

Similarity Analysis of Spatial-Temporal Mobility Patterns for Travel Mode Prediction Using Twitter Data

Zhenhao Cao^a, Zhenyu Shou^b, Xuan Di^{b,c,*}

^a*School of Electronic Information and Electric Engineering, Shanghai Jiao Tong University*

^b*Department of Civil Engineering and Engineering Mechanics, Columbia University*

^c*Data Science Institute, Columbia University*

Abstract

Leveraging the crawled geotagged and timestamped tweets of Twitter users, this study develops a methodological framework to predict massively unreported travel mode choices of Twitter users who have left geotagged and timestamped tweets. The prediction framework is based on the similarity between a user who has not reported her mode choice and the users with known travel modes. To appropriately represent a Twitter user's data, we employ a discretized spatial-temporal probabilistic distribution to characterize the user. A novel convolution-based similarity measure is then proposed to effectively capture the interdependencies of both spatially and temporally adjacent data points. A graph inference model is further established to explore the predictability of people's travel mode choice. To validate the prediction framework, we use the Proposition 1 incident in Austin, TX in 2016 as a case study and leverage relevant data crawled from Twitter. The prediction results validate the effectiveness of both the convolution-based similarity measure and the prediction framework. This work demonstrates the feasibility of using social media data to predict people's mobility choices.

Keywords: Similarity Analysis, Graph Inference Model, Travel Mode Choice, Social Media

1. Introduction

In May 2016, Uber and Lyft pulled out of Austin, TX, as a result of the failure on Proposition 1 (Hampshire et al., 2017; Di et al., 2019), an ordinance defending ridehailing companies' self-regulation. By tracing the historical tweets published on Twitter around May 2016 in Austin, we notice that the proposal of Proposition 1, its defeat, the subsequent suspension of Uber and Lyft, and the launch of several emerging alternative services together triggered a constantly hot discussion on Twitter. A lot of Twitter users reported their opinions, experiences, and travel mode use or switch via tweets. These tweets, which can be crawled from Twitter through its official API, contain geotags and timestamps as well as their new mode choices representing Twitter users' mobility patterns. Leveraging these rich mobility information from tweets, this paper aims to explore the underlying correlation between Twitter users' mobility patterns and their mode choices. A methodological framework of similarity analysis and graph inference models is developed to predict the massively unreported mode choices of those who have left geotagged and timestamped contexts on Twitter. Because social media data can be obtained at low or even no cost, this work can demonstrate a great potential of using social media data for urban mobility pattern analysis and could potentially benefit city planners, policy makers, and commercial corporations.

Travel mode choice prediction has been extensively studied using various discrete choice models, describing individuals decision-making attributable to observable covariates and psychometric measures related to individual attitudes and perceptions (Bhat, 1995; McFadden and Train, 2000; Morikawa et al., 2002; Bhat

*Corresponding author. Tel.: +1 212 853 0435;
Email address: sharon.di@columbia.edu (Xuan Di)

and Dubey, 2014). A few recent studies developed multi-dimensional models to capture people's mode choice in the presence of ride-hailing (Lavieri and Bhat, 2019a,b). Cantillo et al. (2006, 2007) incorporated thresholds to the discrete choice model to capture people's habit inertia when some attributes in a travel mode are changed. Di et al. (2019) exemplified the existence of such inertia in travel mode choice behavior using survey data collected in Austin, TX after the service disruption of Uber and Lyft.

The aforementioned research on mode choice modeling requires data collection, either actively or passively. Most studies relied on survey data to collect respondents' travel choices, which could take a relatively long time and might receive a relatively small pool of responses. Recent years see a growing trend of utilizing location-aware mobile technologies (Xiao and Lo, 2016) to understand travel patterns at both population and individual scales, using GPS trajectories (Krause and Zhang, 2018), connected vehicle basic safety messages (Shou and Di, 2018), call data records (CDRs) (Candia et al., 2008; Wesolowski et al., 2013), smart card data (Ma et al., 2013; Sun et al., 2012), and even credit card transactions (Wu et al., 2005). These types of data are *ubiquitous data*, which conveys the idea of "being everywhere" (Calastri et al., 2017; Goulet-Langlois et al., 2018). Ubiquitous data establishes the knowledge base of people's trip diary in a more cost-effective but reliable way compared with traditional survey, and allow for more flexible data-driven analysis. One popular ubiquitous data is the social media data (Hasnat and Hasan, 2018; Hu and Jin, 2017; Kurkcu et al., 2016), which is gaining its momentum due to its easy accessibility and low cost. Readers interested in different applications of social media data in transportation are referred to Rashidi et al. (2017) for an overview.

Similarity analysis has been widely used to classify travelers, assuming that some characteristics of individuals such as their demographic features or travel mode choices are correlated with their travel patterns (Joh et al., 2002). Various similarity measures are defined based on different characterizations of travel patterns. The travel patterns can be represented as activity sequences (Goulias, 1999; Recker, 1995; Ying et al., 2010; Allahviranloo and Recker, 2013), frequent patterns (Zhao et al., 2018; Goulias, 1999; Shou and Di, 2018), spatial-temporal matrices (Joh et al., 2002) or tensors (Sun and Axhausen, 2016). The sequence alignment (Needleman and Wunsch, 1970; Wilson, 1998) is one of the most commonly adopted similarity measure, in which the distance between two travel sequences is defined as the smallest number of operations required to equalize them. Based on an extended comparison table method (Kruskal, 1983), Joh et al. (2002) proposed a multidimensional alignment approach to incorporate interdependencies between the dimensions or facets underlying travel patterns. Other similarity measures of two travel patterns include Euclidean distance (Joh et al., 2001; Jiang et al., 2013), cosine similarity (Toole Jameson L. et al., 2015), and agenda dissimilarity (Allahviranloo et al., 2014). The aforementioned similarity measures are usually performed on a pair of travelers' single-day activity sequences. When multi-person multi-day activity sequences are available, similarity measures are usually defined on frequent patterns (Ying et al., 2010; Shou and Di, 2018). Frequent patterns are those frequently occurring ordered subsequences, and can be mined by performing some frequent pattern mining algorithms such as Prefix-Span (Pei et al., 2001). However, performing spatial-temporal data mining while maintaining both the spatial and temporal information in frequent travel patterns can be challenging (Atluri et al., 2018).

To overcome such challenge, in this paper, we will first develop a matrix-based representation to characterize travelers' historical activity data in order to preserve both their spatial and temporal mobility information. Then we propose a convolution-based similarity measure to appropriately estimate the similarity of two both spatially and temporally adjacent data points. The major contributions of this paper are as follows:

1. Rather than conducting a survey or using some location-aware technologies to collect people's travel history, which are quite costly, we will utilize geotagged and timestamped tweets to predict people's travel mode choices in response to the service disruption of Uber and Lyft in Austin, TX.
2. A convolution-based similarity measure is proposed to calculate the spatial-temporal similarity between every two users based on a spatial-temporal matrix representation. The concept of convolution is borrowed from image processing, aiming to capture the spatial-temporal similarity of two non-overlapping travel patterns that are otherwise spatially and temporally correlated.
3. Mode entropy of spatially discretized zones and temporally discretized time intervals, which captures

the predictability of mode choices in a zone or in a time interval, is introduced to incorporate the effect of distinct features of zones and time intervals into the similarity measure.

4. A graph inference model is established to explore the correlation between the attribute-level similarities and the travel mode and spatial-temporal similarities.

The remainder of the paper is organized as follows. Section 2 describes the matrix representation of a traveler’s spatial-temporal travel history. Mode entropy of zones and time intervals, which will be used in the definition of similarity to incorporate distinct features of different spatial-temporal locations are also introduced in this section. Section 3 details the convolution based similarity measure. In Section 4, we establish a graph-based inference model to perform prediction on travelers’ mode choices. Section 5 presents a case study orienting the Proposition 1 incident in Austin, TX to validate the effectiveness of the proposed prediction framework. Section 6 concludes the paper and provides some future research directions.

2. Travelers’ spatial-temporal distribution

The workflow of the prediction framework is shown in Figure (1). Based on the spatial-temporal data of multiple travelers, we first construct a spatial-temporal matrix characterizing one’s travel history for each traveler. Leveraging travelers’ mode choices, mode entropy of zones and time intervals are calculated at the same time and will be used to account for distinct features of different zones and time intervals. A convolution-based similarity measure is thereafter developed on the matrix representation of each traveler’s mobility pattern. A graph inference model is then established to explore the correlation between the attribute-level similarities and travel mode and S-T similarities. Prediction of a traveler’s mode choice can then be made through a weighted average fashion atop the similarity values between the traveler and other travelers with known modes.

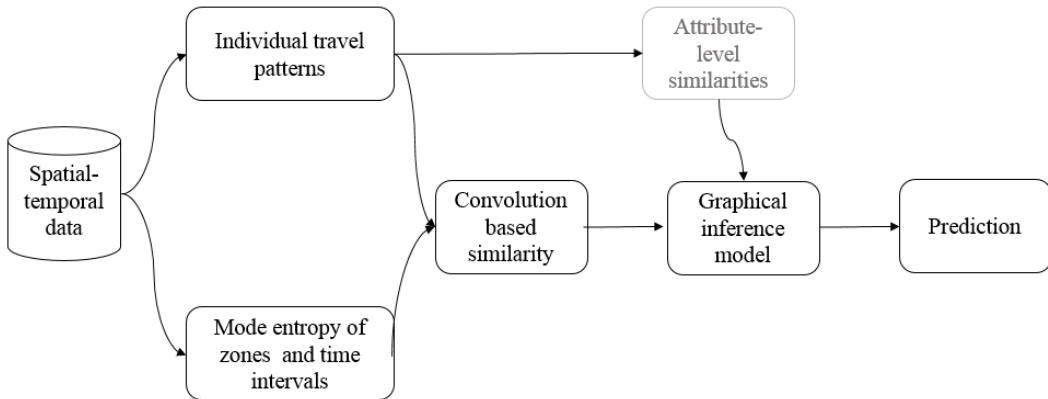


Figure 1: Work flow of similarity analysis and graph inference

2.1. Matrix representation of one’s travel history

A traveler’s mobility states can be snapshotted by the ubiquitous data she generated. Each traveler is supposed to have a ground-truth distribution characterizing her specific mobility patterns over the Spatial-Temporal (hereby called S-T) domain. Such a distribution can be overall sparse but densely concentrated in a few regions. This reveals some important travel-related information such as a traveler’s habitual activity places (e.g., home and workplace), habitual activity times (on daily, weekly, or monthly basis), and some simple but frequent travel motifs. This study aims to find an appropriate representation of a traveler’s S-T data in a tractable form and with as little information loss as possible.

