Rhythm of Transit Stations - Uncovering the Activity-Travel Dynamics of Transit-Oriented Development in the U.S.

Zhuangyuan Fan, Fan Zhang, and Becky P. Y. Loo

Abstract— **Existing transit-oriented-development (TOD) classification studies primarily focus on the static characteristics around transit stations to measure the built environment's density, diversity, and design. As a community development model, time-variant variables, dynamic human activities throughout different times of the day and week matter in further unpacking the characteristics of TODs. Given that this aspect has been under-discussed in most previous TOD literature, this research provides an activity-based framework to classify commuter transit station areas by considering the degree of local vibrancy - the temporal visiting pattern of all points of interest (POIs) that fall within the station areas. We apply a two-step semi-unsupervised clustering algorithm to classify 4,290 station areas from 54 metropolitan areas across the U.S. This method produces 13 distinct station area types. Next, we further examine the connection between station area types and neighborhood travel behavior. A cross-sectional comparison reveals that stations with consistent active morning activities are associated with a higher ratio of commuting by walking and biking and lower automobile usage measured in vehicle miles traveled (VMT). Using stations opened after 2009, we show that active weekend activity patterns are associated with a more significant increase in commuting by public transit.**

Index Terms— **Transit-oriented development (TOD), human dynamics, clustering analysis.**

I. Introduction

TRANSIT-ORIENTED development (TOD) is a well-known community model; its successful implementation has potential benefits at both station and community levels [1]. Policymakers and planners often customize the concept of TOD to achieve planning visions in a given area. While the concept generally refers to high density, mixed-use, and pedestrian-friendly design, implementation measures vary substantially [2], [3]. For example, TODs' environmental and social outcomes are measured through changes in travel behaviors and employment rates of adjacent neighborhoods. Economic outcomes are measured through changes in land values and housing prices. These measures also vary across geographical regions considering the heterogeneous station contexts. The intricate relationship between station areas' development pattern, local communities' life culture, and visitors' potential inflow implies complexity for planners and policymakers interested in assessing TODs' overall characteristics and developing ingredients to further enhance TODs' environmental, social, and economic impact.

Then, how do we decipher this complexity and create a better understanding of TOD performance at the neighborhood level? City planners and researchers have created tools that quantify the TOD characteristics and evaluate their performance. For example, the Center for Transit-Oriented Development (CTOD) developed a Performance-Based TOD typology tool that helps people to assess the performance of transit zones by considering the land use and household vehicle miles traveled (VMT) [4]. Zemp *et al.* [5] applied a hierarchical cluster analysis on 10 indicators to classify 1700 railway stations in Switzerland. Loo *et al.* [6] used principal component analysis on indicators to classify the metro stations in Hong Kong. These studies have produced TOD typologies according to the extended 'D' variables, including but not limited to density, diversity, and design [7]. They also examine the correlation between these variables and changes in travel behaviors such as reduced VMT and increased public transit ridership [6], [8]. Frameworks as such provide valuable insights into resolving the complex context and potential outcome of TODs but also imply limitations. First of all, TOD, as a community development model that connects visitors who use the transit services for commute and local residents who use parks, retails, and restaurants on a daily bases, is a highly temporal concept. However, most existing TOD classification methods resort to static characteristics of transit stations' catchment areas, ignoring how people use the station areas throughout the day and week. Secondly, the social and environmental benefits of TODs hinge on the extent to which the adjacent neighborhood could shift vehicle trips to public transit, biking, or other more environmentally friendly transit modes. Hence, a study of travel behavior is essential.

In view of these limitations, this study first creates an activity-based classification framework to define station areas by how visitors use these areas through time and space. Next, we conduct an experiment using 4,290 fixed-guideway transit stations from 54 metropolitan areas in the U.S. with hourly visiting activities from 701,915 points of interest (POIs). Then,

Manuscript received 26 January 2021; revised 8 July 2021; accepted 15 September 2021. Date of publication 26 October 2021; date of current version 9 August 2022. This work was supported by the National Natural Science Foundation of China under Grant 41901321. The Associate Editor for this article was M. Mesbah. *(Corresponding author: Becky P. Y. Loo.)*

Zhuangyuan Fan and Becky P. Y. Loo are with the Department of Geography, The University of Hong Kong, Hong Kong (e-mail: yuanzf@connect.hku.hk; bpyloo@hku.hk).

Fan Zhang is with the Senseable City Laboratory, Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: zhangfan@mit.edu).

Digital Object Identifier 10.1109/TITS.2021.3115103

1558-0016 © 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Zhejiang Lab. Downloaded on April 18,2023 at 07:32:22 UTC from IEEE Xplore. Restrictions apply.

we quantify the travel behaviors of residents living within these station neighborhood areas between 2009 and 2017 and estimate the types of station area development that contribute more to environmentally friendly transit modes and lower VMT.

To derive the visiting activities within station areas, we use hourly POI visiting pattern data inferred from aggregated smartphone application data. In recent years, the increasingly available location-based services (LBS) data generated from mobile applications have empowered researchers to re-examine transportation planning policies based on how people move and behave at a micro-scale. Ratti *et al.* [9] introduced the concept of *Mobile Landscape* and visualize the movement of people compared to the World Cup Games in Milan [10]. Pei *et al.* [11] analyzed the mobile phone usage data to predict the land use pattern in Singapore. Phithakkitnukoon *et al.* [12] used mobile phone data to infer the social influence of transit mode. These research studies shed light upon the temporal layers of cities and support urban planning decision-making by taking into account activity-travel dynamics happening in cities every day.

Following the same line, this study leverages the available aggregated spatial-temporal location data and explicitly examines the temporal aspects of transit stations. A two-step semi-unsupervised clustering algorithm is applied to classify all stations into typologies based on their associated visitor activities. Here we ask two main research questions: 1) How many different station areas exist from the understanding of how people use the catchment area amenities dynamically? 2) How do different types of stations impact the changes in local travel behaviors?

The rest of the paper is organized in the following manner. Section II reviews previous work on TOD classification and associated travel behaviors. Section III introduces the methodology we used in the classification process. Section IV describes the data and experiment results. Section V concludes our study.

II. Literature Review

A. Existing TOD Classification Frameworks

TOD classification has received much attention in the last two decades. Policymakers and planners primarily adopt normative classification methods to create TOD typologies that consider its contexts. For example, the City of Denver [13] created five different station types along the city's LRT and CRT lines: downtown, urban center, general urban, urban, suburban. These typologies are used so that specific zoning allowances can be designed accordingly. Austin *et al.* [4] created 15 different station types for 3,760 existing transit station areas across the U. S. utilizing the household VMT and land use mix. These studies are normative as they describe the general contexts of each station area and then conduct linear transformation of the factors such as densities, housing types, and available transit services.

Researchers have adopted more non-linear methods to create TOD typologies according to the 'D' variables: density, diversity, and design. One approach is to organize the 'D' variables according to nodal and place-based measures. Nodal functions of a station area include acting as access points to transit or regional transportation networks. Places' functions are neighborhoods featured by mixed-uses, high employment density, diverse amenities, and pedestrian-friendly design [14]–[16]. Along this line, several researchers have been engaged in analyzing the TOD performance of transit stations using a node-place index. For example, Reusser *et al.* [17] applied a hierarchical cluster analysis to classify 1600 rail stations in Switzerland. Here, 11 node-place indicators are included to produce five-station types: smallest, small, mid-sized in populated areas, mid-sized but unstaffed, and large to very-large stations in major urban centers. Zemp *et al.* [5] applied a principal components analysis to classify 1700 railway stations in Switzerland with ten indicators. Their method led to a 7-cluster solution: central stations, large connectors, medium commuter feeders, small commuter feeders, tiny tourist stations, isolated tourism nodes, and remote destinations.

Beyond the node and place models, other researchers also extend the 'D' variables to assessing the TOD inputs against policy expectations. For example, Higgins and Kanaroglou [18] adopted a latent class model-based clustering algorithm to classify 372 rapid transit stations in the Toronto region according to five 'D' variables: density, diversity, distance to transit, design, and destination accessibility. Their work reveals ten station types. They found that stations with high measures such as density, walkability, and mixed-uses were more likely to have higher transit ridership rates, walking and cycling, and household vehicle kilometers traveled (VKT). Moreover, some studies [19], [20] also advanced the 'D' variables with specific station characteristics such as years of operation and generalized travel cost from the station to the downtown area.

One major caveat to the existing literature classifying TODs based on the 'D' variables lies in the fact that TOD, as a transportation and mobility-related concept, is highly temporal. The majority of literature primarily resort to static 'D' features such as density, land use diversity, and local socioeconomic characteristics, thus do not further unpack the temporal characteristics of TODs.

B. Travel Behaviors, TODs, and Land Use

Taken as a form of land use strategy to minimize automobile dependency and maximize public transit ridership, TODs are expected to have an impact on local household travel behavior and address the land use issue simultaneously [1], [14]. Researchers have identified land uses features that are correlated with changes in travel behaviors. They include the level of land use mixing [21], pedestrian amenities [22], diversity of transit modes [1], [23], housing density [24], employment density [25], and accessibility to public spaces [26].

Although these studies have a similar goal to identify critical features in creating and managing TODs, they vary in the measures of outcomes of TODs. Broadly speaking, TODs' expected outcomes include local economic growth and environmental sustainability, and social equity. Many of these benefits hinge on the effects of TOD in supporting more sustainable travel behaviors [27], [28]. The measures

Authorized licensed use limited to: Zhejiang Lab. Downloaded on April 18,2023 at 07:32:22 UTC from IEEE Xplore. Restrictions apply.

of TOD-related travel behaviors include VMT, travel mode, auto-ownership, and transit ridership [7], [19], [29]–[31].

For example, Pan *et al.* [32] used the transit integrated circuit (IC) card data and the cellular signaling data of Shanghai to study the connection between stations and daily passenger volume of stations. They found stations with better commercial development also have larger passenger volumes. Loo *et al.* [19] measured the passenger's weekly patronage in relation to the TOD characteristics and found that context-specific station characteristics have the most significant correlation with passengers weekly patronage. Cai *et al.* [33] developed a visualization method to describe the land use characteristics around station areas and use passenger volumes to validate the derived land-use characteristics. Stiffler [34] used VMT to compare the difference between TOD and non-TOD neighborhoods in Carlsbad, CA.

