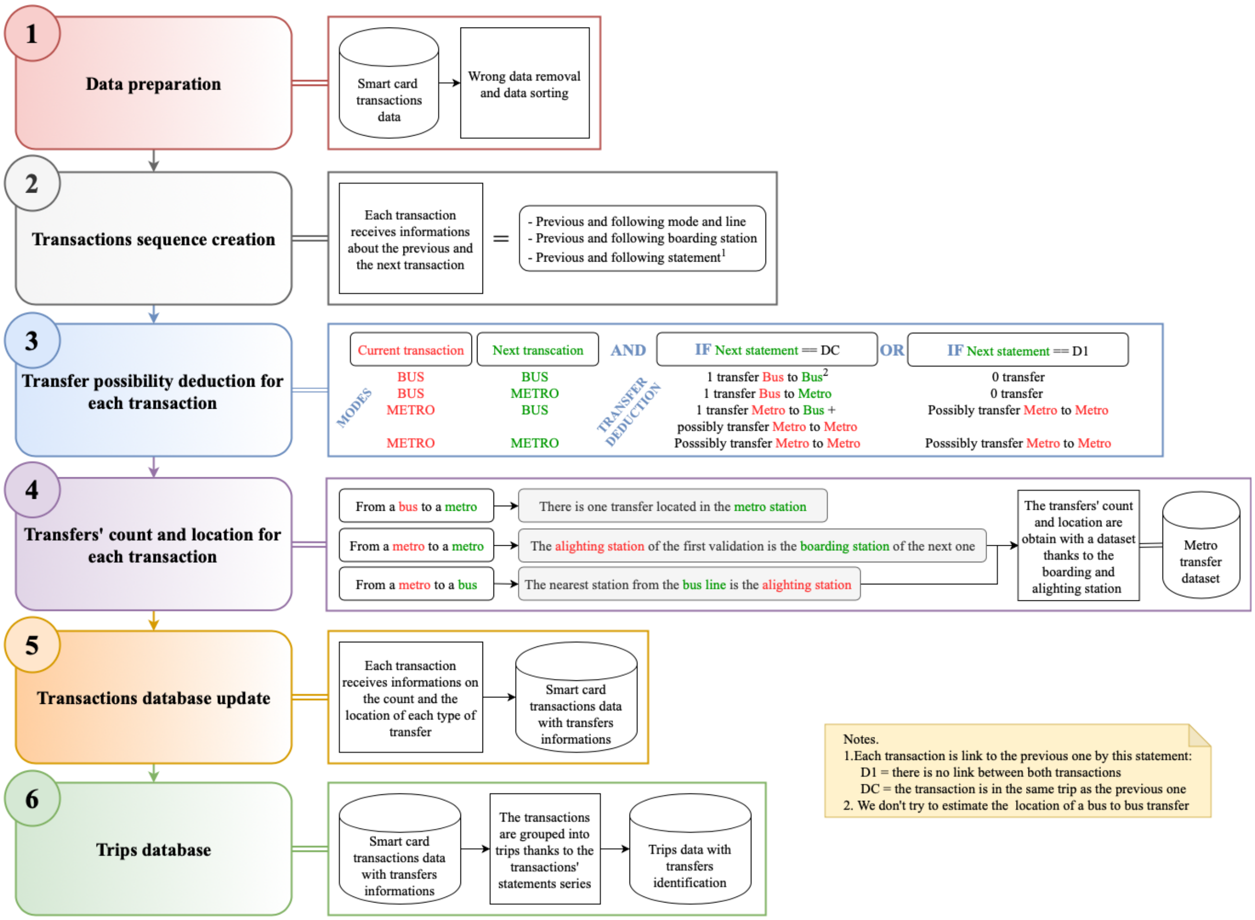

Figure 2. Transactions-to-trip algorithm

1. Data preparation

Not all data can be processed. Before any operation, records with a missing or erroneous data must be removed. Typical cases are:

- One or several data is absent (no identification number, etc.);
- A card is validated on a line that does not exist (systematic error);
- The mode and the line are discordant (ex. metro mode with a bus line number).

