

Research Article

Identification Dockless Bike-Sharing and Metro Transfer Travelers through Mobility Chain

Xiang Li,¹ Qipeng Yan,¹ Yixiong Tang,¹ and Chen Luo ⁰

¹School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 611756, China ²College of Air Traffic Management, Civil Aviation Flight University of China, Deyang 618307, China

Correspondence should be addressed to Chen Luo; luochenkun@cafuc.edu.cn

Received 7 November 2023; Revised 30 January 2024; Accepted 7 March 2024; Published 16 March 2024

Academic Editor: Ziyuan Gu

Copyright © 2024 Xiang Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The burgeoning dockless bike-sharing system presents a promising solution to the first- and last-mile transportation challenge by connecting trip origins/destinations to metro stations. However, the differentiation between metro passengers and DBS riders, as they belong to distinct systems, hinders the precise identification of DBS-metro transfers. This study introduces an innovative method employing mobility chains to establish spatiotemporal relationships, including spatiotemporal conflicts and similarities, among potential users from both systems. This significantly enhances the precision of user matching. An empirical study in Chengdu validates the method's increased accuracy and examines travel patterns, yielding the following insights: (1) Introduction of the mobility chain reduces average matched pairs by 28.27% and improves accuracy by 18.36%. The addition of spatial-temporal similarity further boosts accuracy by 19.32%. (2) Median distances for DBS-metro access and egress transfers are approximately 950 meters. Short trips of 650–750 meters are prevalent, while trips exceeding 1.5 kilometers lead passengers to opt for alternative modes. (3) Temporal patterns reveal weekday peaks at 8:00, 9:00, and 17:00. On weekends, transfers are uniformly distributed, mainly within urban areas. Suburban stations exhibit reduced weekend activity. These findings can provide valuable insights for enhancing DBS bicycle redistribution, promoting transportation mode integration, and fostering urban transportation's sustainable development.

1. Introduction

Dockless bike-sharing (DBS) is a promising solution for improving first-mile and last-mile connectivity to public transportation (PT). By eliminating the need for docking bays, DBS allows bicycles to be parked more conveniently near transit hubs and endpoints, giving users more flexibility and making it easier to integrate with public transit [1, 2]. Consequently, DBS has achieved worldwide prominence owing to its inherent conveniences, typified by notable implementations like Mobike and Halo in China, lime in the USA, Moov in Singapore, and Eazymov in Europe.

However, prevailing studies on integrated DBS and metro utilization typically employ comparable methodologies, whereby rigid geographical boundaries centered on metro stations delineate integrated trips. Expressly, this approach assumes bike-sharing journeys initiating or terminating within the predefined metro area buffer exhibit bike-metro integration, irrespective of actual intermodal transfer [1, 3, 4]. Nevertheless, this imprecise identification approach may misclassify some bike-sharing trips unrelated to metro usage. For example, when metro stations are situated in densely populated residential areas, or adjacent to high DBS using regions like university campuses, relying solely on a fixed geographical buffer cannot effectively distinguish integrated DBS-metro usage from general DBS cycling behavior.

In addition, given that the DBS system and public transportation are operated by distinct entities, they represent typical heterogeneous systems, underscoring the complexity of transfer identification between the two modes. This inherent heterogeneity distinguishes the DBS scenario from the relatively straightforward transfer between buses and metro systems within an urban PT framework. While some studies have investigated shared bicycle-to-metro transfers using docked bike-sharing systems and a unified smart card system shared with public transportation [5, 6], these primarily focus on docked platforms, overlooking the dynamic nature of dockless bike-sharing systems. Hence, the immediate challenge is to ascertain how to employ data from the PT system and DBS to accurately identify DBS-metro transfer travelers with a refined spatial-temporal granularity.

This study establishes an innovative data identification technique that accurately identifies real metro-DBS transfer flows between heterogeneous systems of metro and DBS by constructing distinct mobility chains for public transportation and DBS through multi-source data. The method involves extracting DBS trips associated with metro transfers using buffer zones around metro station entry and exit points. Comprehensive mobility chains for DBS and public transportation are built by integrating bus trip data from a unified smart card system with Traffic Analysis Zones (TAZ) information. By analyzing the spatiotemporal dynamics between these chains, including conflicts and similarities, precise correspondences are established, as exemplified in the context of Chengdu. This research provides essential insights for developing eco-friendly transportation planning and strategic initiatives.

The remainder of this paper is structured as follows: Section 2 reviews relevant literature; Section 3 describes the study areas and data sources; Section 4 details the methodology; Section 5 presents and discusses the results; and Section 6 concludes the study, along with the limitations of the current research and prospects for future investigations.

2. Literature Review

A substantial body of literature examines the integration of DBS and metro transit, spanning diverse topics that include travel behavior of multimodal DBS-metro trips [7], accessibility analyses between DBS and metro networks [8], bike parking capacity surrounding metro stations [9], and forecasting methodologies for DBS-metro transfer demand [10]. This literature review concentrates on integrating DBS with metro transit systems. Specifically, it examines three key areas: (1) identification of potential DBS-metro transfer trips, (2) usage patterns of potential DBS-metro transfer trips, and (3) DBS Mobility Chain and PT Mobility Chain.

2.1. Identification of Potential DBS-Metro Transfer Trips. Previous investigations have employed diverse methodologies to comprehend the interplay between DBS services and transit systems. Several surveys have concentrated on the integration of DBS with the metro network, aiming to discern travelers' preferences in selecting their transportation mode [11, 12]. However, the utilization of survey data is limited by its constrained sample size and restricted spatial coverage, preventing a comprehensive and precise analysis of the geospatial aspects of integrated usage [13]. Another approach to identifying multimodal trips involves matching bike sharing and metro farecard data [14]. However, this methodology only applies to docked systems since DBS does not utilize smartcards.

Fortunately, the Global Positioning System capabilities of DBS allow operators to locate and track bicycles in real time, generating substantial geospatial data. Researchers can leverage these granular location records to pinpoint DBSmetro transfers through spatial analyses of the relative positioning between DBS and metro stations [15]. Several investigations have harnessed historical data from DBS systems, alongside the spatial allocation of dockless shared bikes proximate to metro stations, as a method for identifying DBS-metro transfer journeys. Guo et al. delved into the temporal and spatial utilization tendencies of DBS services in proximity to metro stations by analyzing DBS trip records in conjunction with metro station data. The criterion they employed was the inclusion of the origin or destination location of a dockless shared bike within a buffer zone of 150 meters encompassing each metro station. In such instances, the anticipation is that a transfer trip between DBS and the metro has transpired [15]. Lin et al. examined the service areas of dockless bikes linking to metro stations utilizing dockless trip data. They identified DBS-metro transfers by tracing dockless journeys originating or terminating within 50 meters of a metro entrance [16]. Liu et al. conducted a spatiotemporal analysis contrasting DBS and ridesharing as first-mile/last-mile links to metro systems. Their findings revealed that sociodemographic and built environment variables influenced usage patterns between the two connector modes. Moreover, distinct threshold values were established to delineate transfer journeys for interconnecting metro networks, with a threshold of 50 meters for DBS and 100 meters for ride-sharing, respectively [17]. Yu et al. analyzed DBS accessibility from metro stations by extracting trips originating within 100 meters of any metro entrance. They then assessed the reachability of cycling destinations surrounding metro station areas based on this trip data [18]. Liu et al. determined that a transfer span of 150 meters represents an acceptable threshold between DBS and metro connectivity. Furthermore, their study delved into the spatiotemporal attributes underlying DBS's role as a feeder mode to the metro network [15]. Zhou et al. established a transfer threshold of 300 meters as a viable criterion for achieving effective connectivity between dockless DBS and metro systems. Furthermore, their study encompassed an analysis of the interrelation between the transit system and DBS connectivity [19].

2.2. Usage Patterns of Potential DBS-Metro Transfer Trips. A multitude of studies have been undertaken to probe the analysis of integration behavior between metro systems and DBS services. Employing a nested logit model, Ji et al. ascertained that female travelers, elderly individuals, and low-income commuters exhibited a diminished propensity to embrace bike-sharing as a mode integrated with the metro network. Furthermore, their investigation unveiled that commuters who had experienced bike theft were more prone to embrace bikeshare options integrated within the metro

network [20]. Various individual-level attributes, including socioeconomic factors (such as age, gender, income, and residential location), personal attitude factors (including environmental awareness), and perceptual factors (such as perceived comfort and traffic safety), have been linked to the utilization frequency of DBS services [21]. Kim et al. analyzed metro-bikeshare usage patterns across four dimensions: transfer time, date, location, and access/egress modes. Their findings revealed variations in metro-bikeshare travel behaviors between distinct user groups [22].