These studies have presented valuable insights validating the impacts of TODs with different land-use compositions in different contexts. However, there are three major caveats to the literature mentioned above. First, most empirical studies examining the connection between TODs and travel behaviors are limited to a small region or several cities, thus do not offer much insight in comparing TODs under different cultures and geographical contexts. Second, one of the vital identification challenges for TODs' contribution to neighborhoods' travel behaviors is a positive feedback loop suggested by the tipping model [35]. A station area with many residents preferring public transit in the past is likely to have even more public transit ridership afterward. Third, it is clear that people living within mixed-use regions travel differently, yet it is unclear to what extent stations with TOD characteristics reduce vehicle-dependent travel [36], [37]. Lastly, current literature emphasizes that a successful TOD will facilitate more local activities given their diversity in land use and employment opportunities [38], thus results in a reduction in vehicle-dependent travels. However, it is unclear if station areas that mainly attract visitors from far away would still generate the same effects on travel behavior as those with visitors mainly from the neighborhood areas.

III. METHODOLOGY

To fill the research gaps mentioned above, we present the following research framework. First, leveraging on available large-scale fine-grained foot traffic data derived from smartphone applications (apps), we create an activity-based classification method to classify all transit stations according to their associated visitors' activity. This method creates station types with a two-step clustering process (Figure 1a). Second, we design an empirical experiment to infer the connection between the activity-based station types and the local household travel behavior (Figure 1d). Lastly, we trace back to the original intention of TOD as a community-building model, further differentiate stations by average visitors' home distance, and evaluate the station types effects separately.

A. Two-Step Clustering to Create Station Types

To quantify the temporal aspects of the activity pattern around station areas, we first clarify the process of deriving the aggregated visiting activities for each station area. Then we apply the two-step clustering process to create station types based on the aggregated visiting activities.

1) POIs to Stations Association and Stations' Service Area: We first design a mechanism to find associated POIs for each station and define the station service area. Previous studies have defined the TODs' service area per different rules. For example, Cai *et al.* and Kong *et al.* [33], [39] used the Voronoi diagram to partition their study areas so that each POI is associated with only one station, and they consider the entire study space could be served by at least one station. Loo *et al.* [19] and Schuetz *et al.* [40] defined station areas as circles with different buffer radius based on the local context. Given that many stations included in this data may have various local contexts, we combine the Voronoi diagram and the buffer method to define station areas and find POIs' associated stations. We define a station's service area with two boundaries: a core boundary and a neighborhood boundary. Given all transit stations collection in the study $Stations = \{d_1, d_2, \ldots, d_i, \ldots, d_N\}$, and the total POI collection $POIs = \{p_1, p_2, p_3, \ldots, p_j, \ldots, p_M\}$, where N is the total number of stations included in the study, M is the total number of POIs, and $N \ll M$. We first use a Voronoi diagram to partition a whole study space based on the nearest neighbor principle of each station. Then for each partition, we use the service core area radius r_{core} and the neighborhood area radius $r_{neighbor}$ to define the intersected areas as either a station's service core area $Core(d_i)$ or neighborhood area $Neighborhood(d_i)$. Consequently, a POI within the $Core(d_i)$ is labeled as station d_i's core POI. A POI p_j within $Neighborhood(d_i)$ but outside $Core(d_i)$ is labeled as a station d_i's Neighborhood POI (Algorithm 1). The final defined areas have the following characteristics:

- Each area contains only one station;
- The distances from a point on a shared edge of the areas to the stations on both side of the edge are equal;
- For service core areas, the distance from a point in the area to the station is the shortest and is shorter than r_{core};
- For service neighborhood areas, the distance from a point in the area to the station is the shortest and is shorter than $r_{neighbor}$.

2) Station Areas' Weekly Visiting Pattern: To preserve both hourly visiting patterns and weekly dynamics around a transit station, we use a time-ordered sequence of a whole week (i.e., from Monday to Sunday, 168 hours in total) to represent the average POIs' visiting activities around a station. Then, based on these activity patterns, we create two features to describe each station's activity: 1) Average POI visiting pattern; 2) POI's hourly count of visits.

For all POIs within a station's $Core(d_i)$, we have a collection of POIs $CorePOIi = \{p_1, p_2, p_3, \ldots, p_j, \ldots, p_J\}$, where J is the total number of POIs within $Core(d_i)$. Each POI has its derived activity pattern vector S_{jt}, which indicates the average number of visits to p_j at hour t. To increase the robustness of the method, we average the visiting records of multiple weeks to calculate a one-week visiting pattern for

Authorized licensed use limited to: Zhejiang Lab. Downloaded on April 18,2023 at 07:32:22 UTC from IEEE Xplore. Restrictions apply.

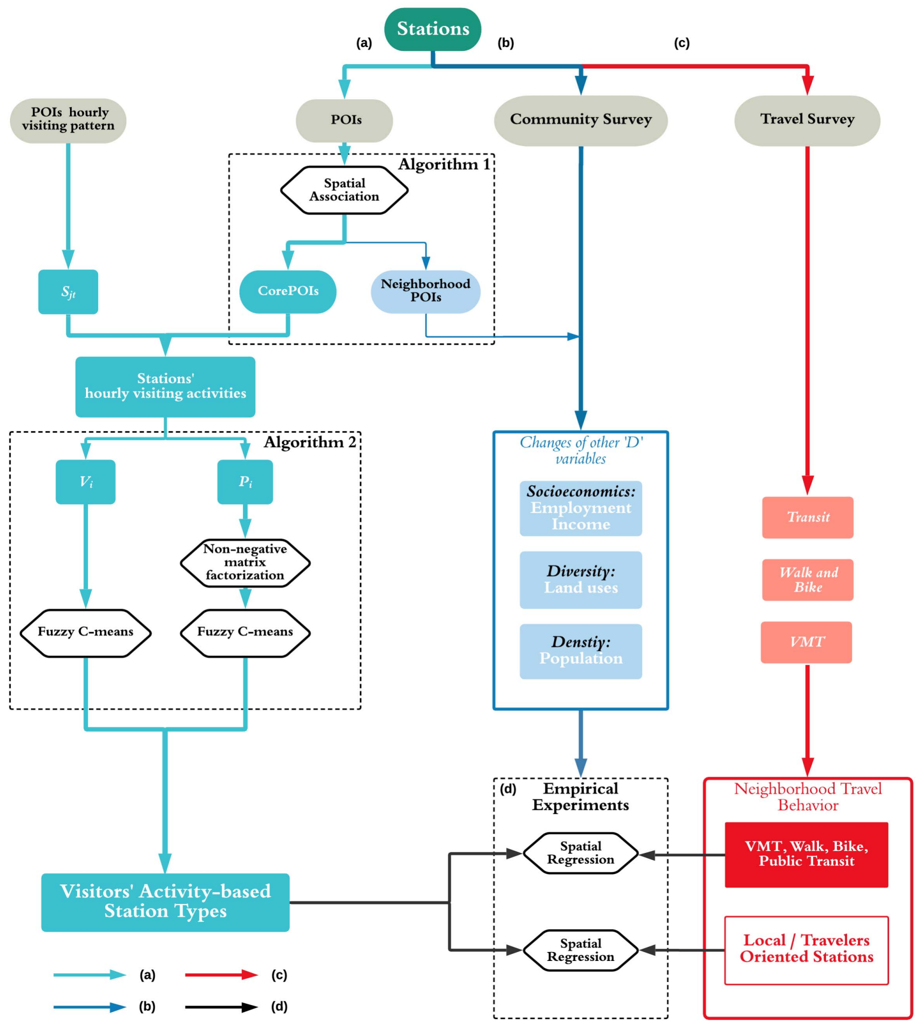


Fig. 1. Proposed research framework. (a) We use each station area's POI hourly visiting pattern data to derive the station types through a two-step clustering process, where S_{jt} shows the hourly number of visitors for each POI; V_i stands for the visit-count vector of each station area; P_i represents the pattern vector of each station area. (b) We obtain other 'D' variables associated with each POI neighborhood area. (c) Neighborhood travel habits are measured by VMT, percentage of workers commuting by public transit, and percentage of workers commuting by walking or biking for each station area. (d) We design an empirical experiment to uncover the effects of different stations on neighborhood travel behaviors.

each POI (Equation 1):

$$S_{jt} = \frac{\sum_{k=1}^{K} R_{jt}}{K}, \quad (j = 1, 2, 3, 4 \ldots J, \;\; k = 1, 2, \ldots, K, \;\; t = 0, 1, 2, \ldots, 167), \quad (1)$$

where R_{jt} is the average number of visitors during hour t in a week. K is the total number of weeks we use in this study. Then we describe the overall visiting pattern within the $Core(d_i)$ as P_{it}, which denotes the average number of visitors visiting POIs that falling within $Core(d_i)$ during hour t (Equation 2):

$$P_{it} = \frac{\sum_{j=1}^{J} S_{jt}}{J}, \quad (j = 1, 2, 3, 4 \ldots J, \;\; i = 0, 1, 2, \ldots, N, \;\; t = 0, 1, 2, \ldots, 167), \quad (2)$$

Based on the pattern vector generated with Equation 2, we construct a visit-count vector V_i to represent the average weekly number of visits for each POI within $Core(d_i)$, given by Equation 3:

$$V_i = log_{10}(1 + \sum_{t=0}^{167} P_{it}), \quad (i = 0, 1, 2, \ldots, N, \;\; t = 0, 1, 2, \ldots, 167), \quad (3)$$

Note that we logarithm transform the total volume over the week for ease of interpretation and stabilization of the variance.

3) Two-Step Clustering Process: After we construct V_i and P_{it} for each $Core(d_i)$, we apply a two-step clustering process to classify all station areas. The visit-count clustering process and the pattern clustering process are separated to isolate the context-specific conditions such as population density, job density, and dedicated land uses. A similar method was used by [11] to simulate land uses in Singapore with aggregated mobile phone usage data.