During the database cleaning, cards records with only one metro transaction in a day (0.75% of the data) are removed because the alighting station cannot be found. After removing all the problematic records, data is sorted by days, cards and hours.

2. Transactions sequence creation

To apply the method proposed by He and al. [25], it is necessary to have, for each transaction, information about the previous and the following ones from the same card, the same day. To do so, the database is modified, and a transaction row receives the mode, the line or the station and the statement code from the previous and the following transactions. This methodology is inspired by the model of Hofmann and O'Mahony [28]. The first record of the day for a card receives a 0 in each 'previous' columns because there is no previous transaction. For the last transaction of the day, to apply the method of He and al. [25], the program adds in each 'following' column information about the first card transaction of the day.

3. Transfer possibility deduction

The algorithm processes a transaction accordingly to the possible transfer types. For the purpose of time resolution optimization, the program cannot try all the possibilities for each transaction. So, a first reading of the database is made to compare each transaction with the following one. During this reading, for each transaction, the algorithm only uses the current mode, the mode of the following transaction and the current statement code. It is then possible to bring out the potential transfer types. The program always considers the possibility of one MM connection when the transaction mode is metro.

4. Transfers' count and location

Transactions presenting similar sets of possible transfer types are then grouped. As presented in the previous section, all the BB transfers are validated, and the BM transfers receive the metro boarding station as transfer location. The proximity between metro and bus is used to locate the MB transfer and then the model of He and al. [25] is applied to estimate the MM transfer locations.

5. Transactions database update

Each transaction now has the transfers' count and location. At the end of this process, the original database is enriched with 8 columns:

- BusBus: 1 if there is a BB transfer; 0 otherwise.
- BusMetro: 1 if there is a BM transfer; 0 otherwise.
- MetroBus: 1 if there is a MB transfer; 0 otherwise.
- MetroMetro: the number of MM transfers; $0 \leq$ MetroMetro ≤ 2.
- BM_loc: the station code of the BM transfer location; 0 if BusMetro = 0.
- MB_loc: the station code of the MB transfer location; 0 if MetroBus = 0.
- MM_loc_1: the station code of the first MM transfer location; 0 if MetroMetro = 0.
- MM_loc_2: the station code of the second MM transfer location; 0 if MetroMetro ≤ 1.

6. Trips database creation

The transaction database is now enriched with the transfers information. The statement code is used to transform the transactions database into a trips database. All the transactions that are linked to the same trip receive the same 'transit ID' that will differ from a trip to another. In this way, all the same trip transactions can be grouped in one record that corresponds to the trip. This transit record is composed of: the transit ID, the number of transactions grouped, the card ID, the date, the time of the first transaction, the chain of modes, the boarding stations chain and the alighting stations chain in the metro, the total transfers' count, the connections' count by type and the location of each transfer. Because the transit ID is retained in both the transactions database and the trips database, all the transactions of a trip can easily be found.

RESULTS

General Characteristics

The transactions-to-trip algorithm is applied on 10,965 OPUS cards from the STM during the 366 days of 2016. This database contains 8,505,189 transactions. After wrong data removal, the algorithm can be used on 97.15% of the original database, being 8,262,842 transactions.

With the transformation of each transactions series into a trip, a database with 4,696,729 transit trip records is finally obtained. There is an average of 1.19 trips per card per day. The trips database shows an average of 1.09 transfers per trip. Moreover, 67.80% of all trips have at least one transfer.

Temporal Analysis

Figure 3 presents the proportion of trips with transfers during the day, according to the transfer type. To do this, trips are grouped by hours and are counted by transfer type. Because the hour of a MM transfer is unknow, the trip hour is defined by the hour of the first transaction of the trip. A trip with several transfer types will be counted once in each category. Conversely, a trip with no transfer (direct trip) is counted in the 'no transfer' category.

Note. A trip with several transfer types is counted several times, therefore the sum in the proportion of trips is more than one (on average 132%).