One traveler j , where $j \in \{1, 2, 3, \dots, N\}$ is the index of the traveler, has an S-T record set whose elements are S-T coordinates (r_l^j, t_l^j) . r_l^j stands for the geographical coordinate, taking the form of

(*latitude, longitude*), t_l^j denotes the corresponding timestamp (simplified to daily basis or other levels), and the subscript l signifies that (r_l^j, t_l^j) is the l^{th} S-T record of traveler j . To map the traveler's mobility information to a powerful descriptive data-structure that preserves her S-T features, we use an S-T density distribution to characterize one traveler's accumulative travel history. A density distribution is theoretically a two-dimensional continuous function over the union area of all travelers, but the intrinsic spatial discreteness of their S-T data makes the zone-level discrete function a better choice. Allahviranloo and Recker (2015) adopted census tracts as the unit of spatial analysis. In practical implementation, Traffic Analysis Zone (TAZ) is a natural choice for area partitioning. Zone set is denoted by $\mathcal{Z} = \{z_1, z_2, z_3, \dots, z_Z\}$, where each $z_k \in \mathcal{Z}$ itself is a closed region in geographical space. In time domain, one calendar day, i.e., 24 hours, can be partitioned by intervals of even length or of uneven length according to some predefined rules (Garrido and Mahmassani, 2000). The time interval set is denoted by $\mathcal{T} = \{\tau_1, \tau_2, \dots, \tau_T\}$. The probability distribution of a traveler's accumulative travel history can thus be discretely represented by a matrix, referred to *S-T matrix* in this work. We denote such an S-T matrix of traveler j as ST^j , i.e.,

$$ST^j = \begin{bmatrix} p_{11}^j & p_{12}^j & p_{13}^j & \dots & p_{1Z}^j \\ p_{21}^j & p_{22}^j & p_{23}^j & \dots & p_{2Z}^j \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{T1}^j & p_{T2}^j & p_{T3}^j & \dots & p_{TZ}^j \end{bmatrix} \quad (1)$$

with each entry as

$$p_{mk}^j = \frac{\sum_{l=1}^{R_j} \mathbb{I}_{t_l^j \in \tau_m} \times \mathbb{I}_{r_l^j \in z_k}}{R_j}$$

where R_j denotes the total number of traveler j 's S-T records, and the indicator function \mathbb{I}_X is 1 when proposition X holds and 0 otherwise. p_{mk}^j is numerically equivalent to the proportion of traveler j 's S-T records falling into zone z_k during time interval τ_m and can be regarded as the maximum likelihood estimate of the probability of traveler j showing up in z_k during τ_m .

To simplify the notation, we define an *S-T field* as an area in the discretized S-T space. More formally, we provide a mathematical definition of *S-T field*.

Definition 1 (S-T Field). *Let \mathbf{x} denote an arbitrary coordinate in the spatial-temporal space. Given a partition of geographic space $\mathcal{Z} = \{z_1, z_2, \dots, z_Z\}$ and a partition of time $\mathcal{T} = \{\tau_1, \tau_2, \dots, \tau_T\}$, the set of all coordinates that satisfy $\mathbf{x} \in z_k$ and $\mathbf{x} \in \tau_m$ forms an S-T field w.r.t. z_k and τ_m .*

2.2. Mode entropy of zones and time intervals

The focus of this study is to build a prediction framework, i.e., to predict a traveler's travel mode, based on some similarity measure. One issue thus comes into play and needs to be addressed before we can appropriately define a similarity measure between two S-T matrices. The issue is: zones are not necessarily equivalent.

Now we use an example to illustrate the reason why two zones are not necessarily equivalent. We consider two zones with one as a downtown zone the other as a suburban zone. Intuitively, the downtown zone is likely a "popular" place in which travelers may have generated a lot of S-T records with different travel modes. Consequently, it is not easy to accurately predict the travel mode of a traveler who visited the downtown zone. On the contrary, in the suburban zone, a few travelers may have used the same travel mode, rendering the travel mode of a traveler who visited the suburban zone quite predictable. Thus, to incorporate the effect of the predictability of each zone into the prediction framework, we borrow the concept of *entropy* from the information theory.

Entropy in the information theory is a natural measure for information quantity or diversity (Martin and England, 2011; Song et al., 2010). Defining the occurrence of certain result in a recurrent trial as *event*, the knowledge of events with a lower probability intuitively carries more information than those with a higher probability. The rationale is that when a very probable event occurs, it provides less new information since its information is very likely to be a repetition. By transplanting the idea, mode entropy of a zone is defined

to measure the mode diversity of the zone. Following the formula of the information entropy, the mode entropy of zone z_k is calculated as

$$E_k^Z = - \sum_{l=1}^C \rho_k^l \log_b \rho_k^l \quad (2)$$

where C is the number of travel modes, $\rho_k^l = \frac{N_k^l}{\sum_{i=1}^C N_k^i}$, N_k^l is the total number of S-T records left by travelers using mode M_l in zone z_k , and the common value for logarithm base b is 2. By its definition, ρ_k^l is the ratio of the total number of S-T records left by travelers using mode M_l in zone z_k to the total number of S-T records falling in zone z_k . ρ_k^l is thus the maximum likelihood estimate of the probability that a traveler who visits zone z_k uses travel mode M_l .

E_k^Z reflects some latent properties of a zone. Highly urbanized regions (e.g., Austin downtown) are normally popular places that multiple services co-occur and compete. With larger entropy, such areas have diversified transportation ecosystem but are less characteristic for the prediction of certain travel mode. An extremely lonely place with only one or two bus routes and few taxis (or ridesharing drivers), on the contrary, has a much more monotonous transportation system. The mode entropy of such zones are smaller and travel mode will be more predictable. In later sections, the mode entropy of zones is used as a confidence score to estimate individual-level pairwise similarity.

Similarly, the mode entropy of time intervals is defined to measure the mode diversity of a given time interval. The mode entropy of time interval τ_m is calculated as

$$E_m^T = - \sum_{l=1}^C \rho_m^l \log_b \rho_m^l \quad (3)$$

where $\rho_m^l = \frac{N_m^l}{\sum_{i=1}^C N_m^i}$, and N_m^l is the total number of S-T records left by travelers using mode M_l in time interval τ_m . By its definition, ρ_m^l is the ratio of the total number of S-T records left by travelers using mode M_l in time interval τ_m to the total number of S-T records falling in τ_m . ρ_m^l is thus the maximum likelihood estimate of the probability that a traveler who commutes in time interval τ_m uses travel mode M_l .

The mode entropy of zones and time intervals can thereafter help us build a zone- and time interval-specific similarity measure by serving as a confidence score to highlight the characteristic zones and time intervals while weaken the perplexing ones.

3. Convolution-Based Spatial-Temporal Similarity

Similarity value between two travelers captures the extent to which the two travelers are alike. Characterizing a traveler's S-T data by a matrix representation, which preserves both the spatial and temporal information, this study aims to define an appropriate similarity measure which is specifically designed for the matrix representation. Before providing the similarity measure, we formulate an axiom below.

Axiom 1 (S-T similarity). *Given two S-T distribution function \mathcal{F} and \mathcal{G} , the similarity between these two distributions (denoted by $S_{\mathcal{FG}}$) should satisfy the following properties:*

1. $0 \leq S_{\mathcal{FG}} \leq 1$.
2. $S_{\mathcal{FG}} = 1$ if and only if the two distributions are congruent, i.e., $\mathcal{F} = \mathcal{G}$.
3. $S_{\mathcal{FG}} = 0$ if and only if the two distributions share no feasible domain*, i.e., $\mathcal{F} \cdot \mathcal{G} = 0$.
4. $S_{\mathcal{FG}}$ is symmetric, i.e., $S_{\mathcal{FG}} = S_{\mathcal{GF}}$.

* Feasible domain of a distribution function \mathcal{F} refers to the domain where \mathcal{F} has a nonzero value.

Denote S_{ij} as the similarity between two S-T matrices, i.e., ST^i and ST^j . A basic similarity definition can be formulated as

$$S_{ij} = \sum_{m=1}^T \sum_{k=1}^N \sqrt{p_{mk}^i \cdot p_{mk}^j} \quad (4)$$

S_{ij} equals the F-norm of the Hadamard product of traveler i 's and traveler j 's S-T matrices.

The similarity defined by Equation (4) treats all zones and time intervals equivalently without taking the property distinctions of zones or time intervals into account. Leveraging the aforementioned mode entropy of zones and time intervals, as shown in Equations (2) and (3), similarity definition is thus modified to be

$$S_{ij} \triangleq \frac{\sum_{m=1}^T \sum_{k=1}^N \frac{\sqrt{p_{mk}^i \cdot p_{nk}^j}}{E_m^T \cdot E_k^Z}}{\sqrt{\sum_{m=1}^T \sum_{k=1}^N \frac{p_{mk}^i}{E_m^T \cdot E_k^Z} \cdot \sum_{m=1}^T \sum_{k=1}^N \frac{p_{mk}^j}{E_m^T \cdot E_k^Z}}}. \quad (5)$$

The denominator is used to scale the value to make sure that the similarity measure still satisfies Axiom 1. In the numerator, the inner numerator follows the basic definition in Equation (4), while the inner denominator is the product of the entropy of a zone and the entropy of a time interval. The rationale is that a zone or a time interval with a more diversified service system is less characteristic and has a larger entropy, which in turn weakens the contribution of this zone or time interval to the overall similarity estimate.

Yet, the spatial-temporal lag effect has not been taken into account so far in the similarity measure defined in Equation (5). For the spatial dimension, lag effect describes the latent associations and dependencies among observations in a geographical space, that is, the fact that neighboring observations affect one another. Again we consider the aforementioned suburban zone (denoted by z_0) where there are one or two bus routes. The neighbor zones of such a lonely place, according to common sense, are normally very likely to hold similar attributes. In addition, an obvious choice for a traveler whose destination is one of the neighbor zones, is to first take the bus to z_0 and then head to the neighbor zone, which clearly makes the traveler a bus user. Such an effect also applies to the time domain. A trip is not a transient process and may easily span adjacent time intervals, which naturally makes temporal lag considerable. We address the effect at the S-T field level. Classic regression methods such as spatial (or local) regression model and Cliff-Ord type spatial models (Cliff, 1981) perform effective analysis on spatial lag and are proven to improve the model fit. A bunch of time-series based models, among which ARIMA regresses on its own lagged values, have also been known as commonly used methods to handle (or even take advantage of) the lag effect. These methods, however, are all based on supervised learning frameworks (such as regression models) and few accounted for the spatial-temporal cross-domain effect. To incorporate such a mechanism into an similarity-based unsupervised method, we use convolution operations and propose a new similarity measure, i.e., "convolution-based similarity".