Authorized licensed use limited to: Zhejiang Lab. Downloaded on April 18,2023 at 07:32:22 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1 Associate POIs With Stations

Input: Stations: all stations in the study; POIs: all POIs in the study

Output: CorePOIs: POIs that fall within each core service area of stations; NeighborhoodPOIs: POIs that fall within each neighborhood area of stations

$Partitions = Voronoi(Stations)$
$CoreBuffer = Buffer(Stations, r_{core})$
$NeighborBuffer = Buffer(Stations, r_{neighbor})$
$Core = Intersect(Partitions, CoreBuffer)$
$Neighborhood = Intersect(Partitions, NeighborBuffer)$
$CorePOIs = \emptyset$
$NeighborhoodPOIs = \emptyset$
for each $d_i \in Stations$ **do**
  $CorePOI_i = Within(POIs, Core_i)$
  $CorePOIs.add(CorePOI_i)$
$POIs.delete(CorePOIs)$ ▷ Remove the POIs that exist in CorePOIs
for each $d_i \in Stations$ **do**
  $NeighborhoodPOI_i = Within(POIs, Neighoborhood_i)$
  $NeighborhoodPOIs.add(NeighborhoodPOI_i)$
return $CorePOIs, NeighborhoodPOIs$

To classify station areas into different categories, we first apply a Fuzzy C-mean (FCM) [41], [42] algorithm to V_i to classify the stations into K_1 clusters by their POIs' average count of visits. We use the elbow method [43] to identify the optimal K_1. For pattern vector P_{it}, we first apply a non-negative matrix factorization (NMF) [44] process to reduce the dimension of the normalized pattern vector P_{it}. Then we use an FCM clustering process to identify K_2 pattern clusters. This process is shown in Algorithm 2.

Algorithm 2 Two-Step Clustering Process

Input: V: Volume vector for each station; P: Pattern vector for each station.

Output: VolumeClusterLabels: The predicted cluster label for each station using visiting volume; PatternClusterLabels: The predicted cluster label for each station using visiting pattern;

$VolumeInertias = \emptyset$
$PatternInertias = \emptyset$
for $i \in range(2, C)$ **do** ▷ C is the largest possible number of clusters
  $fuzzyCmeanVolume = FCM(i)$
  $fuzzyCmeanVolume.fit(V)$
  $inertiaVolume_i \leftarrow computeInertia(fuzzyCmeanVolume, V)$
  $VolumeInertias.add(inertiaVolume_i)$
  $fuzzyCmeanPattern = FCM(i)$
  $fuzzyCmeanPattern.fit(P)$
  $inertiaPattern_i \leftarrow computeInertia(fuzzyCmeanVolume, P)$
  $PatternInertias.add(inertiaPattern_i)$
$K_1 = findElbow(VolumeInertias)$
$K_2 = findElbow(PatternInertias)$
$fuzzyCmeanVolume = FCM(K_1)$
$fuzzyCmeanVolume.fit(V)$
$VolumeClusterLabels = fuzzyCmeanVolume.predict(V)$
$fuzzyCmeanPattern = FCM(K_2)$
$fuzzyCmeanPattern.fit(P)$
$PatternClusterLabels = fuzzyCmeanPattern.predict(P)$
return $VolumeClusterLabels, PatternClusterLabels$

B. Empirical Framework

1) Stations Types and Neighborhood Travel Behaviors: For all station areas, we estimate the effects of station areas' activity types on the local travel behaviors through the following model:

$$Y = \sigma X + \theta D + \beta StationType + \rho \sum_j w_j Y + \epsilon, \quad (4)$$

where Y indicates the travel behavior described by VMT, percentage of workers commuting by public transit and percentage of workers commuting by walking or biking. X includes three types of data: 1) Socioeconomic factors including population density, median household income, and employment rate; 2) the number of years a station has been operated until the current survey time; 3) each station's distance from their closest major metropolitan areas' downtown district. D describes the density of POIs within stations' core area and neighborhood ring. A spatial weight w_j is introduced here to control for the neighborhood spillover effects suggested by the urban invasion theory [45].

Specifically, station area population is the proportional sum of population data from the survey geography intersects the station neighborhood. The station neighborhood's household income, employment rate, VMT, percentage of workers commuting by different modes are calculated by the average statistics weighted by the fraction of the survey base geography that intersects a station's neighborhood (See Appendix C for details).

The density of POIs is measured by the number of POIs that falling into the geographic study units divided by the true area of each associated boundary.

The spatial weights matrix used in this analysis is measured with Queens' contiguity, meaning that all station neighborhood areas sharing a boundary in any direction from the station neighborhood area in question are considered contiguous. If two station neighborhood areas (i and j) share at least one boundary, the contiguity weight is 1, 0 otherwise.

2) Local or Traveler-Oriented Station: In addition, we specifically examine the association between station types with travel behavior by differentiating the aggregated user profiles. To do so, we create a variable $HomeDistance_i$, which measures the average home distance of all visitors visiting

Authorized licensed use limited to: Zhejiang Lab. Downloaded on April 18,2023 at 07:32:22 UTC from IEEE Xplore. Restrictions apply.

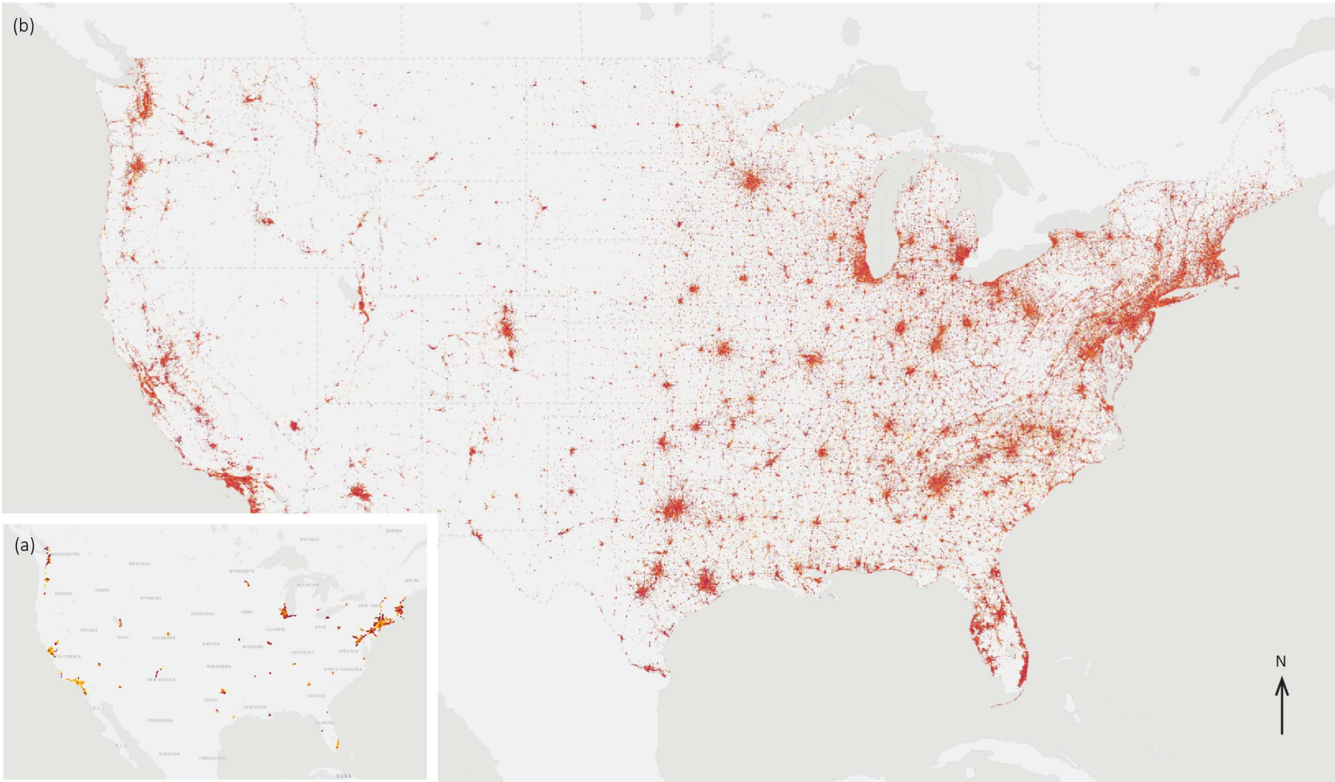


Fig. 2. (a) 4,417 existing fixed guideway transit stations included in the dataset. Yellow represents stations with higher concentration of POIs in the station core-areas; red indicates fewer POIs in the station core-areas. (b)Spatial distribution of 5,393,160 POIs in the original database. The POIs are represented using hexbins in the map. The map is produced with kepler.gl.

a station's core area $Core(d_i)$. For each POI p_j, we have estimated average visitors' home distance l_j, then we further derive the $HomeDistance_i = \frac{\sum_j^J l_j}{J}$. The larger the value of $HomeDistance_i$ is, the more likely it is that most visitors are travelers from distant neighborhoods. For a station's core area $Core(d_i)$, if its associated $HomeDistance_i$ is larger than the median value of all stations, we label it as a traveler-oriented station, local-oriented stations otherwise. After this process, we repeat the model of Equation 4.

3) Stations Types and Changes of Neighborhood Travel Behaviors: For stations that are built between two survey periods, we specifically examine if the establishment of these stations has effects on the neighborhood's travel behavior changes (Equation 5).

$$\Delta Y = \sigma \Delta X + \beta StationType + \rho \sum_j w_j \Delta Y + \epsilon \tag{5}$$

where ΔY indicates the travel behavior changes between two survey periods. ΔX is the change in time-variant socioeconomic factors. w_j is the spatial weight.

IV. Experiment

A. Data Description

1) Transit Station Data: The station data is downloaded from the Transit-Oriented-Development (TOD) database.[1] It contains 4,417 existing fixed guideway stations and 1,583 planned stations in 54 metropolitan areas in the US, as of December 2011 (Figure 2a). The fixed guideway stations included in this study are mostly commuter rail stations with a couple of multi-modal commuter stations such as the Aerial tram in Portland, OR, and commuter ferries in Seattle and San Francisco. These stations are included in the study, given that they play critical roles in the local commute network. We only include the stations currently in operations and have at least one POI associated with the station area in our study. This process left us with 4,309 stations. The location of the data is current as of January 2021. The data provided includes the station name, geolocation of each transit station, associated city, year opened, and the associated transit line.

2) Points of Interest (POI) and Visiting Pattern Data: Both POIs' description and their associated hourly visiting pattern data are provided by the Safegraph Consortium. POIs' detailed descriptions contain the names, geographic coordinates, and category code following the North American Industry Classification System (NAICS) for 12 types of business and service amenities in our study area. Summary of all POI data used in this study is included in the Appendix (A2).

POIs' hourly visiting data are inferred from aggregated anonymized mobile phone data. The location data from mobile user applications are clustered to each known POI for each hour in the entire U.S. The data provided includes each POI's name, unique ID, associated census block group id, hourly number of visitors per week, and average visitors' home distance in meters. In this research, we first query the entire POI database (5,393,160 POIs in total, Figure 2b) to select the POIs that are either within the stations' core areas or

[1] https://toddata.cnt.org/

Authorized licensed use limited to: Zhejiang Lab. Downloaded on April 18,2023 at 07:32:22 UTC from IEEE Xplore. Restrictions apply.