2.3. DBS Mobility Chain and Public Transportation Mobility Chain. It is essential to emphasize that the establishment of a Dockless Bike-Sharing Mobility Chain (DBSMC), which links a user's consecutive biking trips in chronological order, represents a valuable method for uncovering additional insights into user activities when analyzing DBS data [23]. Built upon the foundation of the DBSMC, a comprehensive array of analyses can be undertaken encompassing usage, user dynamics, and integrated evaluations. For example, with a focus on usage patterns, Bordagaray et al. formulated the DBSMC framework to probe bike-sharing system utilization and unveil the actual demand for bike-sharing services. This approach offers substantial utility to policymakers and system operators in tasks spanning demand analysis, service evolution, and optimization strategies [24]. Transitioning to integrated analysis, certain inquiries concentrate on methodological fusion. For instance, Builes-Jaramillo et al. synthesized methods pertinent to the DBSMC with spatiotemporal network analysis, thereby unearthing specific usage trends, such as the reduced propensity of women to engage with the DBS system. Concurrently, alternative studies have focused on harmonizing data within the context of DBSMC. This involves the linkage of transfer trip chains belonging to individual travelers, achieved by correlating transit card data with bike-sharing card data. This integrative approach facilitates the analysis of metro-bike interchange journeys [5, 25]. Regarding urban multimodal travelers, previous research has utilized Public Transportation Mobility Chains (PTMC) to analyze their perception of transfers and intentions related to multimodal trip chains. This includes their considerations regarding transfer frequency, waiting duration, and walking distances [26, 27].

2.4. Research Gap. Most of the mentioned studies characterize the relationship between metro passenger flow and DBS ridership either by analyzing survey data or by evaluating the spatial correlation between DBS O/D points and metro stations. Studies utilizing survey data fail to accurately capture the large-scale metro-DBS transfer passenger flow and are unable to analyze the spatial-temporal relationship between metro and DBS transfers. Similarly, when evaluating the spatial correlation between DBS bike O/D points and metro stations, due to the heterogeneity of the DBS and metro systems, setting any transfer distance threshold between DBS O/D points and metro stations is insufficient to directly substantiate the existence of transfer passenger flow between the metro and DBS systems. Consequently, both approaches may lack robust evidence for accurately identifying metro-DBS transfers. Furthermore, the credibility of the travel patterns of transfer passenger flows between DBS and the metro, as determined by these methods, is also questionable.

3. Study Area and Data Resources

3.1. Study Area. The city of Chengdu, located in China's Sichuan province, constitutes a vital economic center in the Southwest region, encompassing a total area of 14,335 square kilometers, with 949.6 square kilometers comprising developed urban terrain. As of December 2020, official records enumerated Chengdu's population as 20,937,700 residents. Chengdu possesses a comprehensive metro system constituted by 7 numbered lines (1, 2, 3, 4, 5, 7, and 10), possessing a total track length of 518 km and serving 193 stations. In Chengdu, the metro signifies a prevalently utilized transportation mode, with the metro network accommodating an average ridership of 3.75 million daily passengers, highlighting the metro's integral status within the city's transportation infrastructure [28]. With Chengdu's rapid metro expansion, a burgeoning DBS industry materialized and swiftly progressed. In September 2016, Chengdu pioneered its first DBS program, Mobike. By the end of 2020, approximately 985,000 DBS bicycles were in operation with a registered membership of 8.35 million users.

3.2. Data Resources. The data used in this study are as follows:

- DBS trip data. The study sourced a dataset comprising 39.84 million DBS trip records from the Chengdu Transportation Operations Coordination Center. These records, spanning from December 1 to December 16, 2020, are comprehensively detailed Table 1. Data outside the study area was excluded, and incomplete or irregular records were cleansed. Coordinates were standardized, and trips filtered based on a 1 to 60-minute duration criterion.
- (2) *Smart card data.* The study analyzes smart card data from December 1 to December 16, 2020, covering both metro and bus passengers. While metro data provides details on boarding and alighting stations and times, initial bus data only included boarding specifics. Integrating alighting information from bus vehicle trajectory data, as outlined in prior research [28], the consolidated dataset encompasses 58.68 million records, as detailed in Table 2.
- (3) *Public transportation station data.* The PT stations considered in this study consist of metro stations and bus stops. The metro station data was obtained from the Chengdu Rail Transit Group, comprising station coordinates, operating schedules, and entry/exit coordinates. The bus stop data was acquired from the Chengdu Public Transport Group, including bus stop coordinates.

		TTARY	ryambre or records to	T TUD UTP Mara			
DBS trip ID	DBS rider ID	Start time	Start longitude	Start latitude	End time	End longitude	End latitude
11544*****769139511	2864*****691	2020/12/1 19:17:24	103.84933	30.69252	2020/12/1 19:24:47	103.85245	30.70165
11544 * * * * 606788359	2824*****036	2020/12/1 10:05:35	104.00605	30.65451	2020/12/1 10:09:56	104.00273	30.65581
11544 * * * * 606829133	2990*****623	2020/12/1 21:28:12	104.22180	30.57550	2020/12/1 21:33:41	104.22291	30.58507

TABLE 1: Example of records for DBS trip data.

TABLE 2: Example of records for smart card trip data.

Smart card trip ID	Smart card ID	Start station/stop	Start time	End station/stop	End time	Mode
01004****0682147	61001****1340	31602	2020/12/1 07:38:25	30152	2020/12/1 07:51:12	Bus
00002*****6067883	50305****3201	1132	2020/12/1 19:45:33	1149	2020/12/1 20:23:22	Metro
00006****6068291	61000****9700	1158	2020/12/1 13:11:37	1199	2020/12/1 13:34:11	Metro

(4) GIS layers. GIS layers were provided by the Chengdu Municipal Bureau of Planning and Natural Resources, including administrative boundaries, road networks, and TAZ data.

4. Methodology

4.1. Methodology Framework. The study introduces a method for precise identification of DBS-metro transfer travelers through mobility chain analysis and spatialtemporal similarity metrics, also examining their travel patterns. The process begins with screening DBS trips linked to metro transfers via a buffer around metro station entry/ exit points. Using TAZ data, the DBSMC and PTMC are constructed. Initial matched pairs are determined by identifying potential access and egress transfer trips with varying thresholds. Spatial-temporal conflict filtering and similarity metrics are then applied to refine these pairs, addressing challenges like many-to-one matches. The methodology and validation of identification results, along with the analysis of DBS-Metro transfer travelers' travel patterns, are illustrated in Figure 1.

4.2. Obtaining the Initial Matched Pairs

4.2.1. Extracting DBS-Metro Transfer-Related Trips. Previous research has typically used fixed radial buffers of 100-300 meters around metro stations to identify DBSmetro transfers, but this method struggles with stations of diverse sizes [3, 9, 15, 17]. As depicted in Figure 2(a), complex stations served by multiple lines cannot be accurately represented by a single coordinate, and some entry/ exit points may lie beyond 400 meters from the main coordinates. The application of a fixed radial distance centered on a metro station to encompass all station entrances and exits within the buffer would lead to the misclassification of numerous DBS trip data that do not involve DBS-metro transfers. As illustrated in Figure 2(b), this study proposes establishing buffers based on the actual entry/exit coordinates of each station, rather than using a generic buffer for metro stations. This approach ensures accurate extraction of DBS-metro transfer trips tailored to each station's unique layout.

For a DBS trip to qualify as a DBS-metro transfer, its origin or destination must be proximate to a metro station location. By evaluating DBS trip start or end positions relative to metro station coordinates, two transfer types can be delineated: access transfer (arriving at metro stations) and egress transfer (departing metro stations). These are identified based on whether the trip start or end point falls within the entry/exit buffer zone of a given metro station, determined as shown in equations (1) and (2):

$$C_{i}(M_{n}^{k}) = \begin{cases} \text{Access,} & l_{i}^{e} \in \text{buffer}_{150}(M_{n}^{k}), \\ \text{Egress,} & l_{i}^{s} \in \text{buffer}_{150}(M_{n}^{k}), \\ \text{No,} & \text{else,} \end{cases}$$
(1)

where $C_i(M_n^k)$ represents the type of transfer for the DBS trip *i* at the *k* entry/exit points of the metro station *n*. l_i^e represents the end location of DBS trip *i*. l_i^s represents the start location of DBS trip *i*. buffer₁₅₀ (M_n^k) represents the 150-meter buffer zone at the *k* entry/exit point of metro station *n*.

$$\begin{cases} \Omega_1^{\text{Access}}(M_n) = \left\{ i \mid i \in \Omega_{\text{DBS}}, C_i(M_n^k) = \text{Access} \right\}, \\ \Omega_1^{\text{Egress}}(M_n) = \left\{ i \mid i \in \Omega_{\text{DBS}}, C_i(M_n^k) = \text{Egress} \right\}, \end{cases}$$
(2)

where Ω_{DBS} represents the set of all DBS trips. $\Omega_1^{\text{Access}}(M_n)$ represents the set of DBS trip endpoints at the entry/exit of metro station n. $\Omega_1^{\text{Egress}}(M_n)$ represents the set of DBS trip start points at the entry/exit of metro station n.

Given the proximity of metro station entrances and exits, their respective buffer zones can overlap, resulting in a trip being associated with multiple metro entry/exit points. However, it is important to clarify that, within the context of $\Omega_1^{Access}(M_n)$ and $\Omega_1^{Egress}(M_n)$, which are associated with distinct metro stations, a trip is registered only once at a specific metro station and does not have any bearing on the subsequent processes.