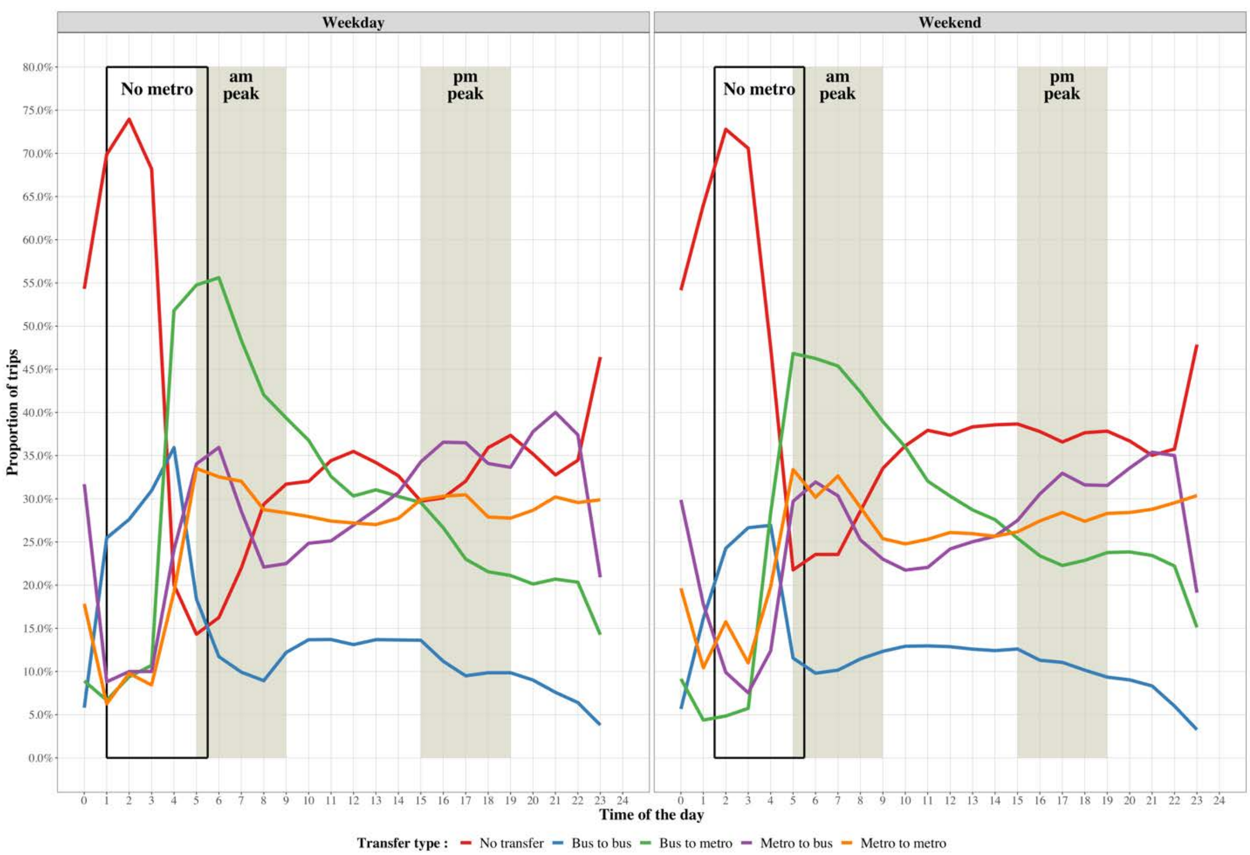

Figure 3. Proportion of trips according to the time of the day, the type of day and the transfer type

The transfer distributions are dependent on the transfer types. During the night, most of the transit trips are direct (without any transfer) and some others are with a BB transfer. This is due to longer bus intervals and metro closing time. Indeed, in Montreal, there is no metro circulation between 1:00 a.m. and 5:30 a.m. on weekdays and between 1:30 a.m. and 5:30 a.m. on Saturday night.

The morning peak (from 5 a.m. to 9 a.m. [29]) is characterized by an increase of the share of trips including a metro transfer. Conversely, the proportion of direct trips and trips with BB transfer is decreasing. This may be because during this time, half of transit trips are for work purposes (according to [30]) and are consequently subject to time constraints. After the morning peak, proportion of trips with a BB transfer stays low. This may be due to the fact that buses are slower than metro and provide less capacity [31].