TABLE I

SUMMARY STATISTICS: NEIGHBORHOOD TRAVEL BEHAVIORS AND SOCIOECONOMIC STATUS OF ALL STATIONS INCLUDED IN THE STUDY

	count	mean	sd	min	max
2017 Panel					
% Commute w. Walk Bike	4265	11.338	12.917	0.000	81.240
% Commute w. Public Transit	4265	21.496	18.593	0.003	88.912
Estimated VMT	4265	25.282	13.744	0.000	72.030
Population Density	4265	15155.739	18769.339	37.090	133796.500
Median Household Income	4265	77814.693	41761.902	49.215	248371.297
Employment (%)	4265	92.190	8.879	0.122	100.000
2009 Panel					
% Commute w. Walk Bike	4268	11.077	12.739	0.000	84.105
% Commute w. Public Transit	4268	21.004	18.256	0.000	80.980
Estimated VMT	4268	28.895	15.715	0.000	84.962
Population Density	4268	13854.307	17912.039	26.746	125242.305
Median Household Income	4268	59312.172	31591.288	1095.776	234977.906
Employment (%)	4268	89.060	9.626	1.350	100.000
Other Factors					
Dist. City Center (km)	4276	13.555	13.714	0.011	84.956
Years of Operation	4290	20.652	2.943	10	22
POI Den. (Core) (#/sqmile)	4290	420.274	718.926	3.303	11397.177
POI Den. (Neighbor) (#/sqmile)	4290	62.468	85.296	0.000	1403.198

Stations in the Puerto Rico regions are removed from the calculation of distance to the nearest major metropolitan downtown area due to missing data from the database. Station areas that have no information of population or median household income are removed from the survey samples.

outside the core areas but inside stations' neighborhood areas. This process left us with 701,915 POIs for the study. Then we include the core POIs' hourly visiting patterns collected in April, May, and June of 2019 to derive the visit-count vector V_i and the pattern vector P_{it} . Figure 3 demonstrates the spatial relationships of the core POIs and the neighborhood POIs.

Although Safegraph's aggregated data covers around 10% of the mobile devices in the U.S., it is still a sample of the entire population. Considering the potential sample bias, we apply a weighting mechanism to adjust the number of visits for each POI based on the ratio of smartphone users to the true population in each census block group (See Appendix for detailed calculation). The adjusted value and the sampled value are 97% correlated.

3) Travel Behavior: Here we use three indicators, VMT, the percentage of workers (16 years and over) commuting by public transit, and the percentage of workers (16 years and over) commuting by walking or biking, to describe the travel behaviors for residents living within each station' neighborhood area. VMT, referring to estimated average daily household miles of vehicle traveled, is collected from the U.S. Department of Transportation 2009 and 2017 National Household Travel Survey. The 2017 survey data was collected between April 2016 and May 2017. The 2009 survey data was collected from March 2008 through May 2009.

Census tract-level percentage of commute by public transit and percentage of commute by walking and biking are obtained from the American Community Survey (ACS 5-year) published in 2009 and 2017 respectively.

4) Demographics and Socioeconomic Factors: ACS also provides population density, employment rates for civilian laborers above 16 years old, and median household income (adjusted by 2017 and 2009 inflation rate respectively) at the census tract level. Summary statistics are shown in Table I.

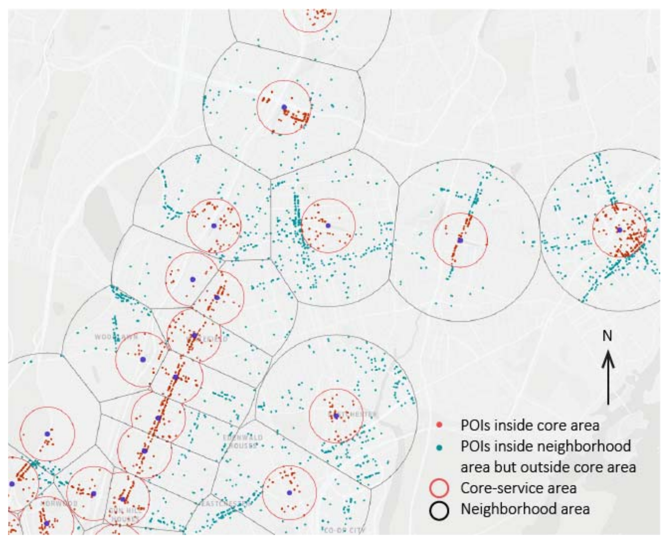


Fig. 3. Red boundaries map the service core area for each station. Red dots are the POIs that fall into each station's core area. Black boundaries map the neighborhood area for each station. Blue dots are the neighborhood POIs.

B. Station Areas Types With Two-Step Clustering

Following the procedures described in the methodology, the results are described below.

1) POI to Station Association: In this experiment, we use 500 meters as the service core area radius r_{core} and 1500 meters as the neighborhood area radius $r_{neighbor}$ considering the typical walking distance and biking distance within 5 minutes, respectively. Figure 3 demonstrates the results graphically. As previous literature used different thresholds to define stations' service areas [19], [40], [46], [47], in the following two-step clustering process, we repeat the entire study with three different r_{core} value: 400m, 500m, and 800m respectively to test the robustness of the station area clustering results (See Appendix B.1).

Authorized licensed use limited to: Zhejiang Lab. Downloaded on April 18,2023 at 07:32:22 UTC from IEEE Xplore. Restrictions apply.

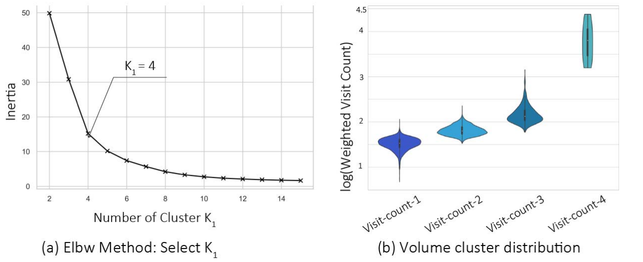


Fig. 4. Visit-count clustering process. (a) Elbow method to identify the optimal K_1 for the visit-count clustering process. (b) Distribution of number of visit in each clusters.

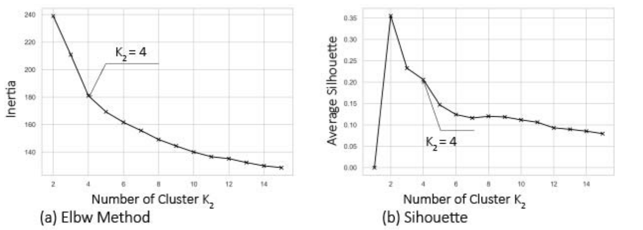


Fig. 5. Pattern clustering process. (a) Elbow method to identify the optimal K_2 for the pattern clustering Process. $K_2 = 4$ is identified as the optimal selection (b) Silhouette method to confirm the K_2 selection.

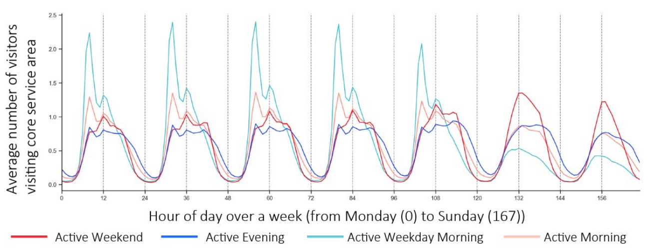


Fig. 6. Count of visits fluctuation of the centers of the three pattern clusters from Monday to Sunday (168 hours). The red line shows stations with more activity during the weekend; the blue line represents stations with daily activities last until later in the night; the light blue line indicates stations with more activities during weekday mornings; the light red line shows stations with subdued activities all week but more activities in the morning.

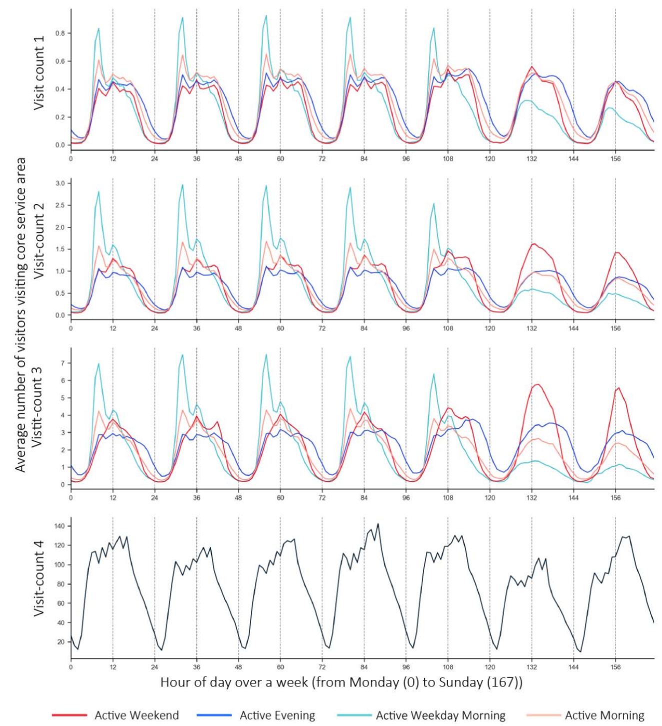


Fig. 7. Count of visits fluctuation of the 13 types of station areas from Monday to Sunday (168 hours).

TABLE II

SUMMARY OF STATION AREA TYPES BY PATTERN AND VOLUME CLUSTERING RESULTS

Visit-count Cluster	Pattern Cluster: Active Weekday Morning	Active Morning	Active Weekend	Active Evening	High Volume	Total
Visit-count 1	410	691	534	726	0	2361
Visit-count 2	235	573	211	533	0	1549
Visit-count 3	93	105	87	79	0	364
Visit-count 4	0	0	0	0	13	13

2) Visit-Count Cluster: First, to classify station areas into different categories, we applied FCM to the visit-count vector V_i to classify the stations into K_1 clusters. The elbow method was used to identify the optimal K_1 as 4 (Figure 4a). The distribution of the number of stations belonging to each cluster is shown in Figure 4b.