Due to the fixed daily operating schedule of each metro station, and the variance in operational times among stations, the initiation or conclusion time of the DBS trip needs to align with the operational timeframe of the transferring metro station. The categorization of this situation can be determined based on the type of DBS trip transfer, as follows:

$$\begin{cases} A_i^{\text{Access}}(M_n) = (t_i^e - T_n^s)(T_n^e - t_i^e), & i \in \Omega_1^{\text{Access}}(M_n), \\ A_i^{\text{Egress}}(M_n) = (t_i^s - T_n^s)(T_n^e - t_i^s), & i \in \Omega_1^{\text{Egress}}(M_n), \end{cases}$$
(3)

where t_i^s and t_i^e represent the start and end times of DBS trip *i*, while T_n^s and T_n^e signify the commencement and conclusion of the operational time of metro station *n*, respectively.

Utilizing equations (1)-(3), this study identifies DBS trips related to access and egress transfers at metro station *n*, as detailed in the following equation:



FIGURE 1: Frameworks to identify DBS-metro transfer travelers.

$$\begin{cases} \Omega_2^{\operatorname{Access}}(M_n) = \left\{ i \mid i \in \Omega_1^{\operatorname{Access}}(M_n), \ A_i^{\operatorname{Access}}(M_n) > 0 \right\}, \\ \Omega_2^{\operatorname{Egress}}(M_n) = \left\{ i \mid i \in \Omega_1^{\operatorname{Egress}}(M_n), \ A_i^{\operatorname{Egress}}(M_n) > 0 \right\}, \\ \Omega_2^{\operatorname{All}} = \bigcup_{n=1}^N \left(\Omega_2^{\operatorname{Access}}(M_n) \cup \Omega_2^{\operatorname{Egress}}(M_n) \right), \end{cases}$$
(4)

where $\Omega_2^{\text{Access}}(M_n)$ and $\Omega_2^{\text{Egress}}(M_n)$ represent the sets of DBS trips in the vicinity of metro station *n* pertaining to access (egress) transfers, respectively. Ω_2^{All} represents the set of DBS trips associated with DBS-metro transfers.

4.2.2. Construct DBS and PT Mobility Chain for Potential DBS-Metro Transfer Travelers. Mobility chains contain multiple subtrips and can reflect the hidden detailed travel information [29]. Meanwhile, mobility chains also furnish

detailed temporal and spatial data, facilitating more precise analyses [23]. Thus, adopting a user-centric approach to construct mobility chains enriches spatial-temporal data for various users, thus enhancing user identification accuracy in both PT and DBS systems. The DBSMC is constructed using OD data from DBS trips, while the PTMC is developed from passenger IDs in the PT system's smart card data. Notably, due to the absence of drop-off stop data for bus trips within the PT system, it was imperative to integrate bus vehicle trajectory data to compensate for this deficiency, a procedure that has been previously accomplished in prior work by the authors [28].

Initially, DBS riders with trips associated with DBSmetro transfers are identified, and their complete set of DBS trips is extracted in chronological order. The compliant DBS rider p_i is represented with all DBS trips as $\Omega_{\text{DBS}}^{p_i}$. For



FIGURE 2: Method of extracting DBS-metro transfer-related trips. (a) Metro station with distant entrances and exits. (b) Extracting metro-DBS transfer-related trips.

metro passengers, all metro trips are compiled chronologically, supplemented by corresponding bus trips within the PT system, denoted for metro passenger p_j as $\Omega_{\text{PT}}^{P_j}$. For a more detailed description, refer to the following equation:

$$\begin{cases} \Psi_{p_{i}} = \left\{ p_{i}^{x_{1}}, p_{i}^{x_{2}}, \cdots, p_{i}^{x_{|X|}} \right\}, & i \in \Omega_{2}^{\text{All}}, x \in X \subseteq \Omega_{\text{DBS}}^{p_{i}}, \\ \Psi_{p_{j}} = \left\{ p_{j}^{y_{1}}, p_{j}^{y_{2}}, \cdots, p_{j}^{y_{|Y|}} \right\}, & j \in \Omega_{\text{metro}}, y \in Y \subseteq \Omega_{\text{PT}}^{p_{j}}, \end{cases}$$
(5)

where p_i represents a DBS rider involved in the DBS-metro transfer trips denoted. The sequence Ψ_{p_i} comprises all DBS trips made by the rider p_i , arranged chronologically. The total count of DBS trips in the sequence Ψ_{p_i} is designated as |X|. Likewise, p_j identifies a metro passenger. Sequence Ψ_{p_j} encompasses all PT trips taken by metro passenger p_j , sorted in chronological order. The total count of PT trips in the sequence Ψ_{p_i} is represented as |Y|.

To facilitate the analysis, PT and DBS trip start/end locations were mapped to TAZ using ArcGIS. This mapping

involved intersecting metro stations and bus stops, and DBS trip OD points, with TAZ boundaries, assigning each location a zone number based on spatial position. This integration allowed examining relationships between PT and DBS trips via their shared TAZ. It also improved accuracy in matching trips to passengers across modes. Linking locations to TAZ enabled subsequent analyses of spatial patterns and DBS-metro transfers in travelers' mobility.

Subsequently, the DBSMC and PTMC were derived from the respective DBS and PT trip sequences. Equation (6) defines the mobility chain of a PT passenger, encompassing the initial or final location at a metro station or bus stop, the corresponding time, and the associated TAZ. Similarly, equation (7) characterizes a DBS rider's mobility chain, incorporating start and end coordinates, timing, and TAZ information.

$$\operatorname{Chain}_{p_{j}}^{\mathrm{PT}} = \left\{ \left[\left(T_{p_{j}^{y_{1}}}^{s}, l_{p_{j}^{y_{1}}}^{s}, z_{p_{j}^{y_{1}}}^{s} \right), \left(T_{p_{j}^{y_{1}}}^{e}, l_{p_{j}^{y_{1}}}^{e}, z_{p_{j}^{y_{1}}}^{e^{y_{1}}} \right) \right], \cdots, \left[\left(T_{p_{j}^{y_{|Y|}}}^{s}, l_{p_{j}^{y_{|Y|}}}^{s} \right), \left(T_{p_{j}^{y_{|Y|}}}^{e^{y_{|Y|}}}, l_{p_{j}^{y_{|Y|}}}^{e^{y_{|Y|}}} \right) \right] \right\}, \tag{6}$$

where $\operatorname{Chain}_{P_j}^{\operatorname{PT}}$ represents the PT mobility chain, encompassing all metro passenger p_j PT trips. $T_{p_j^{y_{|Y|}}}^s$ and $T_{p_j^{y_{|Y|}}}^e$ indicate the start and end times of p_j Y-th PT trip $y_{|Y|}$. $l_{p_j^{y_{|Y|}}}^s$ and $l_{p_j^{y_{|Y|}}}^e$ represent the coordinates of the boarding and alighting locations, either metro stations or bus stops, during the Y-th PT trip $y_{|Y|}$ of p_j . $z_{p_j^{y_{|Y|}}}^s$ and $z_{p_j^{y_{|Y|}}}^e$ correspond to the TAZ numbers of the metro stations or bus stops where boarding and alighting occurred during the Y-th PT trip $y_{|Y|}$ of p_j .

$$\text{Chain}_{P_{i}}^{\text{DBS}} = \left\{ \left[\left(t_{p_{i}^{x_{1}}}^{s}, l_{p_{i}^{x_{1}}}^{s}, z_{p_{i}^{x_{1}}}^{s} \right), \left(t_{p_{i}^{s}}^{e}, l_{p_{i}^{s}}^{e}, z_{p_{i}^{s}}^{e} \right) \right], \cdots, \left[\left(t_{p_{i}^{x_{|X|}}}^{s}, l_{p_{i}^{x_{|X|}}}^{s}, z_{p_{i}^{x_{|X|}}}^{e} \right), \left(t_{p_{i}^{x_{|X|}}}^{e}, l_{p_{i}^{x_{|X|}}}^{e}, z_{p_{i}^{x_{|X|}}}^{e} \right) \right] \right\}, \tag{7}$$

where Chain P_i^{DBS} represents the mobility chain encompassing all DBS trips taken by the DBS rider p_i . $t_{p_i^{x_{|X|}}}^s$ and $t_{p_i^{e_{x_{|X|}}}}^e$ represent the respective start and end times of the X-th DBS trip $x_{|X|}$ taken by p_i . $l_{p_i^{x_{|X|}}}^s$ and $l_{p_i^{e_{x_{|X|}}}}^e$ represent the coordinates at the start and end points of the X-th DBS trip $x_{|X|}$ by p_i . $z_{p_i^{x_{|X|}}}^s$ and $z_{p_i^{e_{x_{|X|}}}}^e$ represent the TAZ numbers associated with the start and end locations of the X-th DBS trip $x_{|X|}$ taken by p_i .

4.2.3. Identification of Potential Access and Egress Transfers. After identifying the DBS-metro transfer-related trips within the entry/exit buffer zone of the metro stations and establishing the DBSMC and PTMC, it is necessary to determine whether DBS rider p_i and metro passenger p_j are indeed transferring.