3.1. Locality Encoding

To encode locality information into the S-T matrix, zones need to be sorted so that geographically close zones are adjacently or nearly arranged. A greedy algorithm is implemented to generate such a zone chain. Figure (2(a)) shows part of the generation result of the Austin case. Likewise, time intervals should also be arranged in a sequence. For example, we have time intervals $[(12 \text{ AM}, 1 \text{ AM}), (1 \text{ AM}, 2 \text{ AM}), \dots, (11 \text{ PM}, 12 \text{ AM})]$ if one day is partitioned into 24 one-hour intervals. Accordingly, locality information is incorporated into the S-T matrix to enable the handling of the lag effect. The closeness of two entries in the matrix indicates the factual closeness of two S-T fields both spatially and temporally. For example, S-T field b in Figure (2(b)) is closer to a than c is, since the time span is shorter. Similarly, S-T field d is closer to a than e is.

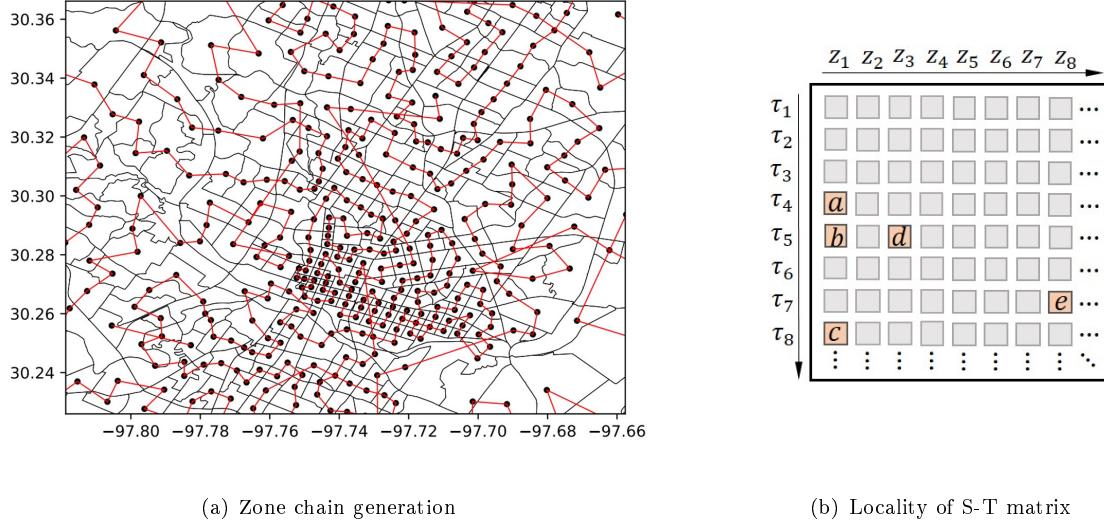


Figure 2: Rearrangement of S-T matrix. (a) shows the process where zones are sorted so that geographically close zones are adjacently or nearby arranged. (b) shows the preservation of locality after the ST mapping.

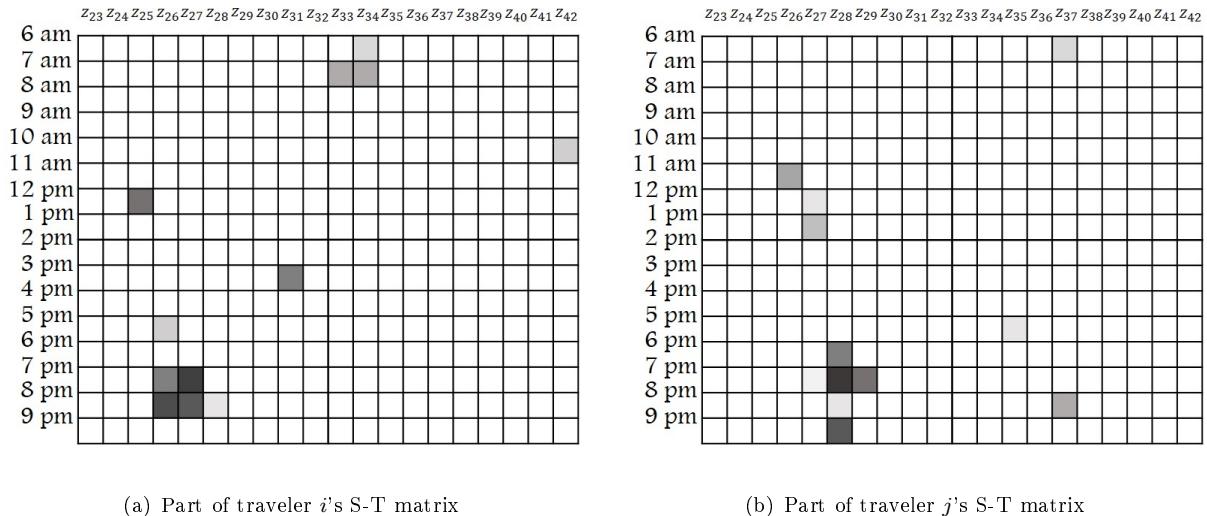


Figure 3: Part of two travelers' S-T matrix. The data were extracted from Twitter users in the case study with Austin. Obvious similarity can be observed from the comparison, however, the calculated similarity score equals 0.000472 (almost 0) according to Equation (5), indicating the similarity estimate error without lag effect taken into consideration.

Figure (3) shows part of two reported Fasten (i.e., a ridesharing service) users' S-T matrices after zone-and interval-rearrangement. Nonzero entries are colored, and a darker color indicates a higher p_{mk}^j value. The continuity of colored squares reveals the existence of the lag effect. Furthermore, conspicuous similarity can be observed from the comparison of the two travelers' S-T matrices. z_{26} to z_{28} from 6 pm to 9 pm and z_{25} to z_{27} at around 12 pm are common active S-T fields of both travelers. The calculated similarity value, however, is only 0.000472 (i.e., a very low similarity value) based on Equation (5), implying a severe estimate error.

3.2. Convolution Operation

Convolution is an operation commonly used in signal processing and now, more popularly, in computer vision. Convolution process adds each element of the raw matrix to its local neighbors, weighted by a small sliding matrix called a convolution kernel. This technique has been frequently applied to image processing for blurring and sharpening, since each pixel's neighboring information can be incorporated and manipulated through the process. Convolutional neural networks also use convolution as a basic operation to extract locality-based features in computer vision problems. Contextual semantics can be learned through convolution operations in natural language processing. We thus borrow such an idea to capture the S-T field-level lag effect in similarity estimation. In other words, we consider not only the comparison of two travelers' corresponding S-T fields, but also the interactions of neighboring S-T fields. For each traveler pair, one traveler's S-T matrix is directly used as the convolution kernel of the other one's, since spatial-temporal lag is a two-way effect in similarity estimate. Due to this nuance, classic convolution operation is tailored so that the model fits realistic scenarios better. We first review classic convolution.

The most general convolution operation takes the form of

$$(R \circledast K)(t) = \int_{-\infty}^{\infty} \Phi(R(\tau), K(t - \tau)) d\tau \quad (6)$$

where both the raw function $R(t)$ and the kernel function $K(t)$ are continuous functions and \circledast is used to distinguish the convolution product from the dot product. If the interaction function $\Phi(\cdot)$ is chosen as $\Phi(x, y) = x \cdot y$, the convolution formula then describes a weighted average of the raw function $R(t)$ where the weighting is given by the kernel function. As t changes, the weighting function emphasizes different parts of $R(t)$. In this sense, $R \circledast K$ represents the result of R being *filtered* by K . Figure (4) illustrates the generalization of convolution to two-dimensional discrete scenarios. A kernel matrix $K \in \mathbb{R}^{L \times W}$ slides over the raw matrix R entry by entry, and at each position (m, k) of R , a filtered value is calculated by

$$R'(m, k) = (R \circledast K)(m, k) = \sum_{x=m}^{m+L} \sum_{y=k}^{k+W} \Phi(R(x, y), K(m + L - x, k + W - y))$$

where R' denotes the filtered matrix. All matrices are represented in function style to show the connection with Equation (6). $\Phi(\cdot)$ represents the atom interaction between two S-T fields, which is also commonly chosen as multiplication so that K still works as a weighting function. Specific filtering effects can be realized by specific designs of the convolution kernel.

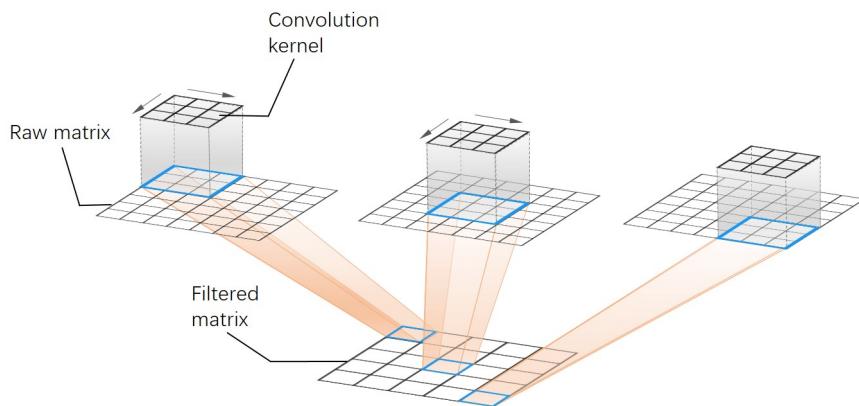


Figure 4: Classic kernel convolution process

However, it is noticeable that the size of the filtered matrix is smaller than the raw matrix, since kernel convolution usually requires values from entries outside boundaries of the matrix. Many edge-handling

methods have been proposed to address the problem. The *extending* method, which fills nearest border entries as far as necessary with 0, is adopted by our model. Figure (5) illustrates the tailored convolution process with *extending layers* taken into consideration. For a matrix pair, i.e., ST^i and $ST^j \in \mathbb{R}^{T \times Z}$, the complete size of their filtered matrix is $(2T - 1) \times (2Z - 1)$, since kernel should have at least one intersected entry with raw matrix throughout the convolution. The complete filtered matrix, whose entries are S-T field-level similarity scores, is denoted by Sim_{ij}^{Max} , i.e., $Sim_{ij}^{Max}(m, k) = (ST^i \otimes ST^j)(m, k)$, then

$$Sim_{ij}^{Max}(m, k) = \sum_{x=m-T+1}^m \sum_{y=k-Z+1}^k \Phi(ST^i(x, y), ST^j(x + T - m, y + Z - k)) \quad (7)$$

where $\Phi(\cdot)$ derives from the form of Equation (5)

$$\Phi(ST^i(x_1, y_1), ST^j(x_2, y_2)) \triangleq \frac{\sqrt{p_{x_1 y_1}^i \times p_{x_2 y_2}^j}}{\sqrt{\sum_{x_1=m-T+1}^m \sum_{y_1=k-Z+1}^k \frac{p_{x_1 y_1}^i}{E_{x_1}^T \times E_{y_1}^Z} \times \sum_{x_2=1}^T \sum_{y_2=1}^Z \frac{p_{x_2 y_2}^j}{E_{x_2}^T \times E_{y_2}^Z}}}, \quad (8)$$

given m and k . Specially,

$$E_{ab} = \sqrt{E_{x_a}^T \times E_{y_a}^Z} \times \sqrt{E_{x_b}^T \times E_{y_b}^Z} \times e^{|E_{x_a}^T - E_{x_b}^T| + |E_{y_a}^Z - E_{y_b}^Z|},$$

and for any traveler j

$$p_{xy}^j = 0, \text{ if } x \leq 0 \text{ or } x > T \text{ or } y \leq 0 \text{ or } y > Z. \quad (9)$$