3) Pattern Cluster: We then identified the pattern dissimilarity of core-area activities by clustering the pattern vector P_{it}. The dimension of pattern vector P_{it} was reduced using NMF. Then we clustered the pattern vectors into K_2 clusters. Here Figure 5a shows that $K_2 = 4$ provides an optimal solution. As the $Visit-count-4$ only contains 13 out of 4290 stations. They are excluded in the pattern clustering process.

Figure 6 plots the four distinctive pattern clusters:

1) **Active Evening**: station areas with highest number of visit between 6 pm to 2 am.
2) **Active Weekday Morning**: station areas with a high peak of visits in the morning through Monday to Friday.
3) **Active Weekend**: station areas with a higher number of visits during the weekend than weekdays.
4) **Active Morning**: station areas that have the second highest visits during the weekday mornings and consistent high visits during the weekend.

So far, with four visit count clusters and four pattern clusters, we generated 13 distinct station area types through the two-step clustering process. Table II lists the number of stations that belong to each pattern and volume cluster. Figure 7 plots the 13 station area types and their associated activities throughout the week.

C. Travel Behavior

1) Stations Types and Neighborhood Travel Behaviors: As the objectives of TODs involve enhancing vibrancy, increasing transit patronage, and reducing car travel, we use three travel behavior-related variables here to understand the station area types derived from the previous steps. Columns 1, 3 and 5 in

Authorized licensed use limited to: Zhejiang Lab. Downloaded on April 18,2023 at 07:32:22 UTC from IEEE Xplore. Restrictions apply.

TABLE III
MAIN RESULTS: EFFECTS OF STATION TYPES ON TRAVEL BEHAVIORS (2017)

	Log(Walk and Bike)		Log(Public Transit)		Log(VMT)	
	(1)	(2)	(3)	(4)	(5)	(6)
Log(POI Den. Core)	0.213***	0.205***	-0.050***	-0.050***	-0.201***	-0.193***
	(0.009)	(0.009)	(0.008)	(0.009)	(0.009)	(0.009)
Log(POI Den. Nei.)	-0.023***	-0.022***	-0.030***	-0.028***	0.151***	0.147***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Years of Operation	-0.007**	-0.005	0.043***	0.042***	-0.012***	-0.013***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Dist. Downtown	-0.012***	-0.011***	0.000	0.001	0.004***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Visit-count 1 (Lowest)						
× Active Weekday Morning		0.033		0.013		-0.230***
		(0.038)		(0.040)		(0.039)
× Active Weekend		-0.060*		0.015		-0.130***
		(0.035)		(0.037)		(0.036)
× Active Evening		-0.004		0.048		-0.038
		(0.032)		(0.033)		(0.033)
Visit-count 2						
× Active Weekday Morning		0.239***		-0.050		-0.312***
		(0.045)		(0.047)		(0.046)
× Active Weekend		0.045		-0.037		-0.109**
		(0.046)		(0.048)		(0.048)
× Active Evening		0.098**		0.018		-0.255***
		(0.035)		(0.037)		(0.036)
× Active Morning		0.169***		0.052		-0.268***
		(0.033)		(0.035)		(0.034)
Visit-count 3						
× Active Weekday Morning		0.329***		-0.097		-0.266***
		(0.067)		(0.070)		(0.069)
× Active Weekend		0.085		0.150**		-0.374***
		(0.067)		(0.071)		(0.070)
× Active Evening		0.338***		0.031		-0.375***
		(0.070)		(0.074)		(0.073)
× Active Morning		0.387***		-0.088		-0.311***
		(0.062)		(0.065)		(0.064)
Socioeconomic factors controlled	✓	✓	✓	✓	✓	✓
Land uses factors controlled	✓	✓	✓	✓	✓	✓
Observations	4244	4244	4244	4244	4244	4244
R-squared	0.534	0.553	0.533	0.535	0.504	0.522

This table reports the regression coefficients of station types by POI-related activities on travel behaviors. Years of operation here refers to the number of years since the station opened. All stations opened before 2000 are listed as 1999. *High Volume* station areas are not included in the regression as they are all major airport areas that have extreme pattern throughout the day. *Visit-count-1 - Active Morning* station area type is the base for comparison. Standard errors in parentheses.
*** denotes a coefficient significant at the 0.5% level, ** at the 5% level, and * at the 10% level.

in Table III are baseline models. We have both densities of core POIs and neighborhood POIs in the models to control stations' impact areas' land use density. The coefficient between the log-transformed number of core POIs and neighborhood POI is 0.1502, which is below the thresholds of collinearity concern.

Here we observe the following results: 1) With all else equal, station areas with a one percent increase in core POIs' density see a 21.3 percent increase in the ratio of workers commuting by walking or biking. 2) One percent increase in neighborhood POIs' density is associated with a 2 percent decrease in commuting by walking, biking, a 3 percent decrease in public transit, and a 15 percent increase in VMT. 3) One-kilometer increase in the distance from the downtown area is associated with a 1.2 percent decrease in commuting by walking or biking and a 0.4 percent increase in VMT. Thus, the conflicting effects of core POI density and neighborhood POI density indicate that station neighborhoods with more concentrated core areas are correlated with lower dependency on cars for the commute.

Columns 2, 4 and 6 compare the effects of different station types on travel behaviors. With all else controlled, all station areas from the *Visit-count* 2 *and 3* groups have a significant higher ratio of residents commuting by walking and biking, and lower VMT except for the *Active Weekend* station areas. Columns 4 and 6 show that the high-volume *Active Weekend* station areas have a higher ratio of residents commuting by public transit and significantly lower VMT.

2) Home-Oriented and Traveler-Oriented Station Areas: Table IV repeats the model of Equation 4 by differentiating visitors' average home distance. In comparison to Table III, we have the following findings: 1) The high-volume *Active Morning* station areas have a significantly higher ratio of residents commuting by walking and biking and lower VMT. This effect maintains across station areas that are local or traveler-oriented. 2) Column 3 indicates that local-oriented *Visit-count 2* station areas have a higher ratio of residents commuting by public transit, especially for *Active Morning* and *Active Evening* station areas. In contrast, column 2 indicates that traveler-oriented station areas have more association with a higher ratio of residents walking and biking to work.

3) Station Area Establishment and Change of Travel Behavior: Lastly, we select all stations opened between 2009 and 2017 and test how different station areas may affect local travel

Authorized licensed use limited to: Zhejiang Lab. Downloaded on April 18,2023 at 07:32:22 UTC from IEEE Xplore. Restrictions apply.

TABLE IV
EFFECTS OF (LOCAL-/TRAVELER-ORIENTED) STATION TYPES ON TRAVEL BEHAVIOR (2017)

	Log(Walk and Bike)		Log(Public Transit)		Log(VMT)	
	Local (1)	Traveler (2)	Local (3)	Traveler (4)	Local (5)	Traveler (6)
Visit-count 1 (Lowest)						
× Active Weekday Morning	0.024	0.022	0.054	0.043	-0.156***	-0.250***
	(0.053)	(0.055)	(0.048)	(0.063)	(0.041)	(0.065)
× Active Weekend	-0.053	0.018	0.022	-0.051	-0.142***	-0.121*
	(0.045)	(0.056)	(0.040)	(0.064)	(0.034)	(0.066)
× Active Evening	0.004	0.018	0.130***	-0.010	-0.034	-0.047
	(0.040)	(0.051)	(0.037)	(0.059)	(0.031)	(0.061)
Visit-count 2						
× Active Weekday Morning	0.108	0.203***	0.118*	-0.018	-0.266***	-0.266***
	(0.078)	(0.058)	(0.071)	(0.067)	(0.060)	(0.069)
× Active Weekend	-0.064	0.043	0.009	-0.018	-0.024	-0.133*
	(0.072)	(0.062)	(0.065)	(0.071)	(0.055)	(0.074)
× Active Evening	-0.018	0.134**	0.166***	-0.054	-0.244***	-0.215***
	(0.050)	(0.053)	(0.045)	(0.061)	(0.038)	(0.063)
× Active Morning	0.014	0.193***	0.157***	0.079	-0.180***	-0.299***
	(0.048)	(0.049)	(0.043)	(0.057)	(0.036)	(0.058)
Visit-count 3						
× Active Weekday Morning	0.149	0.234***	-0.316*	0.032	0.153	-0.276***
	(0.183)	(0.077)	(0.166)	(0.088)	(0.140)	(0.091)
× Active Weekend	-0.032	0.059	0.105	0.251**	-0.158	-0.398***
	(0.131)	(0.082)	(0.118)	(0.094)	(0.100)	(0.097)
× Active Evening	-0.076	0.331***	0.151	0.092	-0.144	-0.344***
	(0.146)	(0.084)	(0.132)	(0.096)	(0.112)	(0.100)
× Active Morning	0.439***	0.269***	0.178*	-0.020	-0.276***	-0.242**
	(0.116)	(0.076)	(0.105)	(0.087)	(0.089)	(0.090)
Socioeconomic factors controlled	✓	✓	✓	✓	✓	✓
Land uses factors controlled	✓	✓	✓	✓	✓	✓
Observations	2121	2120	2121	2120	2121	2120
R-squared	0.399	0.625	0.626	0.436	0.608	0.515

This table reports the regression coefficients of station types by POI-related activities on travel behaviors. Standard errors in parentheses. Years of operation and distance from downtown controlled.
*** denotes a coefficient significant at the 0.5% level, ** at the 5% level, and * at the 10% level.

behaviors. Table V reports the results. Consistent with results in Table III, column 1 in Table V shows that *Visit-count 2* station areas see a significant increase in ratio of residents commute by walking and biking. Column 2 shows that *Active Weekend* station areas with about the medium number of visits per POI (*Visit-count 2*) have seen a significant increase in the ratio of residents taking public transit. Given the small sample size, a sensitivity analysis for adjusting the selection of stations is presented in the Appendix. After the sensitivity analysis, the *Visit-count 1 - Active Weekday Morning* station areas are no longer associated with decreased public transit.

V. DISCUSSION

A. Definition of "Balanced" Station Areas Should Consider the Temporal Dimension

Previous literature has argued for the importance of creating integrated, balanced, and transit-oriented sites [15], [25], [47], [48] so that a station area can attract sufficient patronage on the one hand and support the local community on the other. However, these studies have not addressed if the "balanced" stations measured by static features such as land use and transportation network may have different impacts on the nearby communities or exhibit different temporal characteristics. By incorporating the hourly temporal visitor activities in the station area, we show that even among station areas that have a significantly lower VMT, the high-volume *Active Morning* station areas, characterized by subtle morning peak during the weekdays and consistent visiting volumes during the weekend, are the most successful in encouraging the residents in the local communities to walk or bike to work. Among all the stations built between 2009 to 2017, those *Active Weekend* station areas with a medium number of weekly visits have seen increases in the ratio of commuting by walking, biking, and public transit. Although these stations opened for the sake of connecting with major recreation sites such as parks, stadiums, and shopping malls, accompanied by adjacent developments that are well-connected by public transit, there has been an increase in the use of active transport and transit for commuting (See Appendix Figure A2 for examples).