First and foremost, in the pursuit of identifying DBSmetro transfers, a foundational concept known as matched pair $\langle p_i, p_j \rangle$ is introduced. This concept delineates the plausible connections between users of two distinct modes of transportation: metro passengers and DBS riders.



FIGURE 3: Illustration for identifying access transfer and egress transfer.

Considering the variation in timing between access transfers arriving at the metro station and egress transfers departing from it [7, 30], specific transfer time thresholds for these two distinct scenarios were established, as illustrated in Figure 3.

For access transfer, when the trip of the DBS rider p_i is within the set of access transfer-related trips $\Omega_2^{p_i, \text{Access}}(M_n)$ at metro station n, all the metro passengers whose metro trip start time satisfies the access time threshold σ_1 are recorded to form a matched pair. Afterwards, the number of access transfer identifications for each matched pair at metro station n, formed by a DBS rider p_i with different metro passengers, is computed as outlined in the following equation:

$$\left\langle p_{i}, p_{j} \right\rangle_{\left(M_{n}\right)}^{\operatorname{Access}} = \left\{ \begin{array}{l} \sum_{X_{1}} \sum_{Y_{1}} \left| \left\{ \left\langle p_{i}, p_{j} \right\rangle \mid T_{p_{j}^{Y_{1}}}^{s} - t_{p_{i}^{X_{1}}}^{e} \leq \delta_{1}, X_{1} = \Omega_{2}^{p_{i},\operatorname{Access}}\left(M_{n}\right) \in \Omega_{2}^{\operatorname{Access}}\left(M_{n}\right), Y_{1} = \Omega_{\operatorname{Metro}}^{p_{j}}\left(M_{n}\right) \right\} \right|, \\ 0, \quad \text{else}, \end{array}$$

$$(8)$$

where $\langle p_i, p_j \rangle_{\langle M_n \rangle}^{\text{Access}}$ represents the count of times that matched pair $\langle p_i, p_j \rangle$ has been matched for access transfers at station *n*. X_1 represents the complete set of access transfer-related trips for DBS rider p_i at station *n*. Y_1 represents the complete set of trips for metro passenger p_j at metro station *n*. σ_1 represents the access time threshold, set to 900 seconds [14, 31].

For egress transfer, when the trip of the DBS rider p_i is within the set of egress transfer-related trips $\Omega_2^{p_i, \text{Egress}}(M_n)$ at metro station n, all the metro passengers whose metro trip end time satisfies the egress time threshold σ_2 are recorded to form a matched pair. Afterwards, the number of egress transfer identifications for each matched pair at metro station n, formed by a DBS rider p_i with different metro passengers, is computed as outlined in the following equation:

$$\left\langle p_{i}, p_{j} \right\rangle_{\left(M_{n}\right)}^{\text{Egress}} = \begin{cases} \sum_{X_{2}} \sum_{Y_{2}} \left| \left\{ \left\langle p_{i}, p_{j} \right\rangle \mid t_{p_{i}^{X_{2}}}^{s} - T_{p_{j}^{e}}^{e_{Y_{2}}} \leq \delta_{2}, X_{2} = \Omega_{2}^{p_{i}, \text{Egress}}\left(M_{n}\right) \in \Omega_{2}^{\text{Egress}}\left(M_{n}\right), Y_{2} = \Omega_{\text{Metro}}^{p_{j}}\left(M_{n}\right) \right\} \right|, \\ 0, \quad \text{else,} \end{cases}$$
(9)

where $\langle p_i, p_j \rangle_{(M_n)}^{\text{Egress}}$ represents the count of times that matched pair $\langle p_i, p_j \rangle$ has been matched for egress transfers at metro station *n*. X_2 represents the complete set of egress transfer-related trips for DBS rider p_i at metro station *n*. Y_2 represents the complete set of trips for metro passenger p_j at metro station *n*. σ_2 represents the egress time threshold, set to 600 seconds [14, 31].

Finally, the study calculate the number of access transfers and egress transfers at all metro stations for matched pair $\langle p_i, p_j \rangle$, determining the total count of DBS-metro transfer identifications for DBS rider p_i and metro passenger p_j , as shown in the following equation:

$$\operatorname{sum} \left\langle p_i, p_j \right\rangle = \sum_{N} \left(\left\langle p_i, p_j \right\rangle_{(M_n)}^{\operatorname{Access}} + \left\langle p_i, p_j \right\rangle_{(M_n)}^{\operatorname{Egress}} \right).$$
(10)

4.3. Filtering Invalid Matched Pairs. The preceding two sections have amassed a significant dataset comprising matched pairs of DBS riders and metro passengers. However, it is important to acknowledge the occurrence of "invalid matched pairs," where a single DBS rider may be linked to multiple metro passengers. These instances do not accurately represent genuine transfer trips undertaken by the same individual within both transportation systems. Therefore, two distinct types of conflicts are considered, and

a calculation of spatial-temporal similarity within the matched pairs is undertaken. This process aims to sift out invalid matched pairs and enhance the precision of identifying authentic DBS-metro transfer travelers.

4.3.1. Temporal Conflict. The riding duration of a DBS rider p_i , defined as $(t_{p_i}^s, t_{p_i}^e)$, and the travel duration of a metro

passenger p_j , defined as $(T_{p_j}^s, T_{p_j}^e)$, can be extracted from Chain_{P_i}^{DBS} and Chain_{P_j}^{PT}. When there is an overlap between $(t_{p_i}^s, t_{p_i}^e)$ and $(T_{p_j}^s, T_{p_j}^e)$, it is considered a temporal conflict. The concept of temporal conflict is shown as follows:

$$T_{\text{conflict}} \langle p_i, p_j \rangle = \begin{cases} 1, & \exists \left(t_{p_i^x}^s, t_{p_i^x}^e \right) \cap \left(T_{p_j^y}^s, T_{p_j^y}^s \right) \neq \emptyset, \forall x \in \Omega_{\text{DBS}}^{p_i}, \forall y \in \Omega_{\text{PT}}^{p_j}, \\ 0, & \text{else.} \end{cases}$$
(11)

4.3.2. Spatial Conflict. When p_i and p_j within a matched pair $\langle p_i, p_j \rangle$ represent the same individual, their itineraries across different transportation modes must be connectable. The matched pair can be deemed invalid if under free-flow (V_{free}) speed conditions there exist trips between transportation modes that cannot be connected. Combined with the urban transportation analysis report of the Gaode map to obtain the free flow speed V_{free} of 41.5 km/h in Chengdu [32].

As shown in Figure 4, there exist two matched pairs, $\langle p_{i'}, p_j \rangle$ and $\langle p_i, p_j \rangle$, where $p_{i'}$ first trip involves an access transfer with p_j second trip, and p_i first trip involves an egress transfer with p_j second trip. For access transfers, the speed between the endpoint of p_j first trip and the starting point of $p_{i'}$ first trip should be less than V_{free} .

Simultaneously, the speed between the endpoint of p_j second trip and the starting point of $p_{i'}$ second trip should also be less than V_{free} . However, upon examining the matched pair $\langle p_{i'}, p_j \rangle$, it is evident that p_j second trip reaches its endpoint while the starting point of $p_{i'}$ second trip remains in the vicinity of p_j second trip's starting point, and the speed between them exceeds V_{free} , resulting in a spatial conflict. Consequently, the matched pair $\langle p_{i'}, p_j \rangle$ is deemed invalid. Likewise, for egress transfers, the speed between the endpoint of p_i first trip and the starting point of p_j second trip should be less than V_{free} . At the same time, the speed between the endpoint of p_i second trip and the starting point of p_j third trip should also be less than V_{free} . The concept of spatial conflict is shown as follows:

$$S_{\text{conflict}}\langle p_{i}, p_{j} \rangle = \begin{cases} 1, \quad \exists T_{p_{j}^{Y_{T}}}^{s} - t_{p_{i}^{X_{t}}}^{e} \leq \delta_{1}, \frac{\operatorname{geodesic}\left(l_{p_{j}^{Y_{T-1}}}^{e}, l_{p_{i}^{X_{t}}}^{s}\right)}{\operatorname{time}\left(T_{p_{j}^{Y_{T-1}}}^{e}, t_{p_{i}^{X_{t}}}^{s}\right)} > V_{\text{free}} \operatorname{or} \frac{\operatorname{geodesic}\left(l_{p_{j}^{e}, r}^{e}, l_{p_{i}^{X_{t+1}}}^{s}\right)}{\operatorname{time}\left(T_{p_{j}^{Y_{T}}}^{e}, t_{p_{i}^{Y_{t+1}}}^{s}\right)}\right) > V_{\text{free}}, \\ 1, \quad \exists t_{p_{i}^{X_{t}}}^{s} - T_{p_{j}^{Y_{T}}}^{e} \leq \delta_{2}, \frac{\operatorname{geodesic}\left(l_{p_{i}^{x_{t-1}}}^{e}, l_{p_{j}^{Y_{T}}}^{s}\right)}{\operatorname{time}\left(t_{p_{i}^{x_{t-1}}}^{e}, T_{p_{j}^{Y_{T}}}^{s}\right)} > V_{\text{free}} \operatorname{or} \frac{\operatorname{geodesic}\left(l_{p_{i}^{e}, r}^{e}, l_{p_{j}^{Y_{T+1}}}^{s}\right)}{\operatorname{time}\left(t_{p_{i}^{x_{t-1}}}^{e}, T_{p_{j}^{Y_{T}}}^{s}\right)}\right) > V_{\text{free}}, \\ 0, \quad \text{else}, \end{cases}$$

where $T_{p_j}^{s_{Y_T}} - t_{p_i}^{e_{X_t}} \le \delta_1$ indicates when there is an access transfer between trips X_t and Y_T . $t_{p_i}^{s_{X_t}} - T_{p_j}^{e_{Y_T}} \le \delta_2$ indicates when there is an egress transfer between trips X_t and Y_T . The function geodesic () indicates the distance between two points, which can be obtained using the geodesic function in Python's geopy library [33].