On weekdays, the share of direct trips and trips with a metro transfer (MM, MB and BM) is higher than the share of trips with a BB transfer. However, proportions of trips with a BM transfer are decreasing all day long. One possible explanation is the bus frequency reduction during the day [31]. Because passengers have to wait more for a bus, they favor the metro. The afternoon peak (from 3 p.m. to 7 p.m. [29]) is also visible even if the variations are lower than in the morning peak. It is noticed by a decrease of direct trips and trips with a BB transfer, and an increase of trips with MB and MM transfers. The lower variations can be explained by a longer afternoon peak and because passengers are less constraint on schedule.

On weekends, the distribution trend is the same as on weekdays but with fewer variations and a less pronounced morning peak. Moreover, the morning peak appears to be later, and the afternoon peak is less noticeable. After the morning peak, there are more direct trips.

In conclusion, Figure 3 highlights that constrained trips (trips that are constrained by time, often in peak periods) increase the proportion of trips with a metro transfer while trips outside the peak period (probably trips with less obligation or no time constraint) increase the proportion of direct trips. The low proportion of BB transfers can indicate that passengers prefer connecting with metro than a bus, which is consistent with a study by Iseki and Taylor [9] and Garcia-Martinez and al. [32].

User Behavior During the Year

The previous section shows that the transfer distribution depends on the time, the day and the type. But users have an important role to play, too. Are they all adopting the same behaviors? Are passenger behaviors constant during the year?

To answer these questions, a card dataset is created with this information: month, card number, trips count for this card this month, portion of trips with no transfer, portion of trips with BB, BM, MB and MM transfers. Segmentation is then made on this dataset to cluster typical behaviors together.

The segmentation is made with the k-means method, that is often used in transportation for its efficiency and its capability to create groups with no particular form. Only the proportions of trips are needed, so the Euclidian distance can be used with no need to normalize the data. A first k-means is made with 25 groups to obtain centers. Then a dendrogram applied to these centers shows that the data can be separated into 5 meaningful clusters. Because sometimes there is no transaction on a card during a month, some cards are not included in the segmentation. Another cluster is thus added manually (cluster 0), composed of these cards. Table 1 presents the centers of the six clusters.

Table 1. K-means cluster centers

Cluster	0	1	2	3	4	5
Cards-months [%]	26,66	20,74	12,52	6,66	17,75	15,66
Direct trips [%]	0,00	38,80	72,36	27,32	15,47	12,65
Bus to bus [%]	0,00	6,88	5,68	47,77	7,81	5,09
Bus to metro [%]	0,00	25,90	9,51	20,40	49,45	36,92
Metro to bus [%]	0,00	24,13	9,45	18,92	48,16	38,45
Metro to metro [%]	0,00	25,70	9,49	11,10	14,00	70,87

Looking at Table 1, it is visible that the six clusters express various passengers' behavior:
Cluster 0: People who have a card but don't use public transit in a month. Most of all the observations (26.66%) are in this cluster.
Cluster 1: Passengers that make direct trips (38.80% of their trips) but often have recourse to trips with transfers. Most of the trips' transfers are between a bus and a metro (in the two ways), or between two metro.
Cluster 2: Passengers that make direct trips as often as possible (72.36%). They avoid trips with transfers and when it is not possible, they use a metro connection.
Cluster 3: Passengers that often do transfers (72.68%) namely between two buses. They do less connections with metro. This cluster contains less observations than the others (6.66%).
Cluster 4: Passengers that make a transfer in most of their trips (84.53%). In this cluster, passengers do not do much BB and MM transfers. Their trips mostly include both bus and metro segments.
Cluster 5: Passengers that almost make a transfer at each trip (87.35%). However, they rarely have BB transfers. They often use a BM or MB transfer but, in the majority, they have a MM transfer.