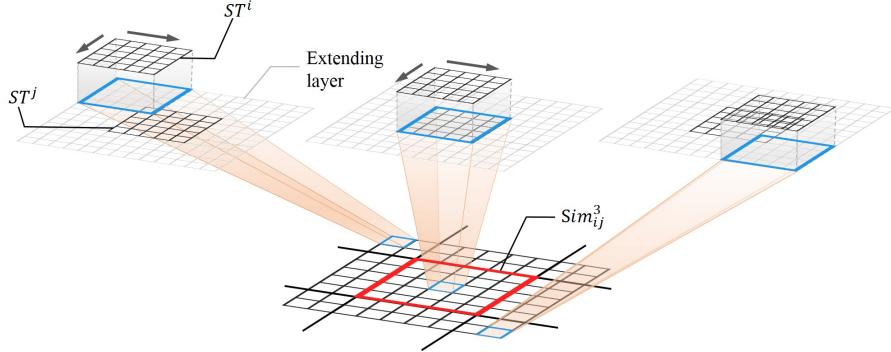


Figure 5: Spatial-temporal convolution process

Similar to Equation (5), the inner denominator in Equation (8) works as a weight to adjust the contribution of each atom interaction. The product of entropies specifies the estimate confidence of the corresponding S-T fields, and the exponential term measures the significance of the interaction. As previously mentioned, entropy reflects the latent attributes of zones and intervals. If the entropy difference $|E_{y_a}^Z - E_{y_b}^Z|$ is very small, one can believe that the two zones are intrinsically similar and strong lag effect may exist. The similarity score obtained from such a comparison are believed to be more meaningful, and therefore gets emphasized by the weight.

The spatial-temporal convolution process naturally bears more marginal entries in Sim_{ij}^{Max} with larger lag effect. For example, the most left-top entry of Sim_{ij}^{Max} comes from the start position of the convolution, where the similarity score of two most distant S-T fields, p_{11}^i and p_{TZ}^j , is calculated. On the contrary, the center entry of Sim_{ij}^{Max} describes the similarity with zero lag effect, since the value comes from the position where only ST^i and ST^j 's same S-T fields interact.

It is convenient to use the marginal degree of Sim_{ij}^{Max} 's entries to describe different lag orders, since the overall lag effect taken into account can be determined by cropping Sim_{ij}^{Max} in specific ways. Normally,

it would be reasonable to capture S-T lag within two- or three-field scope. For example, it increases two travelers' similarity if they show similar distribution in z_6 at 19 pm and in z_7 at 20 pm, respectively, but it would be inappropriate to compare traveler i 's S-T density in z_{300} at 11 am with traveler j 's in z_{500} at 11 pm. To remove unnecessary similarity entries, marginal entries of Sim_{ij}^{Max} are discarded. We provide a formal definition for clarity.

Definition 2 (N -order filtered matrix). *For a complete filtered matrix obtained from ST convolution, i.e., Sim_{ij}^{Max} , we crop its marginal rows and columns. If the remaining matrix preserves at most N -zone spatial lag and N -interval temporal lag, the post-cropping matrix is called an N -order filtered matrix, denoted by Sim_{ij}^N .*

N can be chosen according to the S-T field partitioning rules, i.e., the size of zones and the length of intervals. In implementation, we only preserve zero- to three-order lagged similarity entries, i.e., Sim_{ij}^3 , which is a 5×5 matrix as the bold black one shown in Figure (5). The convolution-based similarity measure is actually a generalization of the basic similarity defined in Equation (5).

Proposition 1. *The similarity measure defined in Equation (5) is a special case of the convolution-based similarity defined in Equation (7).*

Proof. According to Definition 2,

$$\text{Sim}_{ij}^0 = \text{Sim}_{ij}^{Max}\left(\left\lceil \frac{2T-1}{2} \right\rceil, \left\lceil \frac{2Z-1}{2} \right\rceil\right) = \text{Sim}_{ij}^{Max}(T, Z)$$

From Equations (7) and (8),

$$\begin{aligned} \text{Sim}_{ij}^{Max}(T, Z) &= \sum_{x=T-T+1}^T \sum_{y=Z-Z+1}^Z \Phi(ST^i(x, y), ST^j(x+T-T, y+Z-Z)) \\ &= \sum_{x=1}^T \sum_{y=1}^Z \frac{\sqrt{p_{xy}^i \times p_{xy}^j}}{\frac{E_x^T \times E_y^Z \times e^{|E_x^T - E_x^T| + |E_y^Z - E_y^Z|}}{\sum_{x_1=1}^T \sum_{y_1=1}^N \frac{p_{x_1 y_1}^i}{E_{x_1}^T \times E_{y_1}^Z} \times \sum_{x_2=1}^T \sum_{y_2=1}^N \frac{p_{x_2 y_2}^j}{E_{x_2}^T \times E_{y_2}^Z}}} \\ &= S_{ij} \end{aligned}$$

Hence, the similarity measure defined in Equation (5) is a special case of Sim_{ij}^N when $N = 0$, that is, when no lag effect is incorporated. \square

3.3. Similarity Calculation

For an N -order filtered matrix Sim_{ij}^N , we assign weights to its entries and convert it into the normalized final similarity score as

$$S_{ij} = \sum_{x=1}^{2N-1} \sum_{y=1}^{2N-1} w_{xy} \times \text{Sim}_{ij}^N(x, y) \quad (10)$$

where

$$w_{xy} = \frac{e^{\frac{1}{1-\text{Sim}_{ij}^N(x, y)}}}{\sum_{x=1}^{2N-1} \sum_{y=1}^{2N-1} e^{\frac{1}{1-\text{Sim}_{ij}^N(x, y)}}}$$

Proposition 2. *The similarity measure defined in Equation (10) satisfies Axiom 1.*

See proof of Proposition 2 in Appendix A.

4. A multi-layer graph inference model

Although a well-defined similarity measure may be sufficient to disclose whether similar travelers intend to choose similar travel modes, researchers may seek to understand the role of deconstructed factors in traveler's mode inclination, since insights into independent factors facilitate policy-making and adjustment of relevant strategies. For example, TNC operators can adjust their commercial strategies by understanding contributions of different factors to people's travel mode choice behavior. This section explores whether the proposed similarity measure is correlated with independent and concrete factors like radius of gyration and frequent travel patterns. Such factors are empirically considered or proven to be critical elements in travelers' activity pattern modeling and recognition (Kurkcu et al., 2016; Hawelka et al., 2014; Hasnat and Hasan, 2018). To quantify the correlation, a graph-based model is employed to perform statistical inference.

4.1. Graph inference model

The graph model here generally refers to models that demonstrate entity interactions based on graph structures. Graph models are a straightforward and powerful tool to represent causal relationship, interactions, co-occurrences, transitions, etc. They have been distinctively formulated in different tasks and scenarios. Although generalized graph models mostly are neither so sophisticated as neural networks, nor do they emphasize mathematical statistics as Bayesian networks. They have been proven to be concise and effective in previous studies. Table 1 presents typical applications of graph models in the existing studies.

Table 1: Graph model applications in literature

Reference	Graph Type	Graph Essence	Node Meaning	Edge/Arc Meaning	Function
Wolfson et al. (2012)	Directed	Physical network	A facility, e.g., a bus station	an available mode, e.g., bus	Spatial-temporal data management for transportation system
Corman et al. (2012)	Directed	State machine	A railway-related operation	A decision	To optimize the schedule in railway traffic management
Chen et al. (2013)	Directed	State machine	A driving behavior	Behavior transition	To model different driving styles
Tang et al. (2016)	Directed	Time-expanded physical network	A location at a specific time	Transition or stay	To estimate their likely space-time paths and their uncertainties within a transportation network
Lehouillier et al. (2017)	Undirected	Bipartite graph	A possible aircraft maneuver	Maneuver co-occurrence	To maintain the required separation distance between all pairs of aircrafts while minimizing fuel costs

Neural networks appear to be a heuristic type of graph models when graph nodes and edges are regarded as features and combinational relationships. Actually, different from neural networks, graph models have a distinct feature, i.e., the emphasis on *a priori* structural knowledge rather than the learning of a structure. In this sense, graph models avoid complicated interactions between features as well as nonlinear mapping. Thus straightforward interpretability can be well maintained during inference by using a graph model. The simplicity of generalized graph models also effectively controls the number of parameters in a reasonable range. Graph models can resemble Bayesian networks, on the other hand, if directed edges are used to represent causal relationships. The graph model in this work exhibits factor relationships in terms of pairwise similarity.

Figure (6) illustrates the complete structure of the graph model designed in this work. Relevant factors are organized into attribute-level layer (left), latent layer (middle), and integrated level layer (right) to show their semantic hierarchy in the problem. Attribute-level factors refer to the objectified properties that are directly obtainable from observed data. Latent factors describe higher-level factors than simple attributes. The highest-level factors, i.e., integrated factors, are theoretically capable of independently and

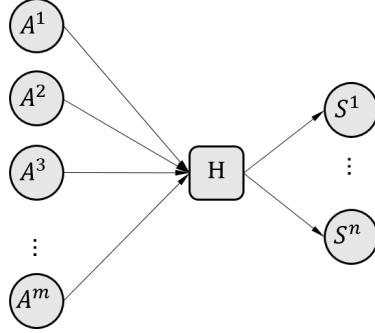


Figure 6: Multi-layer graph model

thoroughly embedding overall features. Attribute-level factors and integrated-level factors are denoted as $\{A^l\}_{l=1}^{l=m}$ and $\{S^l\}_{l=1}^{l=n}$, respectively. For all factors, pairwise similarity rather than a value from feature quantification is used to normalize the form of factor representation (i.e., an $N \times N$ matrix where N is the number of travelers).

4.2. Parameter estimation

The introduction of the hidden layer, i.e., the latent model level, facilitates to solve the graph model via a decomposition-matching procedure, saving any assumption about priors (Oselio et al., 2014). For a three-layered structure (i.e., the graph models used in this work), the inference of the hidden layer H 's component can be decomposed into two parts:

1. The inference of H utilizing the attribute matrices $\{A^l\}_{l=1}^{l=m} \rightarrow H$;
2. The inference of H utilizing the integrated factor matrices $\{S^l\}_{l=1}^{l=n} \rightarrow H$.