4.3.3. Spatial-Temporal Similarity. Individuals exhibit substantial spatial-temporal consistency in travel dynamics and preferences, irrespective of their chosen mode of transportation. This consistency resembles patterns observed in historical areas, temporal, routes, or ODs [34]. Consequently, assessing the similarity of travel characteristics between DBS riders and metro passengers across diverse matched pairs can assist in identifying genuine matches.

To evaluate the spatial pattern similarity between a DBS rider p_i and a metro passenger p_j in a matched pair $\langle p_i, p_j \rangle$, this study define a spatial similarity metric $S_Sim\langle p_i, p_j \rangle$ based on textual similarity principles [35]. To calculate $S_Sim\langle p_i, p_j \rangle$, visit frequency vectors are constructed for p_i and p_j at their respective TAZs. The TAZs co-visited by both are then iteratively traversed, calculating the cumulative cosine similarity at each co-visited TAZ, as demonstrated in the following equation:



FIGURE 4: Illustration for spatial conflict.

$$\begin{cases} \operatorname{len}(Z_p) = \sqrt{\sum_{a_q \in Z_p} w_q^2}, p = p_i \text{ or } p = p_j, \\ \\ S_{-}\operatorname{Sim}\langle p_i, p_j \rangle = \sum_{a_q \in Z_{p_i} \cap Z_{p_j}} \frac{w_q^{p_i} w_q^{p_j}}{\operatorname{len}(Z_{p_i}) \operatorname{len}(Z_{p_j})}, \end{cases}$$
(13)

where Z_p represents the set of TAZ visited by either p_i or p_j , represented as Z_{p_i} or Z_{p_j} , respectively. $w_q^{p_i}$ or $w_q^{p_j}$ denotes the frequency a_q of the q-th TAZ visited by p_i or p_j , where $a_q \in Z_{p_i} \cap Z_{p_j}$. len (Z_p) is the frequency of overall weighted TAZ visits, indicated as len (Z_{p_i}) or len (Z_{p_j}) , for either p_i or p_j . As there must be an access transfer or egress transfer for both p_i and p_j , the TAZs visited by both are not empty, i.e., S_Sim $\langle p_i, p_j \rangle > 0$. When $Z_{p_i} = Z_{p_i}$, S_Sim $\langle p_i, p_j \rangle = 1$. Temporal pattern similarity is pivotal in individual

Temporal pattern similarity is pivotal in individual mobility patterns. The analysis of smart card and DBS trip data reveals mobility patterns of matched DBS riders and metro passengers across various time intervals. A normalized weight assignment is utilized to create a time distribution for each interval, where the weight indicates density. The temporal correlation depends on the similarity of these distributions. Earth Mover's Distance (EMD) [36] is employed to measure the dissimilarity between temporal patterns, capturing the minimal cost to transform one distribution into another. This metric effectively addresses transportation complexities by comparing entire distributions [37]. Thus, the study uses EMD to assess the temporal similarity of matched pairs.

The temporal similarity between DBS rider p_i and metro passenger p_j in matched pair $\langle p_i, p_j \rangle$ at a specific TAZ is determined by overlapping patterns in their respective time distributions, denoted as $h_{p_i} = \{t_{z_{p_i}}^{p_i}, v_{z_{p_i}}^{p_i}\} (z_{p_i} \in Z_{p_i})$ and $h_{p_j} = \{t_{z_{p_j}}^{p_j}, v_{z_{p_j}}^{p_j}\} (z_{p_j} \in Z_{p_j})$. Consider elements in h_{p_i} as "supplies" located at $t_{z_{p_i}}^{p_i}$ and elements in h_{p_j} as "demands" at $t_{z_{p_j}}^{p_j}$. Here, $v_{z_{p_i}}^{p_i}$ and $v_{z_{p_j}}^{p_j}$ represent the quantities of supply and demand, respectively. The Earth Mover's Distance (EMD) is defined as the minimum work needed to facilitate the transportation of supply to meet the demand. To quantify this similarity, a scoring function labelled as *T*_Sim $\langle H_{p_i}, H_{p_i} \rangle$, as detailed in the following equation:

$$\begin{cases} \sum_{Z_{p_{l}},Z_{p_{j}}} f_{z_{p_{l}},z_{p_{j}}} = \min\left\{\sum_{z_{p_{i}}} v_{z_{p_{i}}}^{p_{i}}, \sum_{z_{p_{j}}} v_{z_{p_{j}}}^{p_{j}}\right\}, f_{z_{p_{l}},z_{p_{j}}} \ge 0, \\ d_{z_{p_{l}},z_{p_{j}}} = \left|t_{z_{p_{l}}}^{p_{i}} - t_{z_{p_{j}}}^{p_{j}}\right|, \\ T_{\min\left\langle H_{p_{l}},H_{p_{j}}\right\rangle} = \exp\left(-\operatorname{argmin}_{f_{z_{p_{l}},z_{p_{j}}}} \sum_{Z_{p_{l}},Z_{p_{j}}} f_{z_{p_{l}},z_{p_{j}}} d_{z_{p_{l}},z_{p_{j}}}\right), \end{cases}$$
(14)

where $f_{z_{p_i},z_{p_j}}$ represents the set of flows that symbolize the required transport work. This study calculate the ground distance, denoted as $|t_{z_{p_i}}^{p_i} - t_{z_{p_j}}^{p_j}|$, between positions $t_{z_{p_i}}^{p_i}$ and $t_{z_{p_i}}^{p_j}$ using $|t_{z_{p_i}}^{p_i} - t_{z_{p_j}}^{p_j}|$. It is important to note that $v_{z_{p_i}}^{p_i}$ and $v_{z_{p_j}}^{p_j}$ values are normalized, resulting in $\sum_{Z_{p_i}, Z_{p_j}} f_{z_{p_i}, z_{p_j}}$ being equal to 1. When the two histograms are identical, the Earth Mover's Distance (EMD) between h_{p_i} and h_{p_j} equals 0, yielding a $T_{\text{Sim}}\langle H_{p_i}, H_{p_j}\rangle$ of 1. Conversely, as the EMD increases, the $T_{\text{Sim}}\langle H_{p_i}, H_{p_j}\rangle$ converges toward 0.

To quantify the similarity between DBS rider p_i and metro passenger p_j in the matched pair $\langle p_i, p_j \rangle$, accounting for both temporal and spatial patterns, a novel similarity function, denoted as ST_Sim $\langle p_i, p_j \rangle$, is introduced, defined as follows:

$$\mathrm{ST}_{\mathrm{Sim}\langle p_i, p_j \rangle} = S_{\mathrm{Sim}\langle p_i, p_j \rangle} \times T_{\mathrm{Sim}\langle H_{p_i}, H_{p_j} \rangle}.$$
 (15)

4.4. Determination of DBS-Metro Transfer Travelers from Matched Pairs. After obtaining matched pairs and the spatial-temporal relationships denoted as temporal conflict, spatial conflict, and spatial-temporal similarity between DBS riders and metro commuters, the subsequent steps involve eliminating invalid matched pairs and distinguishing the presence of many-to-one matched pairs. The precise workflow for this procedure is illustrated in Figure 5.

The determination process comprises two primary phases: the first phase is used by temporal conflict and spatial conflict to decrease the quantity of potential matched pairs, which is subsequently followed by spatial-temporal similarity to discern many-to-one matched pairs. This process strikes a balance between the count of identifications and identification accuracy.

5. Results and Discussion

This study employs the open-source unified analytics engine Apache Spark [38] for matched pairs identification. This system is built upon the Hadoop Distributed File System (HDFS), boasting a storage capacity of 68 TB, and comprises 22 nodes. Each node is outfitted with 32 cores and 32 GB of RAM. The entirety of the DBS-metro transfer identification process, spanning data cleansing, PT/DBS mobility chain construction, access transfer and egress transfer identification, as well as spatial-temporal conflict and spatial-temporal similarity calculations, determination of matched pairs is executed through a meticulously orchestrated sequence of spark jobs.

Utilizing the methodology outlined in Section 4, matched pairs were derived, ensuring each smart card ID and DBS rider ID are distinct and unique. The results are shown in Table 3, each matched pair corresponds to a single metro smart card ID paired with one unique DBS rider ID. Notably, all the matched pairs exhibit notably high spatialtemporal similarity values, signifying a substantial degree of spatial-temporal similarity among these matched pairs.