Table 2. Clusters analysis with temporal attributes

Cluster	0	1	2	3	4	5
Trips in peak [%]	//	49,63	52,52	50,84	58,92	58,64
Trips out of peak [%]	//	50,37	47,48	49,16	41,08	41,36
Trips in weekday [%]	//	80,84	82,54	80,83	84,28	84,84
Trips in weekend [%]	//	19,16	17,46	19,17	15,72	15,16

In addition to the transfer types attributes, Table 2 expresses that the daily use of public transit varies according to the clusters. There is no information for cluster 0 because it is composed of cards without transit. Clusters 1 and 3 seem to have the same trends as well as clusters 4 and 5. Cluster 2 however is a central group, with trends that approach clusters 1 and 3. The share of trips made during the peak period is higher for clusters 4 and 5. Trips made by clusters 1 are mostly out of peak. The bigger share of trips in weekdays for clusters 4 and 5 may be due to work trips. These observations confirm results from the previous section: the peak period leads people to use a metro type transfer and more direct trips are out of peak.

To express the changes in behavior during the year, Figure 4 presents the proportion of cards in each cluster depending on the month. The clusters' distribution varies during the year, which means that passengers' behavior with respect to transfer composition of their selected transit routes, is not constant. Furthermore, the variations of the proportions are not identical. The proportion is always higher in cluster 0, and the variations are larger. This cluster has a higher share in January, July and after November. Clusters 1, 2 and 3 show the same trends. Proportions in these clusters are lower in January, increase until the middle of the year (depending on the cluster) and finally decrease until December. Conversely, clusters 4 and 5 have a different pattern. Proportions increase until February and decrease to a minimum in July. Their shares increase again until October and decrease afterwards.