Hidden layer H , which is unknown in both aforementioned inference problems, is estimated as a structural combination (denoted by \mathcal{S}) of attribute matrices (A^m s) and of integrated factor matrices (S^n s), respectively. To simplify the model resolution, we choose \mathcal{S} as a linear combination. It is mentionable that more complex structures are applicable here if prior knowledge can be properly modeled. Since the two problems are dual, the inference of hidden layer H is thus sidestepped by concatenating them. An optimization problem is then formulated.

$$\begin{aligned} & \min_{\alpha, \beta} \mathcal{L}(H_\alpha, H_\beta) \\ \text{s.t. } & H_\alpha = \sum_{l=1}^m \alpha_l A^l, \quad H_\beta = \sum_{l=1}^n \beta_l S^l, \\ & \sum_{l=1}^m \alpha_l = 1, \quad \sum_{l=1}^n \beta_l = 1 \end{aligned}$$

where $\mathcal{L}(\cdot)$ is essentially a loss function that measures the discrepancy between H_α and H_β , i.e., the inferences on both sides. To maintain the convexity of the graph model, \mathcal{L} is chosen as the Frobenius norm of the matrix residual, namely, $\mathcal{L}(H_\alpha, H_\beta) = \|H_\alpha - H_\beta\|_F$ and $\|M\|_F = \sqrt{\sum_{ij} M_{ij}^2}/(T \times Z)$. With Lagrange multiplier method, the object function is converted into

$$\min_{\alpha, \beta} \sum_i^T \sum_j^Z \left(\sum_{l=1}^m \alpha_l A_{ij}^l - \sum_{l=1}^n \beta_l S_{ij}^l \right)^2 + \lambda_1 \left(\sum_{l=1}^m \alpha_l - 1 \right) + \lambda_2 \left(\sum_{l=1}^n \beta_l - 1 \right)$$

Gradient decent is used to estimate the two sets of parameters $\{\alpha\}$ and $\{\beta\}$. As the parameters of attribute-level factors, the values of $\{\alpha\}$ represent the relative significance of the attributes in travelers' travel mode choice behavior.

4.3. Prediction

In practice, it is common that only a small proportion of travelers' travel mode is known. For the rest of travelers, only their S-T data are accessible. For example, in our case study presented in the next section, only hundreds of social media users in Austin disclose their travel mode use through the contents they posted on Twitter. For other Twitter users in Austin, only their check-in data are obtainable. This raises the need to predict the massively unreported travel mode, which is a nontrivial job and can create great value for city planners, policy makers, commercial corporations and other beneficiaries. The prediction task is formulated as follows.

Formulation 1 (Mode prediction). *Given a set of travelers (denoted as \mathcal{U}) whose S-T data and modes are known, this work aims to predict the probability of a traveler $j \notin \mathcal{U}$ using mode M_i on top of her S-T data R_j , i.e., to predict the conditional probability $P(m_j = M_i | R_j)$. Notation P_{ji} is used for short.*

Based on the attribute-level similarities, i.e., A^m s and the derived weights α_m s, the corresponding similarities between traveler j and any traveler $k \in \mathcal{U}$ can be calculated as

$$A_{jk} = \sum_{l=1}^m \alpha_l A_{jk}^l \quad (11)$$

For each $k \in \mathcal{U}$, calculate a weight ω_{jk} based on A_{jk}

$$\omega_{jk} = \frac{e^{A_{jk}}}{\sum_{k' \in \mathcal{U}} e^{A_{jk'}}}$$

The probability of traveler j choosing travel mode M_i is then estimated by

$$P_{ji} = \sum_{k \in \mathcal{U}} \omega_{jk} \times \mathbb{I}_{m_k = M_i}$$

Mathematical expectations of the aggregate adoption of some travel mode (or service) can thus be calculated based on the probability, and individual choice can be estimated by $m_j = \text{argmax}_{M_i} P_{ji}$.

5. Case study with Austin, TX

We will apply the similarity analysis and graph inference developed in the previous sections to predict travelers' travel mode choices for Austinites, after Uber and Lyft suspended services in May 2016.

5.1. Data collection

The data used in this work is crawled from Twitter through its official API. Social media can provide travelers' multi-day S-T data as well as the corresponding contexts indicated by the textual tweet contents. A crawler is implemented to fetch tweets related to travel mode use based on keywords and S-T constraints. Specifically, the request URL adopted by us is

```
https://twitter.com/i/search/timeline?f=tweets&q=query%20from%3Ausername%20since%3A201x-xx-xx%20until%3A201x-xx-xx%20geocode%3Alatitude%2Clongitude%2Ckm&src=typd&max_position=x
```

where 'query', 'username', 'latitude', 'longitude', and 'x' are specified variables or numbers. This URL format enables us to exert constraints on keywords, time intervals, tweet authors, regions, etc., to fetch highly related tweets from the historical tweet pools. Basic keywords used include transportation types, names of service instances as well as their screen names, considering the fact that most of the transportation services own official accounts on Twitter, and users usually directly mention them (@) in tweets when posting relevant contents.

The fetched tweets, however, still contain severe noise such as original or retweeted news reports, political remarks, ironic comments, and perplexing babble without clearly reporting the travel mode use. Purely negative comments on some service also fail to indicate the traveler’s future alternatives. On the other hand, awarded retweets for publicity (e.g., “I just had a trip using GetMe, click and check it out...”) or coupon code (promo code) announcements imply that the travelers are highly likely to be positive adopters of the corresponding travel modes. Some tweet examples are listed in Table 2.

Table 2: Examples of related tweets

Clear attitude		<i>No #uber in Austin means I get to check out #GetMe So far so good #oscon @ Austin-Bergstrom</i>
Perplexing attitude	Political	<i>Austin City Council do you really think GetMe is the answer to the TNC issue? They are incompetent!</i>
	Sarcastic	<i>Hey! check out @GetMe_Austin - the app with the same features, wait time, and ridiculous price as YellowCab. But hey we are all “safe” now!</i>
	Negative	<i>We really need to do a sit in at the austin capitol building bc no uber/lyft is so trash & Fasten, GetMe, RideAustin ect are even more trash</i>
	Pointless babble	<i>#GETME !!!! THATS WAT IM TALKIN BOUT #AUSTIN</i>
News report		<i>As Lyft and Uber leave Austin, GetMe presses newfound advantage http://atxne.ws/1rMiE0g</i>

A hybrid approach, jointly based on rules and keyword bases, is developed to refine the raw tweets. We manually construct multiple keyword bases that collect high-frequency words in distinctive aspects (e.g., city name, politics, sentiment, opinion, action, advertisement, etc.) and determine whether to filter out a tweet by establishing rules with first order logic over the keyword bases. Specifically, keywords in complex aspects like sentiment and opinion are further classified according to their confidence level, which indicates how likely a tweet is an expected one. For a word, its confidence level is obtained by extracting 50 raw tweets containing it and calculating the “hit rate”. Such an operation gives satisfactory results while greatly reducing human efforts. Since the whole rule system gets trivial when observations grow, we show the filtering process by giving a most commonly used paradigm rule, and the execution on five tweets are shown in Table 3.

After filtering raw tweets with the hybrid method, nearly 22.6% were kept in the refinement set. By sampling tweets from both the refinement set and the discarded set, it is found that the hybrid method reaches an accuracy more than 75%. We then manually annotated the reduced data in refinement sets to further remove noise. Tweets in the final set take up nearly 17% of the raw tweets.

5.2. Basic Statistics

For a traveler j , if at least one of her tweets appears in the refinement set, then she is chosen as a sample in this work. Her tweets in the refinement set indicate her exhibited travel mode m_j . For each traveler in the sample set, her historical tweets posted within the observation period of mode m_j are collected. We extract the tweets with geo-tag and obtain the geographical coordinate through a binary search method, since Twitter does not officially provide geo-coordinates according to its new term. After establishing each traveler’s S-T database, the ones with less than 30 valid geo-tags are removed because their S-T record bases are believed to be severely incomplete. In this way, a dataset with 309 samples and 28,019 spatial-temporal records is finally constructed. Five travel modes, namely *Getme*¹, *RideAustin*², *Fasten*³, *Taxi*, and *Public Transit* are selected as travel modes of interest in this case study based on the qualified tweets. Note that public transit includes both bus and *CapMetro*⁴.

¹<http://getmeride.org/>

²<http://www.rideaustin.com/>

³<https://fasten.com/us>

⁴<https://capmetro.org/>

Table 3: A demonstration of the hybrid approach to tweet refinement

Positive sentiment: wonderful, great, happy, safe...
 Negative sentiment: frustrat-, fail, weird, atrocious, drunk...
 Positive action: try, tried, tries, enjoy, work, help, recommend, thank...
 Strongest Positive signal: coupon, promo, code, \$ off, free, flawless...
 Negation: not, n't, no...
 City name: Houston, San Antonio etc.; a city list with Austin taken out.

Rule:

```
If 'http' not in t, {  

  If any word in City_name not in t, {  

    If some word in Positive_sentiment in t, {  

      If (any word in Negation + ' ' + any word in Positive_action) not in t, {  

        If any word in Negative_sentiment not in t, {  

          Store t into refinement set.}}}}}
```

Example Tweet:

1. *Austin has no Uber, and Ride Austin isn't working on my phone :(* (discarded)
 2. *The taxi from the Austin airport is falling apart, doors don't work, seat belts broken, but somehow in this city Uber was a problem?* (discarded)
 3. *Austin, Im in Vegas & just used a service called Uber. Its like a taxi but cheaper/more convenient AND I felt a lil safer. We should get it!* (discarded)
 4. *Great, back to streets full of drunk drivers, atrocious taxi service, and pitiful public transportation. Great job, Austin. #Prop1* (discarded)
 5. *Austin Rides Free with Your Retweets - Fasten Blog [https://blog.fasten.com/austin-rides-free-with-your-retweets/...](https://blog.fasten.com/austin-rides-free-with-your-retweets/)* (accepted)
-

Table 4: Basic statistics

		Statistics	Observation Period
User		309	-
Total check-ins		28,019	-
Zone		239	-
Interval		48	-
Travel modes	Getme	62	05/14/2016 - 12/31/2016
	RideAustin	36	06/16/2016 - 12/31/2016
	Fasten	74	06/01/2016 - 12/31/2016
	Taxi	60	05/14/2016 - 12/31/2016
	Public Transit	77	05/14/2016 - 12/31/2016

Basic statistics are shown in Table 4. For zone division, Traffic Analysis Zone (TAZ) in Austin is adopted and further processed by merging zones with an inter-centroid distance less than 2 kilometers, yielding 239 zones. The time interval set, on the other hand, is obtained by evenly dividing 24 hours into 48 segments with each unit representing 30 minutes, which is a reasonable activity duration. Such interval length is fine-grained enough to differentiate spatial-temporal states and in the meantime not so trivial so that an S-T matrix will not get too sparse.