5.1. Verification of Identification of Matched Pairs. To verify the accuracy of the method proposed in this study for identifying DBS-metro transfer-matched pairs, smart card IDs and DBS rider IDs were obtained from a survey targeting frequent metro and DBS users, primarily including colleagues, family members, and friends, to ensure high reliability. We recruited a total of 448 volunteers who provided their DBS trip data and PT trip data. This dataset includes a total of 10,864 metro trips, among which 5,849 involved access transfers and 5,250 involved egress transfers. Detailed information is presented in Table 4.

Subsequently, a comparative analysis was performed on the identification results obtained through six different methods within the context of the study dataset, evaluating their accuracy when applied to survey data. *Method 1* applied metro station 300 m buffers from Ma et al. [14] and Zhao et al. [5] along with transfer time thresholds used in this study. Building on the foundation of *Method 1*, the mobility chain construction method introduced in the research was applied to establish *Method 3*, which was then further enhanced by integrating spatial-temporal similarity to develop *Method 5*. Similarly, this study uses a method based on station entry/exit points for identification of DBS-metro transfer travelers is *Method 6. Method 4* represents *Method 6* without the utilization of spatial-temporal similarity, and *Method 2* is *Method 4* without the application of mobility chains. In the context of survey data accuracy, it is essential to clarify that when there are many-to-one matched pairs, the accuracy is determined as follows: In *Method 1* through *Method 4*, where effective identification parameters are lacking, accuracy equals 1 if the correct matched pair has the highest and unique identification value. When the highest identification value is nonunique, the accuracy is denoted as 1/n (with *n* representing the number of matched pairs with the highest identification value). In cases where the correct matched pair does not possess the highest identification value, the precision is marked as 0. For *Method 5* and *Method 6*, the determination process aligns with the workflow in Section 4.4.

The performance results of the six methods on both the study dataset and survey data 517 are displayed in Table 5, with the best performance emphasized in bold font. It is evident that when applied to survey data, the method surpasses all others, achieving an accuracy exceeding 0.96, while the average number of matched pairs is less than 9.

Method 1 performs the least effectively among all the methods. While Method 1 is a commonly used approach for identifying DBS-metro transfer-related trips, its accuracy is limited because it relies solely on spatial and temporal thresholds and lacks the capability to filter the identified matched pairs. This is evident from the values of *Identified* DBS trips and Identified DBS riders. Method 1 exhibits the highest figures for Identified DBS trips and Identified DBS riders, indicating that it incorrectly categorizes numerous nontransfer trips and riders. Shifting to Method 2, which utilizes station entry/exit points, there is a significant reduction in the number of identified transfer trips and riders. The average number of matched pairs within the survey data shows notable improvements in both quantity and accuracy. It is worth noting that while Method 2 identified fewer transfer riders compared to Method 1, it identified 36,885 transfer riders that Method 1 did not. This emphasizes that adopting a station-based method may result in some omissions.

Regarding the impact of introducing mobility chains and spatial-temporal similarity, both station-based and entry/exitbased methods exhibit a reduction in the average number of matched pairs in survey data and an increase in accuracy. When comparing Methods 1 and 3 and Methods 2 and 4, the introduction of the mobility chain resulted in a 28.27% decrease in the average number of matched pairs within the survey data, coupled with an 18.36% increase in accuracy. In addition, when comparing Method 3 with Method 5, and Method 4 with Method 6, accuracy increased by 19.32% upon implementing spatial-temporal similarity to address the many-to-one matching pair issue. These findings collectively underscore the effectiveness of the methods employed in this research in resolving many-to-one matched pair challenges in DBS-metro transfer identification, ultimately enabling accurate identification of DBS-metro transfer travelers.

5.2. Analysis of the Results of the DBS-Metro Transfer Travelers. This study identified 2,499,809 DBS-metro transfer trips using *Method* 6, with 1,251,405 access transfers and 1,248,404 egress transfers comprising approximately equal



FIGURE 5: Workflows of deciding DBS-metro transfer travelers from matched pairs.

Matched pairs (smart card ID, DBS rider ID)	Number of DBS-metro transfer	Number of access transfer	Number of egress transfer	ST_Sim
<6100100021583387, 9a9b4a3dc8dfec6>	62	28	34	0.845
<6100100022252170, ef2110d2944ebc2>	56	28	28	0.912
<6100000100543053, 5b202a01f8684a9>	55	23	32	0.788
<6100010095193072, f17fc93511a7040>	55	29	26	0.938
<9094927857395712, bc0bcac29536e2d>	54	27	27	0.944
<9346061748873216, 5c186baa2592af6>	53	24	29	0.822
_		_	_	_

FABLE 3: Identified DBS-metro matched pa
--

TABLE 4: Survey of passengers' trip information.

Travel mode	Passenger count	Days of travel	Number of trips	Cycling distance (km)	Cycling durations
Metro	448	9.64	22.25	_	_
Access transfer	432	8.75	13.54	0.84	512
Egress transfer	415	8.32	12.65	0.78	526

TABLE 5: Comparison of different identification methods.					
(hour)	Identified DBS trips	Identified DBS	Survey data's average matched		

Method	Cost (hour)	Identified DBS trips	Identified DBS riders	Survey data's average matched pairs	Accuracy of survey data
Method 1	10.5	4979648	1324041	22.92	0.272
Method 2	8.9	2384959	981265	16.31	0.545
Method 3	15.8	4979648	1211156	18.48	0.471
Method 4	14.7	2384959	883138	10.25	0.713
Method 5	16.9	4979648	1211156	18.48	0.688
Method 6	18.3	2384959	883138	10.25	0.882

proportions. Daily averages reached 156,238 DBS-metro transfer cycling trips. Statistics on the number of users who took the transfer trips (Table 6) show that a total of 883,138 riders completed transfers during the study period. While the distribution of riders' transfer trips followed an exponential pattern, 39,035 riders still completed 10 or more transfers, indicating a stable user base overall.

To analyze individual-specific matched transfer trips, personal travel logs were derived and visualized, with six examples showcased in Figure 6. In this illustration, metro trips are represented by blue dotted line segments, DBS trips by red dotted line segments, and bus trips by solid black line segments. The diagram encompasses all metro, bus, and DBS trips spanning 16 days. The DBS-metro transfer activities of individuals are easily discerned through the adjacency of metro and DBS trips.

Individuals 2, 3, 5, and 6 exhibit regular commute patterns using DBS for consistent first/last mile with metro trips. They regularly employ DBS for their first- and last-mile connections to the metro. Many of their metro trips are accompanied by corresponding DBS trips, revealing specific symmetries between their morning and afternoon/evening trip sequences. Among these individuals, both Individual 2 and Individual 5 consistently opt for access transfers and egress transfers during their morning commutes but use DBS services irregularly in the afternoon. Individual 3 maintains a fixed OD point for morning trips, predominantly relying on access transfers. However, his afternoon trips lack such consistency. Conversely, Individual 6 consistently employs egress transfers for both morning and evening trips.

On the other hand, Individuals 1 and 4 likely do not conform to the traditional office worker commuting pattern, yet they utilize DBS as a practical solution for addressing first- and last-mile transportation challenges. Individual 1, in particular, selects DBS as a transfer method at various metro stations. Individual 4, with fewer travel days, exhibited DBS transfers at multiple metro stations on December 7, with no other DBS trips observed during the study period.

5.3. Comparison of Travel Patterns of DBS-Metro Transfer Travelers

5.3.1. Travel Distance and Travel Duration. To elucidate the distinctions in travel patterns between access transfers and egress transfers within DBS-metro transfers, this study initial step involved a comparative analysis of the travel time and distance associated with these two types of trips. The

travel distance of each trip was determined using the Manhattan distance, a widely accepted approach in transportation research [39], calculated from the coordinates of the origin and destination points. In addition, the travel duration for each trip was directly computed by subtracting the trip's start time from its end time.

Figure 7 presents the probability density distributions of travel distances and durations associated with DBSmetro transfer trips. For DBS-metro access transfers, the median travel distance is 943.27 meters, while for egress transfers, it is 959.77 meters. The travel distance intervals with the highest occurrence rates are 650 meters to 700 meters for access transfers and 700 meters to 750 meters for egress transfers. Beyond 1.5 kilometers, employing a DBS involves considerable physical effort, which makes metro passengers more inclined to opt for alternative transportation modes to reach the metro station [17, 40]. Within a travel distance of 1.5 kilometers, DBSmetro transfer is the choice of 76.69% for access transfers and 75.61% for egress transfers. It is particularly noteworthy that there is a significant difference between the mode and median of the cycling distances for access transfer and egress transfer. The reasons for this phenomenon may be attributed to several factors. Firstly, the urban geographical layout significantly influences cycling distances. Certain areas become hotspots for cycling due to geographical advantages or concentrated city functions, leading to an increased frequency of short-distance cycling. Additionally, the distribution of cycling distances exhibits asymmetry; while short-distance cycling constitutes the majority, the presence of longer-distance cycling raises the median distance. Lastly, the diversity in starting and ending point choices and travel preferences among different user groups is a key factor causing variations in cycling distances, reflecting the heterogeneity of travel needs. This observation aligns with findings from prior research. In terms of travel durations, the median duration for access transfer trips is 412 seconds, while for egress transfer trips, it is 443 seconds. Notably, most of these trips have durations of less than 840 seconds. To be precise, access transfer trips with durations under 840 seconds constitute 85.59% of the total, and similarly, egress transfer trips with durations under 840 seconds make up 82.92%.