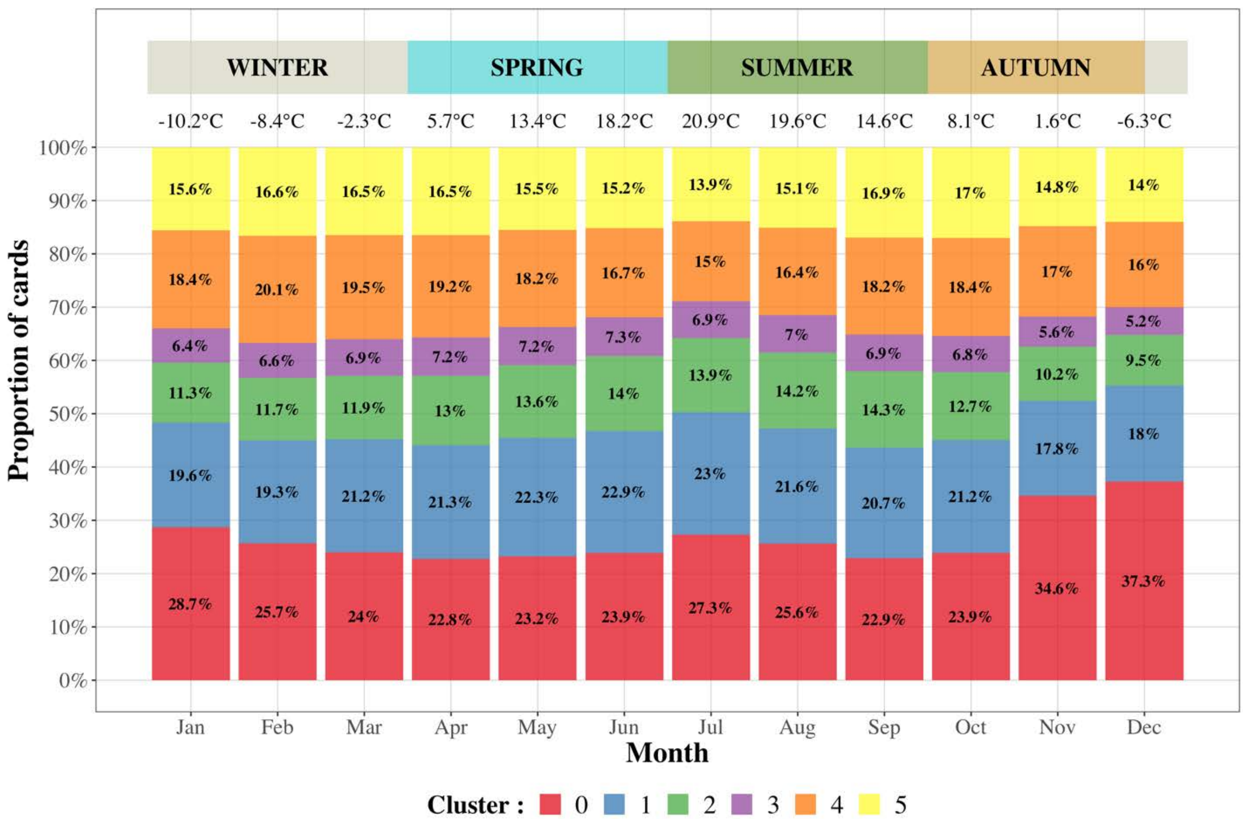

Figure 4. Proportion of cards in each cluster per month

These variations during the year bring the learning that passengers' transit transfers patterns are dependent on the month. First of all, the proportions' distribution in cluster 0 can be explained by the winter and summer holidays. People leave Montreal and therefore don't use their card. Temperature can explain the peak in winter too. Fewer people travel when it is cold. Clusters 1, 2 and 3 refer to passengers who are often using a bus, with or without transfer. The lower proportion in winter can then be explained by the meteorological conditions (as expresses by the average temperature for each month on Figure 4 [33]). Because of the snow and the temperature, they are less willing to wait or walk outside to take a bus [7]. Weather can also explain the higher proportion of cards in clusters 4 and 5 during winter. Passengers who don't want to wait for a bus take metro and use transfers with metro. In summer, because the weather is less difficult, some people choose to walk and wait to take buses and use direct trips. So, patterns between clusters are reversed. Further research including weather data will be used to learn more about the changes in behavior due to this factor.

Another way to evaluate the behavioral changes is to count the number of clusters where a card is found during the year. If a passenger's card is only in one cluster, it means that his transit transfers pattern doesn't change during the year. Inversely, if a passenger's card is in several clusters, it means that there are some changes in behavior during the entire year. Table 3 presents the share of cards that have different numbers of clusters throughout the year.

Table 3. Proportion of cards belonging to several clusters

Number of clusters	1	2	3	4	5	6
Cards portion [%]	8.13	28.37	36.39	21.28	5.28	0.56

From Table 3 it is visible that there are some passengers that are in only one cluster during the year. Looking into the database, it appears that most of these passengers are in clusters 4 and 5 with respectively 35% and 34% of the cards. So, the more regular passengers are people who often have recourse to transfers and most of the time with a metro. In third position there is cluster 2 with 14%. It means that there are passengers who make direct trips and don't change route throughout the year. This regularity can be due to passenger preference or the absence of any other route choice.

Most of the cards are in two or three clusters (65%). That means that the majority of passengers change their behavior during the year but not always. On the contrary, only 27% of the users do not have a regular behavior (more than 4 clusters in a year).

Knowing the passengers' transit transfers patterns is useful for planners because it brings some learnings about how passengers are using the network according to the period of the year. This information helps to adapt the offer according to passengers' behavior. A possible improvement could be the implementation of a pricing system consistent with the habits (for example, for passengers of cluster 3, create a subscription that make only bus access possible).

CONCLUSION

This study has proposed a methodology to count and locate the transit transfers in a network using data from a SCACF system. This method leans on the transformation of a transactions database into a trips database by linking series of transactions together and determining the alighting stop of each boarding. The results show that the algorithm can estimate the number of transfers for 97.15% of the transactions in the case study. The analysis of the share of trips with each transfer type expresses a link between the time of the day and the use of transfers. These changes can be explained by both the users' preferences and the way the network is organized. Passengers use metro transfer for constrained trips and direct or BB transfers trips otherwise. To go further, a k-means method has been applied to obtain groups of typical behavior. Six categories of passengers' transfers composition were found. Some users always have recourse to a transfer, and the type varies within the clusters. Other users maximize the number of direct trips. Finally, regarding the clusters' composition for each month, we have demonstrated that, for the majority of them, passengers' transfers composition is changing over a year. Possible explanations for this change in behavior are the weather that makes more difficult the use of buses in winter and the holidays. Another explanation is that maybe passengers use other routes to go somewhere else. Looking at the corridors use distribution will make verifications possible. Knowing the reason for the trip can also be useful for further research.