In addition, in the results, stations opened earlier have seen their neighborhood households with fewer VMT and a higher proportion of commute with public transit. This result implies the time required to establish a vibrant urban environment and change people's travel habits.

B. Spatial Distribution of the Temporal Pattern

We mapped the geographic distribution of the 13 types of station areas in six selected cities (Figure 8): San Francisco, Portland, Chicago, Dallas, New York City, and Philadelphia.

Authorized licensed use limited to: Zhejiang Lab. Downloaded on April 18,2023 at 07:32:22 UTC from IEEE Xplore. Restrictions apply.

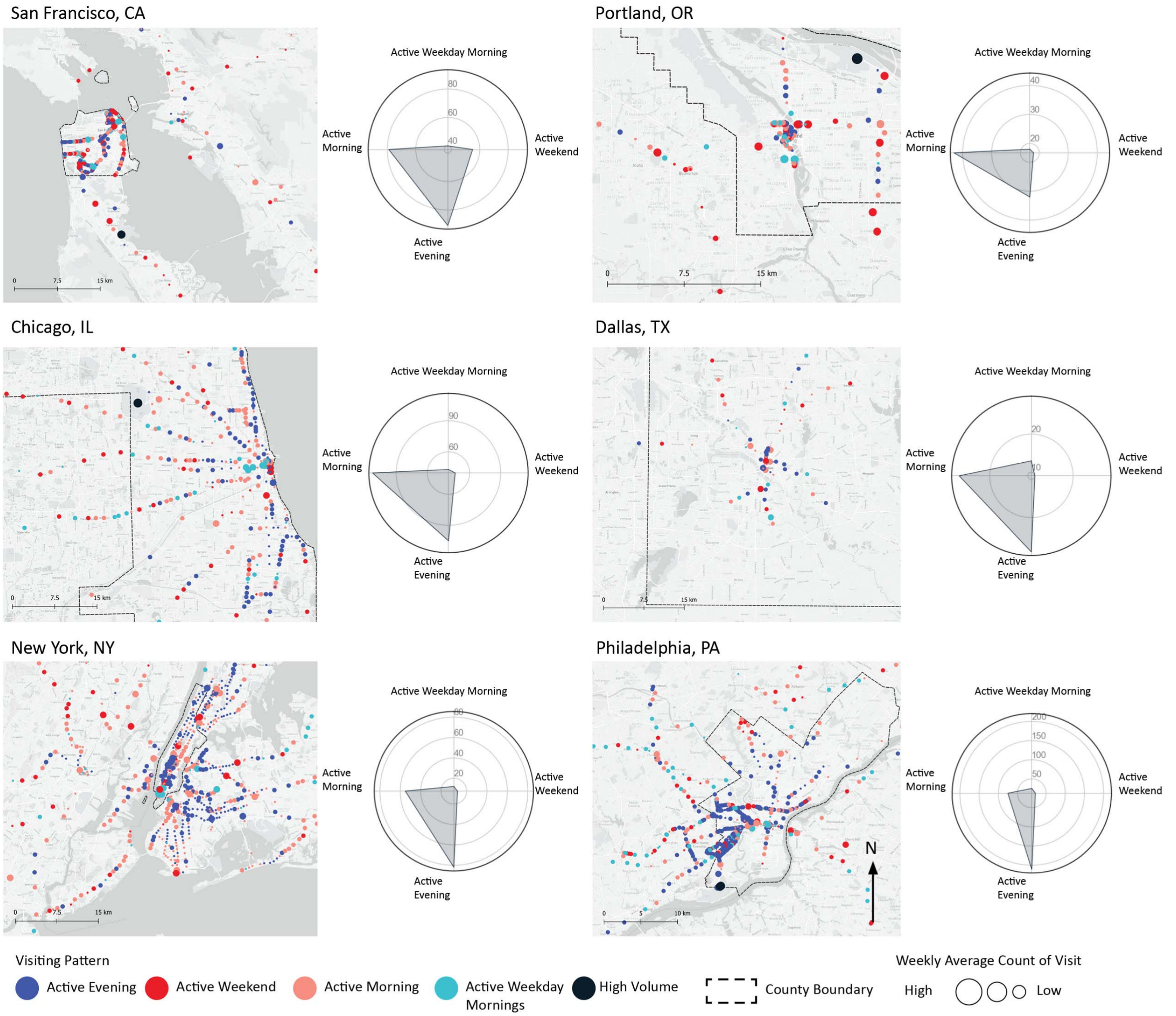


Fig. 8. Geographical distribution of station types by temporal pattern. We selected six metropolitan areas and their associated county to present station types' geographical distribution. The color of the dots represent the station types, and the radar charts show the number of pattern cluster within each county.

We list three main observations here: 1) The pattern distribution resonates with different urban structures. For example, we observe high volume and *Active Weekday Morning* type of station areas in the downtown area except for Dallas, which indicates intense work-only-related activities. Chicago has *Active Weekend* type of stations both in the suburban area and across the city center. They represent two recreational functions: the city center has a series of museums, theaters, and parks connecting as a cultural-recreation band, while the suburban area offers more natural recreation activities. University clusters, major sight-seeing locations, and transfer stations are found within high-volume *Active Mornings* station areas, as activity patterns in these areas do not have dramatic differences between weekdays and weekends. 2) People with different backgrounds and social statuses have different travel patterns and might demand transit service at different times. New York City has *Active Evening* stations clustered in the Bronx and Queens County, and Philadelphia has a similar pattern in the West, where most residents living in these areas provide services to the region - they have longer working hours as well as more late night commutes.

C. From Spatial Mixed-Use to Temporal Mixed-Use

Researchers and planners have extensive studies on the benefit of developing spatially and functionally mixed-use neighborhoods. Our study shows that with land-use density and spatial characteristics considered, the "mixed-use" across time still contributes to the goals of TOD. However, how to build on the current TOD models to include the temporal measure of urban vibrancy is a question for further study.

Authorized licensed use limited to: Zhejiang Lab. Downloaded on April 18,2023 at 07:32:22 UTC from IEEE Xplore. Restrictions apply.

TABLE V
EFFECTS OF STATION TYPES ON CHANGE OF TRAVEL BEHAVIORS

	Changes of Travel Behavior 2017 - 2009		
	Log(Walk or Bike) (1)	Log(Public Transit) (2)	Log(VMT) (3)
Visit-count 1 (Lowest)			
× Active Weekday Morning	0.163	-0.309**	0.056
	(0.131)	(0.126)	(0.099)
× Active Weekend	-0.008	-0.153	-0.011
	(0.151)	(0.147)	(0.114)
× Active Evening	0.114	0.003	0.190**
	(0.115)	(0.111)	(0.091)
Visit-count 2			
× Active Weekday Morning	0.288*	-0.081	0.215*
	(0.156)	(0.151)	(0.118)
× Active Weekend	0.277*	0.348**	0.049
	(0.164)	(0.159)	(0.124)
× Active Evening	0.376**	-0.060	0.099
	(0.142)	(0.137)	(0.108)
× Active Morning	0.243*	-0.079	0.019
	(0.137)	(0.133)	(0.104)
Visit-count 3			
× Active Weekday Morning	0.314	0.243	0.135
	(0.247)	(0.239)	(0.187)
× Active Weekend	0.296*	-0.241	-0.016
	(0.174)	(0.169)	(0.132)
× Active Evening	0.192	-0.090	-0.065
	(0.201)	(0.193)	(0.150)
× Active Morning	0.254	0.189	-0.066
	(0.303)	(0.295)	(0.229)
Δ Socioeconomic Factors	✓	✓	✓
Observations	134	134	134
R-squared	0.070	0.288	0.631

This table reports the regression coefficients of station types by POI-related activities on changes of travel behaviors. Years of operation and distance from nearest downtown areas are controlled. Standard errors in parentheses. Only stations that are opened after 2009 are included in the model.
*** denotes a coefficient significant at the 0.5% level, ** at the 5% level, and * at the 10% level.

VI. CONCLUSION

Despite the rich literature on TOD classification methods, few studies have examined TOD's temporal activities or studied the relationship between the visitors' visiting patterns and neighborhoods' travel behavior. This paper leverages the revealed visiting pattern of associated POIs within 4,290 fixed guideway station areas across 54 metropolitan areas in the U.S. to propose an activity-based station classification framework. Using the aggregated count of visits and visiting patterns of a week, we classify all stations in this study into 13 distinct station area types. This measure captures the dynamics of station areas across different times of day and week.

Researchers from the field of transport geography and urban planning have articulated the importance of mixed land uses [22], [23] and employment density [32] in curating a successful TOD. In relation, we demonstrated that among stations built between 2009 to 2017, those with specific weekend activities have encouraged more increased walk, bike, and public transit trips to work from 2009 to 2017. Thus, studying the impacts of different activity-based station types on household behavior changes is crucial for understanding how much potential impact we could further exert by developing station areas that encourage more non-routine travels.

In addition, we revisited the community concept of TOD and conducted a further experiment to understand the station types on TODs with different visitors' compositions. Our models illustrate that local-oriented station areas and traveler-oriented station areas could have lower VMT, but they support working trips by public transit and walking or biking differently.

We believe this work is significant to urban policymakers in evaluating TODs across different regions and transport researchers in throwing light on how recently available large-scale fine-grained mobility data could offer further insights in developing important planning concepts.

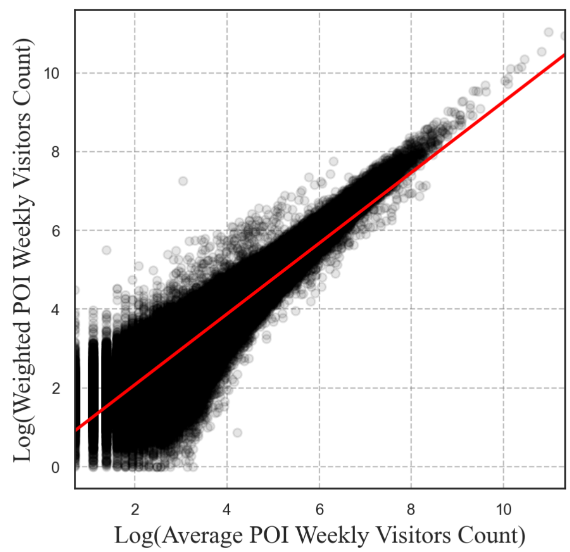


Fig. 9. Scatter plot of relationship between weighted visitors and non-weighted visitors.