The geo-coordinates are plotted in Figure (7(a)). Yellow dots sketch the outline of the city of Austin and red dots pin the coordinates of each geo-tagged tweet posted by travelers in the sample set (i.e., the remaining 309 travelers). Figure (7(b)) presents the distribution of each zone's number of S-T records. There are 49 zones with less than 5 check-in records, which is as expected since a large area is shown to contain very few and sparse check-ins according to Figure (7(a)). Note that the scale of the horizontal axis

is not uniform for the purpose of better visualization and interpretation. Zones with 0 check-in records have already been excluded from the statistics. Figure (7(c)) presents the temporal distribution of the number of S-T records (note that here we use 24 1-hour time intervals for the purpose of better visualization). The distribution reveals people's activity intensity. 4 AM in the early morning is the nadir of people's activity intensity while it soars at around 9 AM after a gradual increase. Another surge appears at 7 PM, followed by a stable decrease afterwards.

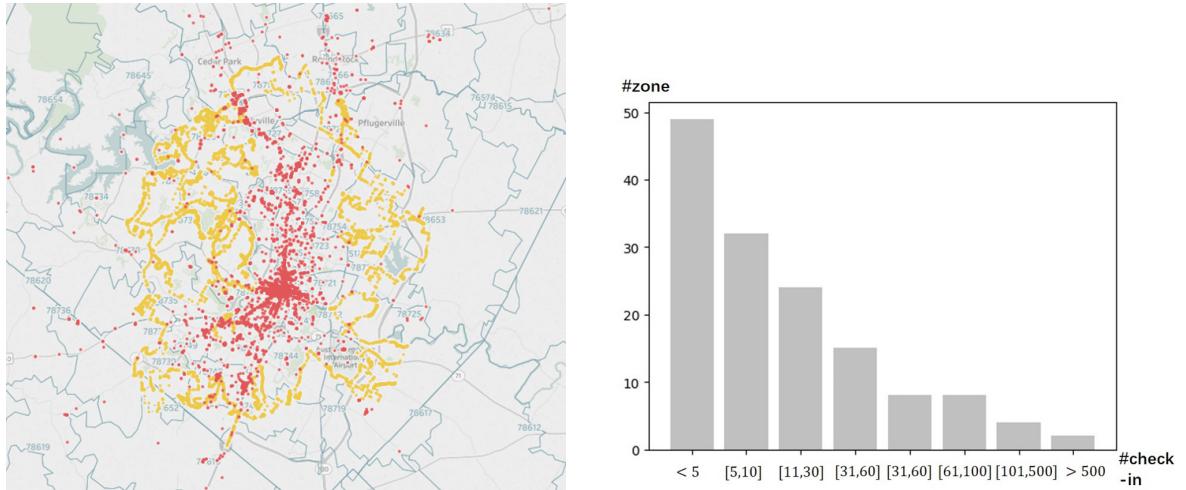


Figure 7: Visualization of S-T record distribution

5.3. Spatial-temporal Similarity

5.3.1. User base and sample set

The number of a traveler's S-T records (i.e., geo-tags or check-ins) left in Twitter obeys a power-law distribution, which is shown in Figure (8(a)). The value ranges from 1 to 34,998. The distribution is plotted based on all Twitter users in Austin who have ever left S-T records within the observation period

(i.e., May 2016 to December 2016). The number of such records reaches 205,591. Note that only a very small proportion of them (i.e., 309 in total) explicitly reported their travel mode choices and collectively constitute the sample set in this study, despite the large traveler base.

Figure (8(b)) shows the distributions of the number of S-T records for both the user base and the sample set. More than 70% users have less than 100 S-T records in both the user base and the sample set. By organizing S-T data in terms of zones, another distribution, i.e., the distribution of the number of zones in which a user leaves S-T data, is obtained and plotted in Figure (8(c)). As is shown, 59.02% users in the user base and 56.96% in the sample set take no more than 10 zones as their main activity area, and 32.35% in the user base and 35.60% in the sample set have 10 to 20 "frequent" zones.

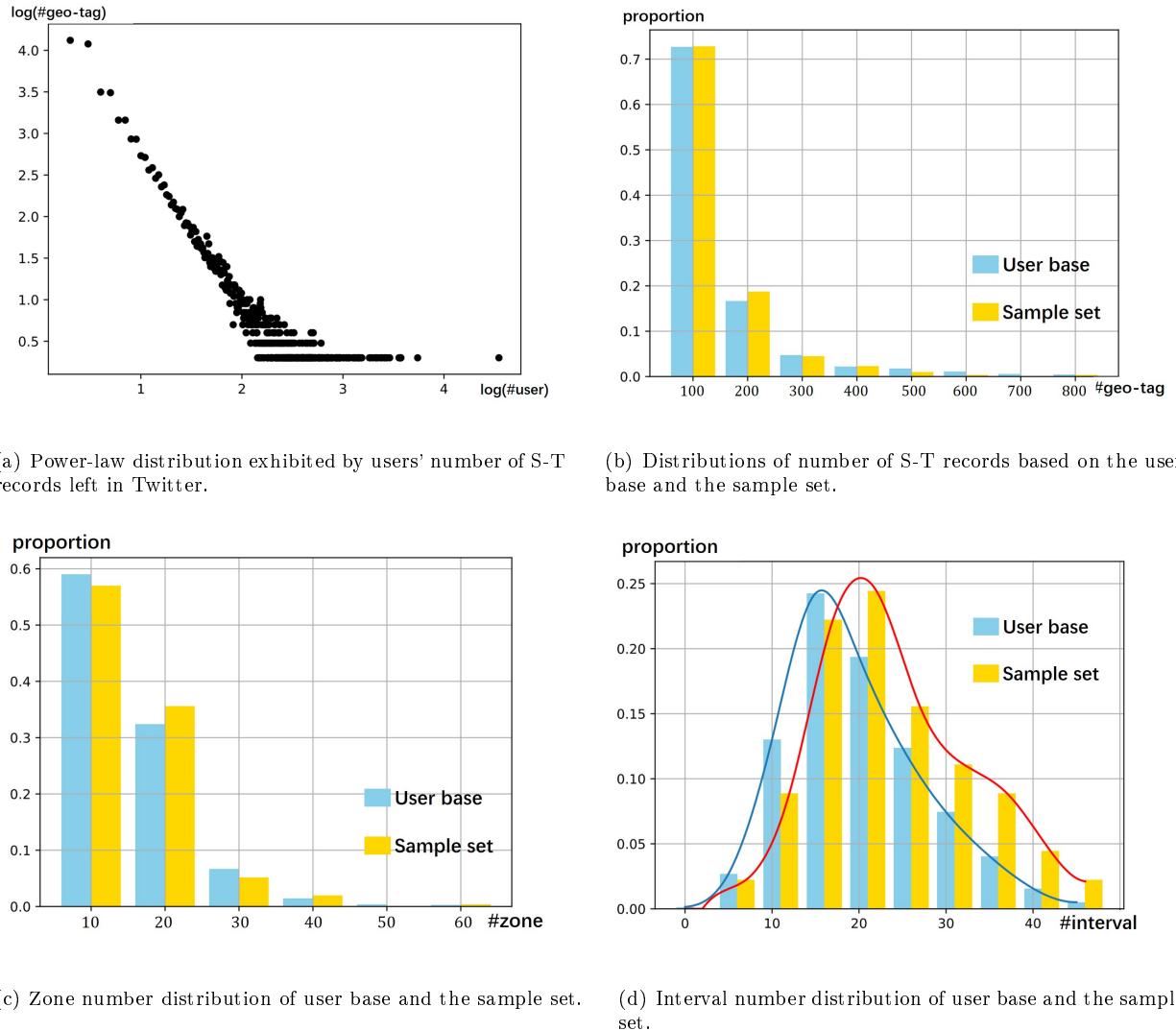


Figure 8: Individual-level statistics

Figure (8(d)) plots the distribution of the number of intervals in which a user has ever conducted activities. In terms of overall trend, the distribution of the user base slightly lags that of the sample set. Most people in the user base have about 15 activated intervals while the sample set reaches its peak value at 20. The phenomenon can be explained by the theory of *survivorship bias* (Brown et al., 1992) because the travelers in the sample set, i.e., those who explicitly informed of their travel mode use, are more likely

to be active Twitter users and in general have the tendency of oversharing. They thus have slightly more active intervals compared with the user base.

5.3.2. Spatial distribution of travel modes

Figure (9) shows the distribution of the five travel modes over 816 raw TAZs in and around Austin, with darker colors representing larger values of service density. Denser travel demands at downtown places and the airport is a common point shown by all five travel modes. Fasten shows an obvious above-average service density at the airport area (see the darkest zone in the southeast part of the figure). Dense service areas reach farthest for Fasten whereas shortest for RideAustin. Public transit takes on a reasonably uniform density throughout the city. A base distribution, which calculates user ensemble's spatial distribution, is also shown for reference.

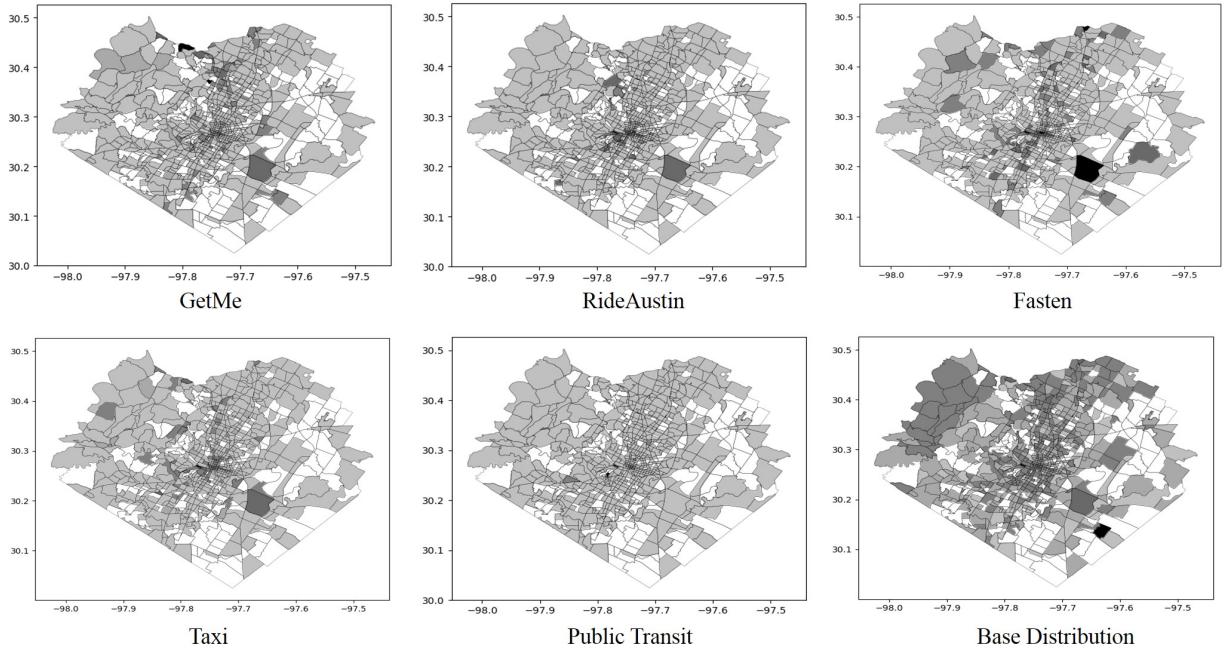


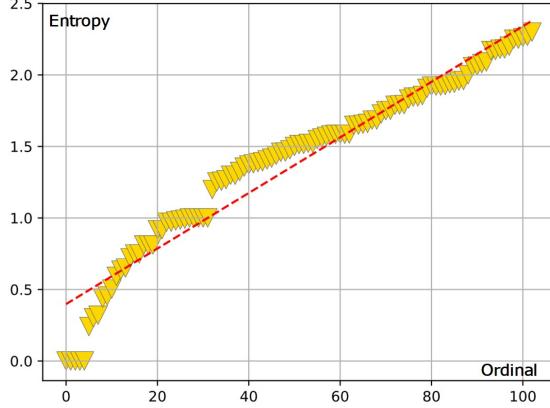
Figure 9: Density distribution of different travel modes and a base distribution