This research brings some information on the users' patterns during an entire year. Also, the presented algorithm is based on a normalized smart card database. Consequently, it can be applied to any other network equipped with an SCAFC system. This will allow future comparison with other cities in the world.

However, this study is limited by the absence of location of the bus boarding stops. This creates a lack into the original database that prevents the estimation of the bus to bus transfer location and leads to a bias on the MB and MM transfers' location. We are not currently able to verify whether or not the location is correct. In further research, will try to validate these results with the Origin-Destination survey. Also, because transfers' type distribution and passengers' behavior seem to be linked with exogenous factors, we will add meteorological information into the database, take street furniture into account as well as transit service frequency.

ACKNOWLEDGMENTS

The authors wish to thank the *Société de transport de Montréal* (STM) for providing the data and the *Mitacs Accélération* organization with *Keolis* and the *Société Nationale des Chemins de fer Français* (SNCF) for providing funding.

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design: B. Disson, M. Trépanier, C. Morency; data collection: M. Trépanier, C. Morency; analysis and interpretation of results: B. Disson, M. Trépanier, C. Morency; draft manuscript preparation: B. Disson. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

[1] A. Ceder and S. J. Chowdhury, "Interchange attributes influencing public-transport user's intention to use routes with transfers," *Transp. Res. Board*, 2012.

[2] F. Devillaine, M. Munizaga, and M. Trépanier, "Detection of activities of public transport users by analyzing smart card data," *Transp. Res. Rec.*, vol. 2276, no. 1, pp. 48–55, 2012.

[3] M.-P. Pelletier, M. Trépanier, and C. Morency, "Smart card data use in public transit: A literature review," *Transp. Res. Part C Emerg. Technol.*, vol. 19, no. 4, pp. 557–568, 2011.

[4] C. Morency, M. Trépanier, and B. Agard, "Measuring transit use variability with smart-card data," *Transp. Policy*, vol. 14, no. 3, pp. 193–203, 2007.

[5] A. Dubos-Golain, M. Trépanier, and C. Morency, "Understanding transit use patterns in Montreal," CIRRELT, 2017.

[6] M. Trépanier, K. M. Habib, and C. Morency, "Are transit users loyal? Revelations from a hazard model based on smart card data," *Can. J. Civ. Eng.*, vol. 39, no. 6, pp. 610–618, 2012.

[7] M. Trépanier, C. Morency, B. Agard, E. Descoimps, and J. Marcotte, "Using smart card data to assess the impacts of weather on public transport user behavior," 2012.

[8] V. R. Vuchic, *Urban transit: operations, planning, and economics*. John Wiley & Sons, 2005.

[9] H. Iseki and B. D. Taylor, "Not all transfers are created equal: Towards a framework relating transfer connectivity to travel behaviour," *Transp. Rev.*, vol. 29, no. 6, pp. 777–800, 2009.

[10] Z. Guo and N. H. M. Wilson, "Assessing the cost of transfer inconvenience in public transport systems: A case study of the London Underground," *Transp. Res. Part Policy Pract.*, vol. 45, no. 2, pp. 91–104, Feb. 2011.