5.3.2. Temporal Usage Patterns. Figure 8 illustrates the daily distribution of DBS-metro transfer trips over the study period. Notably, the number of these trips is considerably lower on weekends compared to weekdays. Among the weekdays,

TABLE 6: Number of DBS-metro transfer rider transfers.

Number of rides	Number of DBS riders	Percentage (%)	Number of rides	Number of DBS riders	Percentage (%)
1	403693	45.61	6	39682	4.48
2	160803	18.17	7	30894	3.49
3	85078	9.61	8	17538	1.98
4	50007	5.65	9	15051	1.70
5	43357	4.90	≥10	39035	4.41
	Total			883138	



1248 Metro station code ● – –● Metro trip ■ Bus trip ▲ – · –▲ DBS trip





FIGURE 7: Distributions of travel distance and travel time for DBS-metro transfer trips.



FIGURE 8: Distribution of daily cycling for DBS-metro transfer trips.

December 11 records the highest number of DBS-metro transfers, approaching 195,000 trips, while December 3 has the fewest, with less than 150,000 trips. Turning to the weekends, December 6 witnesses the most DBS-metro transfers at 136,818, and December 13 has the lowest, with less than 10,0000 transfers. In terms of the ratio between access transfers and egress transfers, the difference between the two remains consistently within 5% throughout the study period, except for December 5, when the difference reached 5.6%. This variation can likely be attributed to reduced commuting patterns during the weekend, resulting in irregular DBS-metro transfer behaviors among individuals.

In Figure 9, heat maps have been generated to facilitate a comparative analysis of the temporal patterns of DBS-metro transfer trips. Within the figure, the intensity of coloration denotes the degree of usage, with darker shades indicating higher levels of activity. On weekdays, both access transfers and egress transfers in the DBS trip patterns exhibit distinctive morning peaks, occurring at 8:00 and 9:00, respectively, as well as evening peaks at 17:00 and 18:00, respectively. It is noteworthy that the morning peak period for access transfers reveals two discernible peaks, whereas egress transfers exhibit only one. This observation has implications for future transport management strategies; it suggests the need to augment bicycle availability in residential areas before the morning peak to cater to the first-mile demand. Subsequently, close attention should be paid to the management of DBS parking facilities near stations to prevent fleet congestion. Preceding the evening peak, transport managers should focus on increasing bicycle availability near metro stations to meet the last-mile demand. On weekends, access transfers and egress transfers demonstrate a more evenly distributed pattern. Nevertheless, access transfers continue to exhibit significant morning peak-hour peaks, signifying a preference among commuters for cycling to metro stations.

5.3.3. Spatial Usage Patterns. Given the aggregation of the DBS-metro transfer trips by station, the spatial distributions of the total number of trips on weekdays and weekends could be obtained, as shown in Figure 10. Overall, the DBS-metro transfer is mainly concentrated in the urban area on both weekdays and weekends, which may be due to two factors. First, a few DBS bikes are distributed in the suburbs, so it is difficult for passengers to find them easily and quickly. Second, the densities of metro stations in the suburbs is low, and the distances between the origins of passengers and metro stations or between metro stations and their final destinations are long; therefore, it is very physically demanding to use DBS to connect metro stations in the suburbs.

An analysis of DBS-metro transfers at various stations reveals that on weekdays, such transfers occur at all stations. However, on weekends, notably suburban stations like Tianfu New Area and Xinjin, located at the termini of Line 1 and Line 10, respectively, experience a complete absence of transfers, both in access and egress categories. Conversely, metro stations situated in suburban areas such as Xipu Station (at the terminus of Line 2), Petroleum University, and Chengdu Medical College (at the terminus of Line 3) exhibit consistently higher transfer volumes, both on weekdays and weekends. This phenomenon can likely be attributed to the former station's proximity to primarily residential areas, resulting in elevated commuting activity during the weekdays but a reduced availability of bicycles on weekends. In contrast, the latter station enjoys its proximity to regions with substantial DBS usage, including university campuses, where cycling remains prevalent even on weekends. Considering these findings, future operators and transportation managers should contemplate augmenting DBS supply in the vicinity of suburban metro stations with relatively low transit accessibility.



FIGURE 9: Dynamics of the number of DBS-metro transfer trips on weekdays and weekends. (a) Dynamics of number of DBS-metro transfer trips on weekdays. (b) Dynamics of number of DBS-metro transfer trips on weekends.



FIGURE 10: Spatial distribution of DBS-metro transfer trips on weekdays and weekends.

6. Conclusion and Discussion

This study presents an innovative data identification method employing mobility chains to establish associations between metro passengers and DBS riders sharing the same identity. The process commences by extracting DBS trips related to DBS-metro transfers through the buffer zone based on metro station entry/exit points. Subsequently, integration with bus trip data from the same smart card system and TAZ data enables the establishment of DBS and PT mobility chains. These mobility chains serve as the foundation for analyzing spatiotemporal conflicts and similarities among potential correspondences between DBS riders and metro passengers. The accuracy of the identification method was validated through rigorous testing with survey data and individual travel logs. This approach, in contrast to the metro station-based method, significantly enhances matching accuracy in many-to-one scenarios. The introduction of the mobility chain led to a remarkable 28.27% reduction in the average number of matched pairs within the survey data and an 18.36% improvement in accuracy. Furthermore, implementing spatial-temporal similarity to resolve many-to-one matching pairs resulted in an additional 19.32% accuracy increase. In summary, this method amalgamates transfer identification, mobility chain construction, spatial-temporal filtering, and similarity metrics to precisely identify DBS-metro transfer travelers, facilitating in-depth analysis of multimodal travel behavior.

Furthermore, based on the identification results, a spatial-temporal analysis was conducted on the travel patterns of DBS-Metro transfers. The analysis revealed median travel distances of 943.27 meters for DBS-metro access transfers and 959.77 meters for egress transfers. Notably, travel distances of 650-700 meters for access transfers and 700-750 meters for egress transfers are the most common. Beyond 1.5 kilometers, the physical effort required for using DBS leads to passengers choosing alternative transport options. Regarding temporal patterns, weekday access transfers exhibit two morning peaks at 8:00 and 9:00, while egress transfers have one at 17:00. This suggests the need for bicycle availability adjustments near metro stations for firstmile and last-mile demand. During weekends, the temporal distribution of access transfers is notably more uniform than that of egress transfers. This suggests a preference for cycling to metro stations. Spatially, DBS-metro transfers are mainly concentrated in urban areas due to DBS availability and metro station density. However, suburban stations like Tianfu New Area and Xinjin have low transfer activity on weekends. In contrast, stations near residential areas or university campuses show higher transfer volumes. Future transport strategies should consider enhancing DBS supply near suburban metro stations and facilitating bike availability adjustments based on travel patterns to meet varying demands.

However, this study is subject to certain limitations. Firstly, when validating the survey data results, the scale of the survey data is relatively small, consisting of only 448 individuals. This is significantly less in comparison to the substantial numbers of metro passengers and DBS riders, which could potentially introduce biases. Secondly, the empirical data in this study are derived exclusively from a 16-day dataset of DBS and PT trips in Chengdu. Consequently, the analysis is confined to this specific timeframe, failing to encompass potential variations across distinct months or seasons. Furthermore, while this research has effectively examined the spatial and temporal distribution patterns of travelers on DBS-metro transfer trips and travel behaviors, it has refrained from delving into the underlying causal mechanisms. Hence, future research endeavors will prioritize the refinement of identification methods, the expansion of sample sizes for empirical investigations, and a more exhaustive exploration of urban travelers' DBSmetro transfer patterns along with their associated determinants.

Data Availability

The data used to support the findings of this study have not been made available because the authors have signed the confidentiality agreement with the data providers.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was funded by the Chengdu Key Research and Development Support Program for Technological Innovation and Development Projects (Grant no. 2022-YF05-00302-SN), Young Scientists Fund of the National Natural Science Foundation of China (Grant no. 52002127), and Sichuan Science and Technology Program (Grant nos. 2022YFG0197 and 2022JDR0324).

References

- X. Ma, Y. Ji, Y. Yuan, N. Van Oort, Y. Jin, and S. Hoogendoorn, "A comparison in travel patterns and determinants of user demand between docked and dockless bike-sharing systems using multi-sourced data," *Transportation Research Part A: Policy and Practice*, vol. 139, pp. 148–173, 2020.
- [2] J. Lazarus, J. C. Pourquier, F. Feng, H. Hammel, and S. Shaheen, "Micromobility evolution and expansion: understanding how docked and dockless bikesharing models complement and compete – a case study of san francisco," *Journal of Transport Geography*, vol. 84, Article ID 102620, 2020.
- [3] P. Zhao and S. Li, "Bicycle-metro integration in a growing city: the determinants of cycling as a transfer mode in metro station areas in beijing," *Transportation Research Part A: Policy and Practice*, vol. 99, pp. 46–60, 2017.
- [4] Z. Gan, M. Yang, T. Feng, and H. J. P. Timmermans, "Examining the relationship between built environment and metro ridership at station-to-station level," *Transportation Research Part D: Transport and Environment*, vol. 82, Article ID 102332, 2020.
- [5] D. Zhao, W. Wang, G. P. Ong, and Y. Ji, "An association rule based method to integrate metro-public bicycle smart card

data for trip chain analysis," Journal of Advanced Transportation, vol. 2018, Article ID 4047682, 11 pages, 2018.