It is mentionable that all Twitter users who made geo-tagged tweets within the observation period have contributed to the estimation of base distribution, apart from the 309 samples who reported their mode choices. Independent of travel modes, base distribution plays a role as the background density distribution. In this sense, the subtraction of the base distribution from the density distribution of a travel mode transforms the service density from an absolute value to a relative value, emphasizing the net service quantity of a travel mode in some zones.

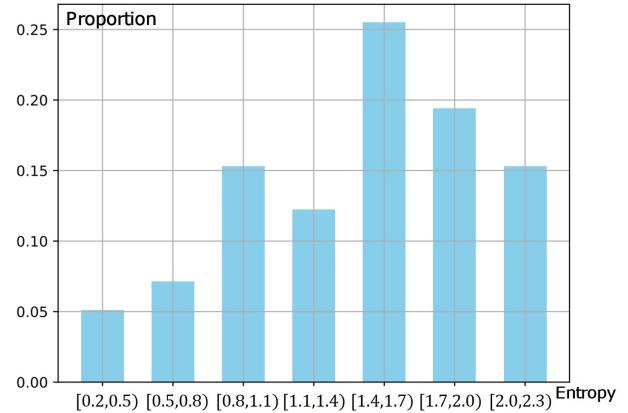
5.3.3. Mode entropy

Entropy of each zone and each interval is calculated based on the data of the sample set used in this case study. To intuitively show the distribution of the mode entropy of zones, the values are sorted and sequentially scattered as is shown in Figure (10(a)). Note that zones with entropy 0 have been removed. An approximate linear incremental tendency can be observed. Figure (10(b)) presents the distribution by segmenting the value range. Nearly 5% zones have very low entropy values (less than 0.5), representing a relatively high predictability of the travel mode choice of travelers who take the zone as a habitual activity area. Over a quarter of the zones have an entropy value between 1.4 and 1.7, taking up the largest proportion out of all segments.

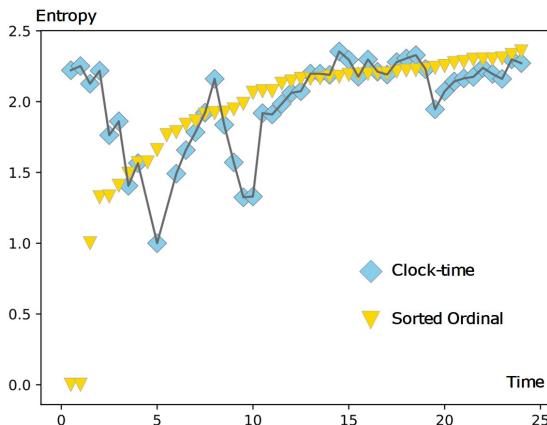
Figure (10(c)) presents the variation of the mode entropy of time intervals. The entropy value reaches a nadir at 5 AM, and exhibits generally higher values in evenings than in mornings. An increase followed by a sudden decrease of the value can be observed between 5 AM and 10 AM, specifying the rush hour in the morning when a variety of services emerge and compete to meet people's home-to-work travel need. Figure (10(d)) shows that the majority of time intervals have an entropy value ranging from 1.7 to 2.4.



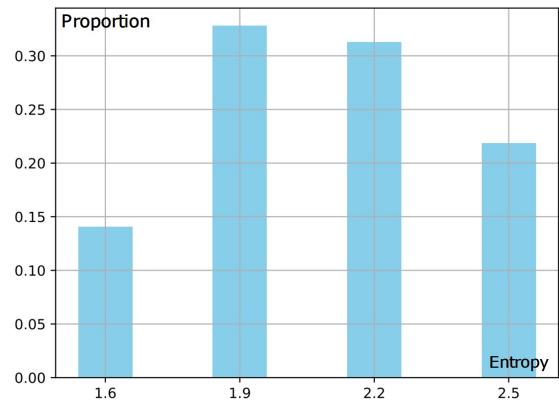
(a) Mode entropy of zones (sorted and scattered). An approximate linear incremental tendency is observed. Zones with entropy 0 are removed from the figure.



(b) Distribution of mode entropy of zones. Low entropy value represents a relatively high predictability of the travel mode.



(c) Mode entropy of time intervals (sorted and scattered). Intervals with entropy 0 are removed from the figure.



(d) Distribution of mode entropy of time intervals

Figure 10: Distributions of the mode entropy of zones and time intervals

Figure (11) shows the entropy value of each S-T field over an ST matrix. The value of each entry represents the synthetic entropy of an S-T field (denoted by $SE(m, k)$), which is calculated by the multiplication of the entropy of the corresponding zone and that of the corresponding time interval, namely $E_m^Z \cdot E_k^T$. Redder areas signify S-T fields with a higher entropy value, meaning that multiple services are observed in the S-T field, while blue areas indicate a lower entropy value. The reddest bars in the middle corresponds to downtown regions. The eide blue bar on the right corresponds to a suburban area of the city of Austin. Obvious continuity can be observed from the heatmap. Instead of being filled by random colors, color gradients are formed in several areas throughout the heatmap. The whole matrix is shown to be divided by a few blue bars, since there always exist devious places in even the most prosperous region of a city.

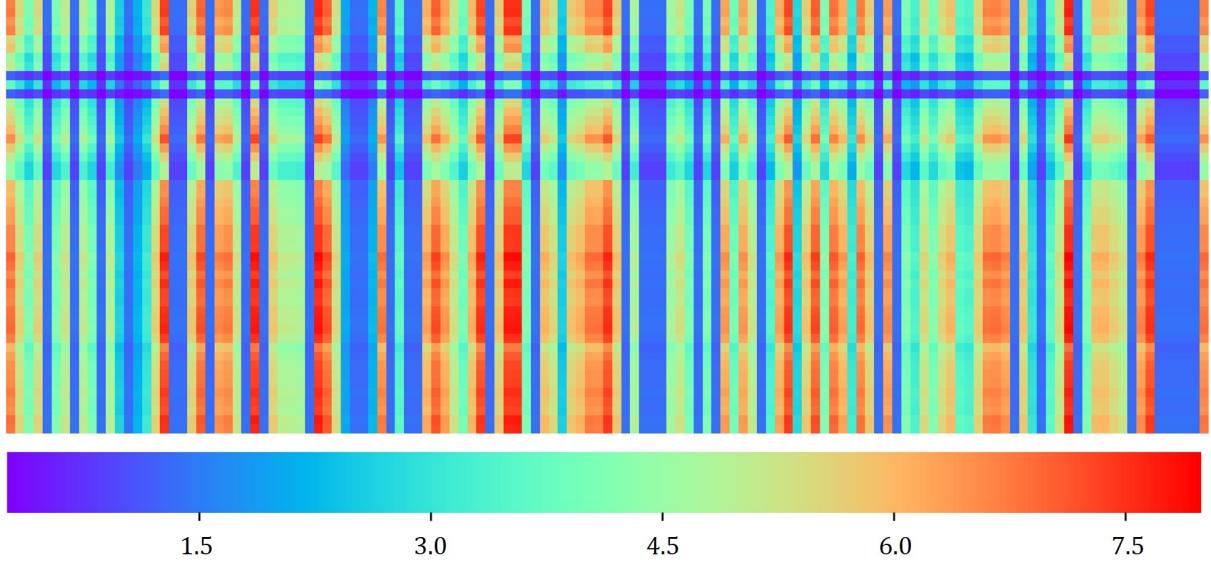


Figure 11: Entropy heatmap over a ST matrix

5.3.4. Similarity Calculation

Pairwise similarity between every two travelers in the sample set is then calculated with the convolution-based method. We plot the distribution of the calculated similarity values in Figure (12). The proportion of similarity scores between 0.3 and 0.4 reaches the peak value. The proportion of similarity scores within the range from 0 to 0.1 reaches 0.0735, indicating some travelers are highly different, with no overlapping footprints even after considering lag effects.

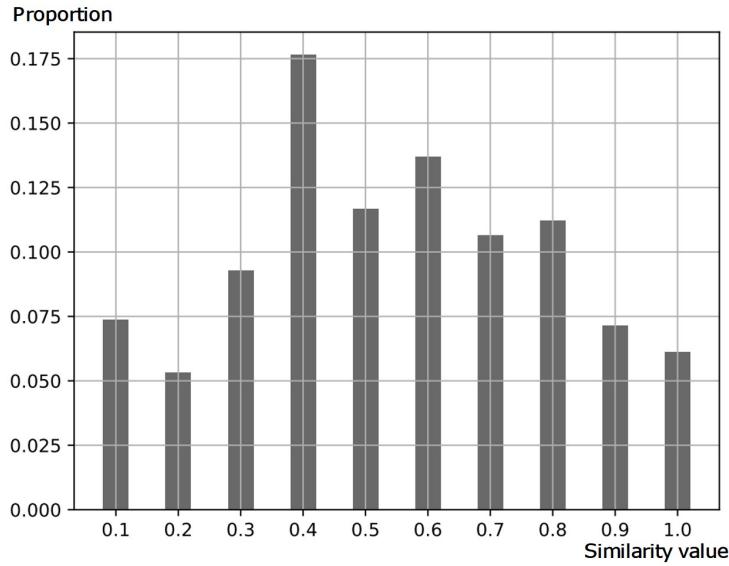


Figure 12: Similarity distribution

5.4. Graph inference models

In the case study, we include five attributes, namely, spatial distribution, temporal distribution, radius of gyration, trip frequency, and frequent travel patterns, resulting in five attribute-level factors, i.e., temporal

distribution similarity A^T , spatial distribution similarity A^Z , similarity of radius of gyration A^R , travel frequency similarity A^F , and similarity of frequent patterns A^P . As for the integrated-level factors, a discrete factor, i.e., similarity of travel mode choice S^I , and a continuous one, i.e., spatial-temporal similarity S^{ST} are jointly utilized. To explore the correlation between the attribute-level factors and the integrated-level factors, we develop three inference models, as shown in Figure (13).