[11] B. De Keizer, K. T. Geurs, and G. H. Haarsman, “Interchanges in timetable design of railways: A closer look at customer resistance to interchange between trains,” *Proc. Eur. Transp. Conf. Glasg.*, pp. 1–13, 2012.
[12] S. Hernandez and A. Monzon, “Key factors for defining an efficient urban transport interchange: Users’ perceptions,” *Cities*, vol. 50, pp. 158–167, 2016.
[13] D. Lois, A. Monzón, and S. Hernández, “Analysis of satisfaction factors at urban transport interchanges: Measuring travellers’ attitudes to information, security and waiting,” *Transp. Policy*, vol. 67, pp. 49–56, 2018.
[14] C. A. Cheriyan and S. Sinha, “Assessment of Transit Transfer Experience: Case of Bangalore,” *Transp. Res. Procedia*, 2019.
[15] M. Wardman, “Public transport values of time,” *Transp. Policy*, vol. 11, no. 4, pp. 363–377, 2004.
[16] H. Iseki and B. D. Taylor, “Style versus service? An analysis of user perceptions of transit stops and stations,” *J. Public Transp.*, vol. 13, no. 3, p. 2, 2010.
[17] S. Chowdhury, A. A. Ceder, and B. Schwalger, “The effects of travel time and cost savings on commuters’ decision to travel on public transport routes involving transfers,” *J. Transp. Geogr.*, vol. 43, pp. 151–159, 2015.
[18] Z. Guo, “Assessment of the transfer penalty to transit trips in Downtown Boston: a GIS-based disaggregate modeling approach,” PhD Thesis, Massachusetts Institute of Technology, 2003.
[19] A. Grapperon, B. Farooq, and M. Trépanier, “Information fusion of smart card data with travel survey,” CIRRELT, 2016.
[20] H. Nishiuchi, T. Todoroki, and Y. Kishi, “A fundamental study on evaluation of public transport transfer nodes by data envelop analysis approach using smart card data,” *Transp. Res. Procedia*, vol. 6, pp. 391–401, 2015.
[21] Y. Sun and P. M. Schonfeld, “Schedule-based rail transit path-choice estimation using automatic fare collection data,” *J. Transp. Eng.*, vol. 142, no. 1, 2015.
[22] L. Sun, J. G. Jin, D.-H. Lee, and K. W. Axhausen, “Characterizing Multimodal Transfer Time Using Smart Card Data: the Effect of Time, Passenger Age, Crowdedness, and Collective Pressure,” *Transp. Res. Board*, 2015.
[23] W. Li, Q. Luo, Q. Cai, and X. Zhang, “Using Smart Card Data Trimmed by Train Schedule to Analyze Metro Passenger Route Choice with Synchronous Clustering,” *J. Adv. Transp.*, vol. 2018, 2018.
[24] M. Trépanier, N. Tranchant, and R. Chapleau, “Individual trip destination estimation in a transit smart card automated fare collection system,” *J. Intell. Transp. Syst.*, vol. 11, no. 1, pp. 1–14, 2007.
[25] L. He and M. Trépanier, “Estimating the destination of unlinked trips in transit smart card fare data,” *Transp. Res. Rec.*, vol. 2535, no. 1, pp. 97–104, 2015.
[26] Société de transport de Montréal, *La mobilité des Montréalais*. 2016.
[27] Société de transport de Montréal, “Règlement R-105 concernant les conditions au regard de la possession et de l’utilisation de tout titre de transport pour les services de transport offerts par ou pour la Société de transport de Montréal,” *Les règlements et politiques*, 01-Sep-2017. [Online]. Available: http://www.stm.info/fr/a-propos/gouvernance-d-entreprise/les-reglements-et-politiques/reglement-r-105-artm. [Accessed: 26-Jul-2019].
[28] M. Hofmann and M. O’Mahony, “Transfer journey identification and analyses from electronic fare collection data,” in *Proceedings. 2005 IEEE Intelligent Transportation Systems, 2005.*, 2005, pp. 34–39.

[29] Agence Métropolitaine de Transport, *L'enquête Origine-Destination 2013, mobilité des personnes dans la région de Montréal*. 2013.
[30] Agence Métropolitaine de Transport, *La mobilité des personnes dans la région de Montréal, Faits saillants*. 2013.
[31] V. R. Vuchic, "Transit systems operations and networks," in *Urban transit: operations, planning, and economics*, Wiley, Ed. John Wiley & Sons, 2005, pp. 3–295.
[32] A. Garcia-Martinez, R. Cascajo, S. R. Jara-Diaz, S. Chowdhury, and A. Monzon, "Transfer penalties in multimodal public transport networks," *Transp. Res. Part Policy Pract.*, vol. 114, pp. 52–66, 2018.
[33] Ville de Montréal, "Climat, Principales données climatiques." [Online]. Available: http://ville.montreal.qc.ca/portal/page?_pageid=6897%2C67875636&_dad=portal&_schema=PORTAL. [Accessed: 18-Jul-2019].