- [6] Y. Liu, Y. Ji, T. Feng, and H. Timmermans, "Understanding the determinants of Young commuters' metro-bikeshare usage frequency using big data," *Travel Behaviour and Soci*ety, vol. 21, pp. 121–130, 2020.
- [7] Y. Liu, Y. Ji, T. Feng, and Z. Shi, "A route analysis of metrobikeshare users using smart card data," *Travel Behaviour and Society*, vol. 26, pp. 108–120, 2022.
- [8] L. Liu, L. Sun, Y. Chen, and X. Ma, "Optimizing fleet size and scheduling of feeder transit services considering the influence of bike-sharing systems," *Journal of Cleaner Production*, vol. 236, Article ID 117550, 2019.
- [9] Z. Zhan, Y. Guo, R. B. Noland, S. Y. He, and Y. Wang, "Analysis of links between dockless bikeshare and metro trips in Beijing," *Transportation Research Part A: Policy and Practice*, vol. 175, Article ID 103784, 2023.
- [10] Y. Lv, D. Zhi, H. Sun, and G. Qi, "Mobility pattern recognition based prediction for the subway station related bike-sharing trips," *Transportation Research Part C: Emerging Technologies*, vol. 133, Article ID 103404, 2021.
- [11] C. Zhou, X. Zhao, D. Xie, and J. Bi, "Understanding individuals' choice-making mechanism in trip chains of shared autonomous vehicles," *Travel Behaviour and Society*, vol. 33, Article ID 100619, 2023.
- [12] F. Schneider, W. Daamen, and S. Hoogendoorn, "Trip chaining of bicycle and car commuters: an empirical analysis of detours to secondary activities," *Transportmetrica: Transportation Science*, vol. 18, no. 3, pp. 855–878, 2022.
- [13] X. Liu, W. Li, Y. Li, J. Fan, and Z. Shen, "Quantifying environmental benefits of ridesplitting based on observed data from ridesourcing services," *Transportation Research Record*, vol. 2675, no. 8, pp. 355–368, 2021.
- [14] X. Ma, Y. Ji, M. Yang, Y. Jin, and X. Tan, "Understanding bikeshare mode as a feeder to metro by isolating metrobikeshare transfers from smart card data," *Transport Policy*, vol. 71, pp. 57–69, 2018.
- [15] Y. Guo, L. Yang, Y. Lu, and R. Zhao, "Dockless bike-sharing as a feeder mode of metro commute? The role of the feeder-related built environment: analytical framework and empirical evidence," *Sustainable Cities and Society*, vol. 65, Article ID 102594, 2021.
- [16] D. Lin, Y. Zhang, R. Zhu, and L. Meng, "The analysis of catchment areas of metro stations using trajectory data generated by dockless shared bikes," *Sustainable Cities and Society*, vol. 49, Article ID 101598, 2019.
- [17] X. Liu, J. Fan, Y. Li, X. Shao, and Z. Lai, "Analysis of integrated uses of dockless bike sharing and ridesourcing with metros: a case study of Shanghai, China," *Sustainable Cities and Society*, vol. 82, Article ID 103918, 2022.
- [18] S. Yu, G. Liu, and C. Yin, "Understanding spatial-temporal travel demand of free-floating bike sharing connecting with metro stations," *Sustainable Cities and Society*, vol. 74, Article ID 103162, 2021.
- [19] X. Zhou, Q. Dong, Z. Huang, G. Yin, G. Zhou, and Y. Liu, "The spatially varying effects of built environment characteristics on the integrated usage of dockless bike-sharing and public transport," *Sustainable Cities and Society*, vol. 89, Article ID 104348, 2023.
- [20] Y. Ji, Y. Fan, A. Ermagun, X. Cao, W. Wang, and K. Das, "Public bicycle as a feeder mode to Rail transit in China: the role of gender, age, income, trip purpose, and bicycle theft experience," *International Journal of Sustainable Transportation*, vol. 11, no. 4, pp. 308–317, 2017.

- [21] J. Wang, M.-P. Kwan, W. Cao, Y. Gong, L. Guo, and Y. Liu, "Assessing changes in job accessibility and commuting time under bike-sharing scenarios," *Transportmetrica: Transportation Science*, vol. 20, pp. 1–17, 2022.
- [22] K. Kim, "Investigation of modal integration of bike-sharing and public transit in seoul for the holders of 365-day passes," *Journal* of Transport Geography, vol. 106, Article ID 103518, 2023.
- [23] R. Xin, J. Yang, B. Ai, L. Ding, T. Li, and R. Zhu, "Spatiotemporal analysis of bike mobility chain: a New perspective on mobility pattern discovery in urban bike-sharing system," *Journal of Transport Geography*, vol. 109, Article ID 103606, 2023.
- [24] M. Bordagaray, L. dell'Olio, A. Fonzone, and A. Ibeas, "Capturing the conditions that introduce systematic variation in bike-sharing travel behavior using data mining techniques," *Transportation Research Part C: Emerging Technol*ogies, vol. 71, pp. 231–248, 2016.
- [25] X. Ma, S. Zhang, Y. Jin, M. Zhu, and Y. Yuan, "Identification of metro-bikeshare transfer trip chains by matching docked bikeshare and metro smartcards," *Energies*, vol. 15, no. 1, p. 203, 2021.
- [26] D. Guo, E. Yao, S. Liu, R. Chen, J. Hong, and J. Zhang, "Exploring the role of passengers' attitude in the integration of dockless bike-sharing and public transit: a hybrid choice modeling approach," *Journal of Cleaner Production*, vol. 384, Article ID 135627, 2023.
- [27] Y.-H. Cheng and W.-C. Tseng, "Exploring the effects of perceived values, free bus transfer, and penalties on intermodal metro-bus transfer users' intention," *Transport Policy*, vol. 47, pp. 127–138, 2016.
- [28] X. Li, Q. Yan, Y. Ma, and C. Luo, "Spatially varying impacts of built environment on transfer ridership of metro and bus systems," *Sustainability*, vol. 15, no. 10, p. 7891, 2023.
- [29] X. Feng, H. Sun, J. Wu, Z. Liu, and Y. Lv, "Trip chain based usage patterns analysis of the round-trip carsharing system: a case study in Beijing," *Transportation Research Part A: Policy* and Practice, vol. 140, pp. 190–203, 2020.
- [30] D. Zhao, G. P. Ong, W. Wang, and X. J. Hu, "Effect of built environment on shared bicycle reallocation: a case study on nanjing, China," *Transportation Research Part A: Policy and Practice*, vol. 128, pp. 73–88, 2019.
- [31] Y. Liu, T. Feng, Z. Shi, and M. He, "Understanding the route choice behaviour of metro-bikeshare users," *Transportation Research Part A: Policy and Practice*, vol. 166, pp. 460–475, 2022.
- [32] Gaode Maps, "Traffic analysis report of major cities in China," 2020, https://www.199it.com/archives/1194887.html.
- [33] GeoPy, 2023, https://geopy.readthedocs.io/en/stable/.
- [34] X. Zhao, Y. Zhang, Y. Hu, Z. S. Qian, H. Liu, and B. Yin, "Modeling relation proximity of passengers using public transit smart card data," *IEEE Intelligent Transportation Systems Magazine*, vol. 14, no. 1, pp. 163–172, 2022.
- [35] X. Zhao, Y. Zhang, Y. Hu et al., "Interactive visual exploration of human mobility correlation based on smart card data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 8, pp. 4825–4837, 2021.
- [36] Y. Rubner, C. Tomasi, and L. J. Guibas, "The Earth mover's distance as a metric for image retrieval," *International Journal* of Computer Vision, vol. 40, no. 2, pp. 99–121, 2000.
- [37] O. Pele and M. Werman, "A linear time histogram metric for improved SIFT matching," in *Proceedings of the Computer Vision – ECCV 2008*, D. Forsyth, P. Torr, and A. Zisserman, Eds., pp. 495–508, Springer, Berlin, Germany, 2008.

- [38] M. Zaharia, R. S. Xin, P. Wendell et al., "Apache spark: a unified engine for big data processing," *Communications of the ACM*, vol. 59, no. 11, pp. 56–65, 2016.
- [39] L. He, B. Agard, and M. Trépanier, "A classification of public transit users with smart card data based on time series distance metrics and a hierarchical clustering method," *Transportmetrica A Transport Science*, vol. 16, no. 1, pp. 56–75, 2020.
- [40] Y. Guo and S. Y. He, "Built environment effects on the integration of dockless bike-sharing and the metro," *Transportation Research Part D: Transport and Environment*, vol. 83, Article ID 102335, 2020.