APPENDIX

A. Representatives of the Visit Data

The mobile phone data, although having a relative large sample size, is still a sample from the true population. So we use a weighting mechanism (post-stratification) based on the ratio of smartphone users to the true population in the census block group to adjust the number of visitors that visit each POI. We define the weight for individual user i to be:

$$\lambda_i = \frac{\tilde{N}_{b_i}}{\sum_i^J \tilde{N}_{b_i}} \times \frac{\sum_i^J N_{b_i}}{N_{b_i}}, \tag{A.1}$$

where N_{b_i} is the population of census block group b and $\tilde{N}_{b_i}$ is the number of users in our sample with home locations in block group b. J is the total number of users that are in the sample. Then we adjust the number of visitors that visit a POI by multiplying the weight λ_i. Here we plot the relationship between the *Log(Visitors before re-weighting)* and the *Log(Visitors after re-weighting)* in Figure 9 The correlation between the two is around 0.97.

B. Sensitivity Analysis

1) Station Area Radius: We conduct a sensitive analysis to study the effects of different station area radius on the clustering results. We consider 400 meter, 500 meter, and

Authorized licensed use limited to: Zhejiang Lab. Downloaded on April 18,2023 at 07:32:22 UTC from IEEE Xplore. Restrictions apply.

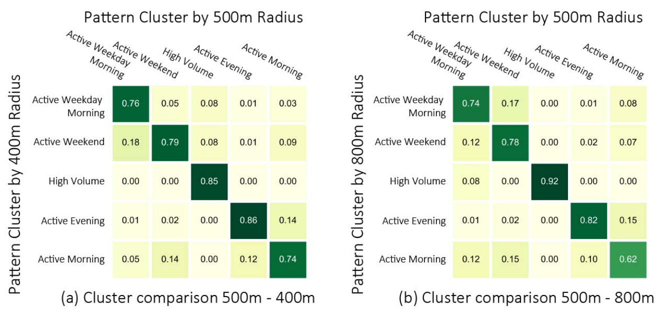


Fig. A1. Percentage of station areas remained with the original cluster analysis after changing the upper-limit of core service area radius from 500m to 400m (a) and 800m (b) respectively.

TABLE A1

SENSITIVITY TEST: EFFECTS OF STATION TYPES ON CHANGE OF TRAVEL BEHAVIORS

	Change of Travel Behavior 2017 - 2009		
	Log(Walk or Bike) (1)	Log(Public Transit) (2)	Log(VMT) (3)
Dist. Downtown	0.003	0.025***	0.004
	(0.005)	(0.004)	(0.003)
Years of Operation	0.317**	0.049	-0.033
	(0.140)	(0.130)	(0.094)
Δ Log(Pop.Den.)	-0.012	-0.487	0.175
	(0.327)	(0.297)	(0.216)
Δ Log(Income)	-0.383	-0.778**	0.269
	(0.290)	(0.279)	(0.197)
Visit-count 1 (Lowest)			
× Active Weekday Morning	0.152	-0.239	0.168
	(0.168)	(0.158)	(0.114)
× Active Weekend	0.166	-0.126	0.017
	(0.199)	(0.190)	(0.136)
× Active Evening	0.260	-0.015	0.251**
	(0.158)	(0.149)	(0.108)
Visit-count 2			
× Active Weekday Morning	0.470**	-0.249	0.005
	(0.229)	(0.219)	(0.156)
× Active Weekend	0.275	0.438**	0.053
	(0.216)	(0.203)	(0.147)
× Active Evening	0.431**	0.067	0.139
	(0.183)	(0.171)	(0.124)
× Active Morning	0.496**	-0.123	0.064
	(0.216)	(0.200)	(0.144)
Visit-count 3			
× Active Weekday Morning	0.559*	0.486	0.343
	(0.327)	(0.309)	(0.224)
× Active Weekend	0.634	-0.373	0.201
	(0.437)	(0.409)	(0.311)
× Active Evening	0.131	0.093	-0.099
	(0.248)	(0.229)	(0.167)
× Active Morning	0.265	0.309	-0.063
	(0.327)	(0.309)	(0.225)
Observations	74	74	74
R-squared	0.238	0.591	0.209

This table reports the regression coefficients of station types by POI-related activities on changes of travel behaviors. Years of operation here refers to the number of years since the station opened. All stations opened before 2000 are listed as 1999. Standard errors in parentheses. Only stations that are opened after 2010 are included in the model. *** denotes a coefficient significant at the 0.5% level, ** at the 5% level, and * at the 10% level.

800 meters all valid definitions as the station core area radius r_{core}. Then we repeat the two-step clustering process to create station types by the three radius respectively. Figure A1 shows the results for the sensitivity analysis when we adjust the r_{core}. Over 74% station areas stay in the same pattern cluster group when we adjust the radius from 500 meter to 400 meter. Over

Fig. A2. Examples of *Active Weekend* station areas that were opened after 2009.

70% station areas stay in the same pattern cluster group when the radius is adjusted from 500 meter to 800 meter with an exception of the *Active Morning* type.

2) Definition of New Stations: Given the fact that the first survey in our study was conducted from 2008 to 2009, there might be some stations that were open to operate in 2009 but its impact took place before that. Thus we repeat the model of Equation 5 by removing the stations that are opened in 2009, and report the results in Table A1.

C. Station Neighborhood Area Variables Calculation

Station neighborhood area's demographic, socioeconomic and travel behavior variables are weighted by the fraction of census tract that intersects a stations' neighborhood area. Use population and employment rate as an example, for each station's neighborhood $Neighborhood_i$, we obtain a collection of census tract intersections $C_i = \{c_1, c_2, c_3, \ldots, c_l, \ldots, c_L\}$. L indicates the total number of census tracts intersected with $Neighborhood_i$. For each census tract c_l that intersects with a station neighborhood $Neighborhood_i$, we obtain the A_l, which is the original land area of each census tract c_l, a_{il}, which indicates the intersection area of c_l with $Neighborhood_i$, and $NeighborhoodA_i$, which is the land area of each station neighborhood. We then calculate a proportion $p_{il} = a_{il}/A_l$ and weight $w_{il} = a_{il}/NeighborhoodA_i$. Then we derive population and employment rate of each

Authorized licensed use limited to: Zhejiang Lab. Downloaded on April 18,2023 at 07:32:22 UTC from IEEE Xplore. Restrictions apply.

TABLE A2
POI SUMMARY BY NAICS CODE

Two-Digit NAICS	Industry	Number of POIs (Neighbor Area)	Number of POIs (Core Area)
11	Agriculture	10	19
21	Mining	12	1
22	Utilities	221	66
23	Construction	662	352
31	Manufacturing	2947	2568
32	Manufacturing	563	510
33	Manufacturing	833	485
42	Wholesale Trade	2498	1057
44	Retail Trade	67372	47716
45	Retail Trade	23846	17701
48	Transportation	2052	1865
49	Transportation	2150	2210
51	Information	5304	4916
52	Finance	21530	17582
53	Real Estate	7902	5039
54	Professional	6206	4789
55	Management	5	3
56	Administrative	3575	2123
61	Educational Services	13451	8730
62	Health Care	72914	48314
71	ArtsRecreation	22521	15058
72	Food Services	64746	68997
81	Other Services	65816	44175
92	Public Administration	2843	2351
	Not Known	8268	7041
Total		398247	303668

station neighborhood as:

$$Population_i = \sum_{l}^{L} Population_l * p_{il}$$
$$Employment_i = \frac{\sum_l^L Employment_l * w_{il}}{\sum_l^L w_{il}}, \quad \text{(A.2)}$$

D. Examples of Stations Built Between 2009 to 2017

To confirm the findings, we select three examples of *Active Weekend* stations opened after 2009 to demonstrate the changes (Figure A2).

E. POI Summary

Breakdown of POI summary by types are shown in Table A2.

REFERENCES

[1] J. Jacobson and A. Forsyth, "Seven American TODs: Good practices for urban design in transit-oriented development projects," *J. Transp. Land Use*, vol. 1, no. 2, pp. 51–88, Nov. 2008.

[2] J. Zhou *et al.*, "Can TODness improve (expected) performances of TODs? An exploration facilitated by non-traditional data," *Transp. Res. D, Transp. Environ.*, vol. 74, pp. 28–47, 2019.

[3] A. Ibraeva, G. H. de Almeida Correia, C. Silva, and A. P. Antunesa, "Transit-oriented development: A review of research achievements and challenges," *Transp. Res. A, Policy Pract.*, vol. 132, pp. 110–130, Feb. 2020.

[4] M. Austin *et al.*, "Performance-based transit-oriented development typology guidebook," Transp. Res. Board, Tech. Rep., 2010. [Online]. Available: http://reconnectingamerica.org/assets/Uploads/2010_performance basedtodtypologyguidebook.pdf

[5] S. Zemp, M. Stauffacher, D. J. Lang, and R. W. Scholz, "Classifying railway stations for strategic transport and land use planning: Context matters!" *J. Transp. Geogr.*, vol. 19, no. 4, pp. 670–679, Jul. 2011.

[6] B. P. Y. Loo, A. H. Cheng, and S. L. Nichols, "Transit-oriented development on greenfield versus infill sites: Some lessons from Hong Kong," *Landscape Urban Planning*, vol. 167, pp. 37–48, Nov. 2017. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S01692 04617301172

[7] R. Cervero and K. Kockelman, "Travel demand and the 3Ds: Density, diversity, and design," *Transp. Res. D, Transp. Environ.*, vol. 2, no. 3, pp. 199–219, Sep. 1997.

[8] M. A. Berawi, G. Saroji, F. A. Iskandar, B. E. Ibrahim, P. Miraj, and M. Sari, "Optimizing land use allocation of transit-oriented development (TOD) to generate maximum ridership," *Sustainability*, vol. 12, no. 9, p. 20, 2020.

[9] C. Ratti, A. Sevtsuk, S. Huang, and R. Pailer, "Mobile landscapes: Graz in real time," in *Location Based Services and TeleCartography*. Springer, 2007, pp. 433–444.

[10] F. Calabrese, M. Colonna, P. Lovisolo, D. Parata, and C. Ratti, "Real-time urban monitoring using cell phones: A case study in Rome," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 1, pp. 141–151, Mar. 2011.

[11] T. Pei, S. Sobolevsky, C. Ratti, S.-L. Shaw, T. Li, and C. Zhou, "A new insight into land use classification based on aggregated mobile phone data," *Int. J. Geograph. Inf. Sci.*, vol. 28, no. 9, pp. 1988–2007, 2014.

[12] S. Phithakkitnukoon, T. Sukhvibul, M. Demissie, Z. Smoreda, J. Natwichai, and C. Bento, "Inferring social influence in transport mode choice using mobile phone data," *EPJ Data Sci.*, vol. 6, no. 1, pp. 1–29, Dec. 2017.

[13] H. M. Taki *et al.*, "Planning TOD with land use and transport integration: A review," *J. Geosci., Eng., Environ., Technol.*, vol. 2, no. 1, pp. 84–94, 2017.

[14] H. Dittmar and G. Ohland, *The New Transit Town: Best Practices in Transit-Oriented Development*. Washington, DC, USA: Island Press, 2012.

[15] L. Bertolini, "Spatial development patterns and public transport: The application of an analytical model in The Netherlands," *Planning Pract. Res.*, vol. 14, no. 2, pp. 199–210, May 1999.

[16] D. Olaru *et al.*, "Place vs. node transit: Planning policies revisited," *Sustainability*, vol. 11, no. 2, p. 477, 2019.

[17] D. E. Reusser, P. Loukopoulos, M. Stauffacher, and R. W. Scholz, "Classifying railway stations for sustainable transitions–balancing node and place functions," *J. Transp. Geogr.*, vol. 16, no. 3, pp. 191–202, May 2008.

[18] C. D. Higgins and P. S. Kanaroglou, "A latent class method for classifying and evaluating the performance of station area transit-oriented development in the Toronto region," *J. Transp. Geogr.*, vol. 52, pp. 61–72, Apr. 2016. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S096669231600034X

[19] B. P. Y. Loo, C. Chen, and E. T. H. Chan, "Rail-based transit-oriented development: Lessons from New York city and Hong Kong," *Landscape Urban Planning*, vol. 97, no. 3, pp. 202–212, Sep. 2010. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0169204610001283

[20] J. Zhou, Y. Yang, and C. Webster, "Using big and open data to analyze transit-oriented development: New outcomes and improved attributes," *J. Amer. Planning Assoc.*, vol. 86, no. 3, pp. 364–376, 2020.

[21] G. S. Rutherford, E. McCormack, and M. Wilkinson, "Travel impacts of urban form: Implications from an analysis of two Seattle area travel diaries," in *Proc. Urban Design, Telecommuting Travel Forecasting Conf.*, Transportation Research Board, Oct. 1996. [Online]. Available: http://www.mendeley.com/catalog/travel-impacts-urban-form-implications-analysis-two-seattle-area-travel-diaries/

[22] R. Cervero, "Built environments and mode choice: Toward a normative framework," *Transp. Res. D, Transp. Environ.*, vol. 7, no. 4, pp. 265–284, Jun. 2002.

[23] P. Calthorpe, W. Fulton, and W. B. Fulton, *The Regional City*. Washington, DC, USA: Island Press, 2001.

[24] R. Cervero, "Transit-based housing in california: Evidence on ridership impacts," *Transp. Policy*, vol. 1, no. 3, pp. 174–183, Jun. 1994.

[25] D. Belzer, S. Srivastava, J. Wood, and E. Greenberg, "Transit-oriented development (TOD) and employment," Transp. Res. Board, Tech. Rep., 2011. [Online]. Available: http://reconnectingamerica.org/assets/Uploads/TODandEmploymentFINALMay2011.pdf

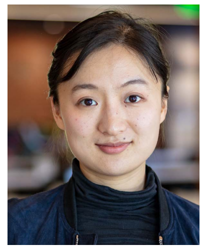
[26] S. Handy, "Smart growth and the transportation-land use connection: What does the research tell us?" *Int. Regional Sci. Rev.*, vol. 28, no. 2, pp. 146–167, 2005.

[27] M. Kamruzzaman, D. Baker, S. Washington, and G. Turrell, "Advance transit oriented development typology: Case study in Brisbane, Australia," *J. Transp. Geogr.*, vol. 34, pp. 54–70, Jan. 2014.

[28] J. Niles and D. Nelson, "Measuring the success of transit-oriented development: Retail market dynamics and other key determinants," in *Proc. APA Nat. Planning Conf.*, vol. 28, Seattle, WA, USA, 1999, p. 2003.

Authorized licensed use limited to: Zhejiang Lab. Downloaded on April 18,2023 at 07:32:22 UTC from IEEE Xplore. Restrictions apply.

[29] L. D. Frank and G. Pivo, "Impacts of mixed use and density on utilization of three modes of travel: Single-occupant vehicle, transit, and walking," *Transp. Res. Rec.*, vol. 1466, pp. 44–52, Jan. 1994.
[30] D. M. Levinson and A. Kumar, "Density and the journey to work," *Growth Change*, vol. 28, no. 2, pp. 147–172, Mar. 1997.
[31] X. Weizhou, W. Shusheng, and H. Fumin, "Influence of land use variation along rail transit lines on the ridership demand," in *Proc. 2nd Int. Conf. Intell. Comput. Technol. Autom.*, vol. 3, 2009, pp. 633–637.
[32] H. Pan, J. Li, Q. Shen, and C. Shi, "What determines rail transit passenger volume? Implications for transit oriented development planning," *Transp. Res. D, Transp. Environ.*, vol. 57, pp. 52–63, Dec. 2017. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1361920917301189
[33] Z. Cai, G. Sun, X. Su, T. Li, L. Guo, and Z. Ding, "Analysis of land use characteristics around urban rail transit stations," *IEEE Trans. Intell. Transp. Syst.*, early access, May 6, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9088211/, doi: 10.1109/TITS.2020.2989811.
[34] N. L. Stiffler, "The effect of transit-oriented development on vehicle miles traveled: A comparison of a TOD versus a non-TOD neighborhood in Carlsbad, CA," California Polytech. State Univ., San Luis Obispo, CA, USA, Tech. Rep., 2011.
[35] T. C. Schelling, "Models of segregation," *Amer. Econ. Rev.*, vol. 59, no. 2, pp. 488–493, 1969.
[36] R. Crane, "The influence of urban form on travel: An interpretive review," *J. Planning Literature*, vol. 15, no. 1, pp. 3–23, Aug. 2000.
[37] P. L. Mokhtarian and D. Van Herick, "Viewpoint: Quantifying residential self-selection effects: A review of methods and findings from applications of propensity score and sample selection approaches," *J. Transp. Land Use*, vol. 9, no. 1, pp. 9–28, Apr. 2016.
[38] M. Kamruzzaman, L. Wood, J. Hine, G. Currie, B. Giles-Corti, and G. Turrell, "Patterns of social capital associated with transit oriented development," *J. Transp. Geogr.*, vol. 35, pp. 144–155, Feb. 2014.
[39] X. Kong, F. Xia, K. Ma, J. Li, and Q. Yang, "Discovering transit-oriented development regions of megacities using heterogeneous urban data," *IEEE Trans. Computat. Social Syst.*, vol. 6, no. 5, pp. 943–955, Oct. 2019.
[40] J. Schuetz, G. Giuliano, and E. J. Shin, "Can a car-centric city become transit oriented? Evidence from Los Angeles," *Cityscape*, vol. 20, no. 1, pp. 167–190, 2018.
[41] J. C. Bezdek, W. Full, and R. Ehrlich, "FCM: The fuzzy c-means clustering algorithm," *Comput. Geosci.*, vol. 10, nos. 2–3, pp. 191–203, 1984.
[42] N. R. Pal and J. C. Bezdek, "On cluster validity for the fuzzy c-means model," *IEEE Trans. Fuzzy Syst.*, vol. 3, no. 3, pp. 370–379, Aug. 1995.
[43] T. M. Kodinariya and P. R. Makwana, "Review on determining number of cluster in K-means clustering," *Int. J.*, vol. 1, no. 6, pp. 90–95, 2013.
[44] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 13, 2001, pp. 556–562.
[45] E. W. Burgess, *The Growth of the City: An Introduction to a Research Project*. London, U.K.: Routledge, 2015.
[46] E. Guerra and R. Cervero, "Is a half-mile circle the right standard for TODs?" UC Berkeley, Berkeley, CA, USA, Tech. Rep., Apr. 2013. [Online]. Available: https://escholarship.org/uc/item/0m47w77k
[47] D. S. Vale, "Transit-oriented development, integration of land use and transport, and pedestrian accessibility: Combining node-place model with pedestrian shed ratio to evaluate and classify station areas in Lisbon," *J. Transp. Geogr.*, vol. 45, pp. 70–80, May 2015.
[48] P. Chorus and L. Bertolini, "An application of the node place model to explore the spatial development dynamics of station areas in Tokyo," *J. Transp. Land Use*, vol. 4, no. 1, pp. 45–58, 2011.



Zhuangyuan Fan received the master's degree in landscape architecture from the University of Pennsylvania in 2015 and the Master of City Planning, concentration in Urban Information System (UIS), degree from Massachusetts Institute of Technology (MIT) in 2020. She is currently pursuing the Ph.D. degree with the Department of Geography, The University of Hong Kong. She is also a Research Fellow with the Senseable City Laboratory, MIT.

Fan Zhang received the B.S. degree from Beijing Normal University, Zhuhai, in 2012, and the M.Sc. and Ph.D. degrees from The Chinese University of Hong Kong in 2013 and 2017, respectively. He is currently a Senior Post-Doctoral Associate with the Senseable City Laboratory, Massachusetts Institute of Technology. His research interests include spatio-temporal data mining, urban visual intelligence, and data-driven approaches for urban studies.

Becky P. Y. Loo is currently a Professor with the Department of Geography, The University of Hong Kong. She is also the Director of the Institute of Transport Studies and the Founding Co-Director of the Joint Laboratory on Future Cities. Her main research interests are sustainable transportation, future cities, and smart technologies.

Authorized licensed use limited to: Zhejiang Lab. Downloaded on April 18,2023 at 07:32:22 UTC from IEEE Xplore. Restrictions apply.