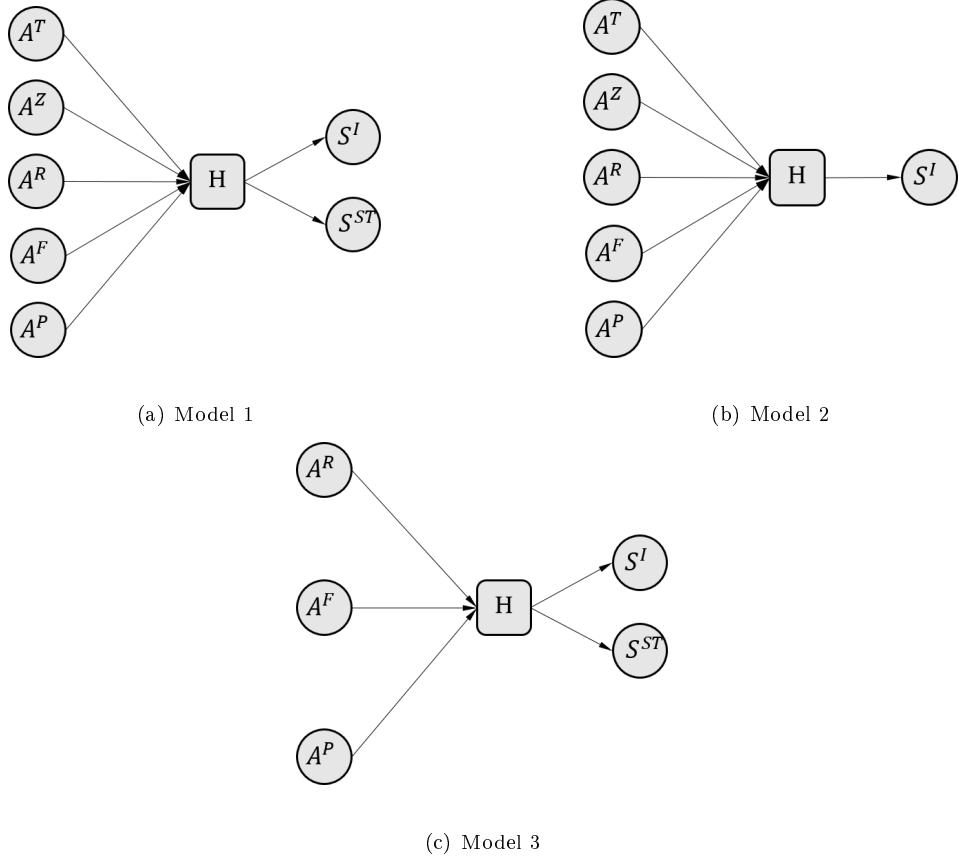


Figure 13: Three alternative graph models

5.4.1. Attribute-level similarity

5.4.1.1. Spatial distribution similarity. A more granular representation of a traveler's mobility pattern can be constructed by simply marginalizing either the spatial or the temporal dimension out of the traveler's S-T matrix. If one sums the S-T matrix over the temporal dimension, i.e., to discard the temporal information recorded in the matrix, then traveler j 's *spatial distribution vector* can be obtained as

$$D_j^Z = [p_1'^j, p_2'^j, p_3'^j, \dots, p_Z'^j] \quad (12)$$

where $p_k'^j = \sum_{m=1}^T p_{mk}^j$. $p_k'^j$ is actually the proportion of traveler j 's S-T records falling in zone z_k and can be regarded as the maximum likelihood estimate of the probability of traveler j appearing in zone z_k . Then the similarity of spatial distribution between two travelers i and j is calculated as

$$A_{ij}^Z = \frac{D_i^Z \cdot D_j^Z}{\|D_i^Z\| \times \|D_j^Z\|}$$

The similarity definition is straightforward since it calculates the cosine similarity of two travelers' spatial distribution vectors.

5.4.1.2. Temporal distribution similarity. Similarly, traveler j 's *temporal distribution vector* can be obtained by summing her S-T matrix over the spatial dimension, i.e., to discard the spatial information,

$$D_j^T = [p_1''^j, p_2''^j, p_3''^j, \dots, p_T''^j] \quad (13)$$

where $p_m''^j = \sum_{k=1}^Z p_{mk}^j$. Then the time distribution similarity between two travelers i and j is calculated as

$$A_{ij}^T = \frac{D_i^T \cdot D_j^T}{\|D_i^T\| \times \|D_j^T\|}$$

5.4.1.3. Radius of gyration. Radius of gyration measures the spread of a traveler's locations around her habitual location (Hawelka et al., 2014). It is intuitively used as an indicator of how far and how often a user commutes (Hasnat and Hasan, 2018). Radius of gyration is defined as the standard deviation of the spatial distances between the traveler's S-T coordinates and her center of mass:

$$R_g = \sqrt{\frac{1}{n} \sum_{i=1}^n (r_i - r_m)^2}$$

The definition were modified by previous studies for different applications. Kurkcu et al. (2016) took each traveler's strongest density location containing the most points as a likely home location and calculated the standard deviation of distances between S-T coordinates and the traveler's home location. Such an approach is adopted by this work. Similarity of R_g between two travelers i and j is thereafter defined as

$$A_{ij}^R = cl(R_g^i, R_g^j)$$

where the closeness function $cl(\cdot)$ is

$$cl(x, y) = 1 - 2 \times |sig(x - y) - 0.5| \quad (14)$$

and $sig(\cdot)$ is adapted from the *sigmoid* function as

$$sig(x) = \frac{1}{1 + e^{-\gamma x}},$$

so that the similarity of radius of gyration is mapped into the interval $[0, 1]$. γ is a parameter which can be tuned to make the mapping function better adapt to R_g 's distribution over all samples.

5.4.1.4. Travel frequency similarity. Travel frequency can be interpreted as an indicator of trips' temporal regularity or variability and also a measure of travelers' level of usage of the transportation system (Ortega-Tong, 2013). The authors also reported that the travel frequency is one of the clearest differences between classified individual travel patterns. Data observed during longer timespan is believed to more accurately describe people's habitual travel frequency, considering the low-resolution of some ubiquitous data type. Trip frequency is thus calculated based on the complete observation period of the corresponding travel mode. In common cases, the length of the observation period is supposed to align for all modes. However, this may not always be realistic, i.e., the length of the observation period for different travel modes may be very different, especially in some extreme scenarios, e.g., when there is/was recent disruption in the transportation system. The observation timespan of a travel mode M_i is denoted as t_i . Then traveler j 's travel frequency is simply defined as the ratio of the number of S-T records to the length of the observation period of the mode that traveler j takes, i.e.,

$$Fre_j = \frac{R_j}{\sum_{i=1}^C t_i \times \mathbb{I}_{m_j=M_i}}$$

where m_j is traveler j 's primary travel mode. Similarity of travel frequency between two travelers i and j , A_{ij}^F , is thereafter defined as

$$A_{ij}^F = cl(Fre_i, Fre_j)$$

where $cl(\cdot)$ is the closeness function defined in Equation (14).

5.4.1.5. Similarity of frequent travel patterns. Frequent sequential pattern mining is a natural way to analyze multi-day travel patterns because it is able to uncover people's frequently occurring ordered subsequences. However, the low resolution of some ubiquitous data (e.g., social media data) usually jeopardizes the continuity and recognizability of travel patterns. Bigram travel pattern mining, i.e., mining patterns of length 2, is employed to adapt to the low-resolution context. To accurately extract a true habitual bigram travel pattern, we define *habitual transition*.

Definition 3 (Habitual transition). *Given two S-T records, i.e., r_1 and r_2 , of a traveler j , the bigram sequence formed by r_1 and r_2 is called a habitual transition if and only if the following conditions hold:*

1. *The time gap between r_1 and r_2 is less than 4 hours.*
2. *One of the following holds true:*

- *r_1 and r_2 occurred at least three times in traveler j ' S-T records, respectively, and r_1 and r_2 co-occurred for at least one day.*
- *r_1 and r_2 co-occurred for at least two days.*

Note that condition 2 indicates that r_1 and r_2 are both repetitive, i.e., r_1 and r_2 both occurred at least twice in traveler j ' S-T records. Rigorously repetitive S-T records may be scarce for some data source. In other words, it may actually be rare that one traveler has two data points at exactly the same time of a day and at the same geographical point. We thus relax repetitive records to be the S-T records falling into the same S-T field more than once. This essentially partitions one traveler's S-T data by S-T fields to further determine habitual transitions. In implementation, the condition is further relaxed by considering the spatial-temporal lag effect. That is, the six nearest S-T fields are traversed to find repetitive records. Each habitual transition can be visualized as a line segment on an S-T matrix. Two travelers' habitual transitions are plotted on the same S-T matrix to better show their relative layout (see Figure (14)). Typical interpretations are as follows.

1. Approximately overlapping lines (Figure 14(a)) indicate that similar spatial transitions have taken place during similar periods. These two travel behaviors are regarded to be approximately identical in such case.
2. Approximately parallel lines (Figure 14(b)) demonstrate the circumstance where mobility trajectories are similar but the two transitions were conducted at different times.
3. If two segments share almost the same start zone and end zone but span time intervals of significantly different length (Figure 14(c)), then for the two transitions, their mobility trajectories are similar but the travel modes are highly likely to be different.
4. If two segments have an overlap but different start and/or end points, then the trajectory of one of the travelers is part of that of the other traveler, or their trajectories have common parts. This would be a typical and nontrivial scenario for public transit users.

Segments with lighter colors in Figure (14) symbolically indicates the necessity to consider lag effect. The visualization indicates that a natural way to calculate the similarity is to quantify how different two segments are based on their relative layout. We integrate the slope, intercept, and the segment length to define such a similarity of frequent patterns as

$$A_{ij}^P = \sum_a \sum_b q_a^i \cdot q_b^j \cdot cl(slope_a, slope_b) \cdot cl(inter_a, inter_b) \cdot cl(len_a, len_b)$$

where $slope_a$ denotes the slope of the line segment a , $inter_a$ stands for the intercept of the line segment a , len_a stores the length of the segment a , Tr_i denotes the habitual transfer set of traveler i , and q_a^i represents the sum of the entry values corresponding to a 's two endpoints in the S-T matrix, i.e., $q_a^i = p_{mk}^i + p_{nl}^i$ when a 's two endpoints are in zone z_k and time interval τ_m and in zone z_l and time interval τ_n , respectively.

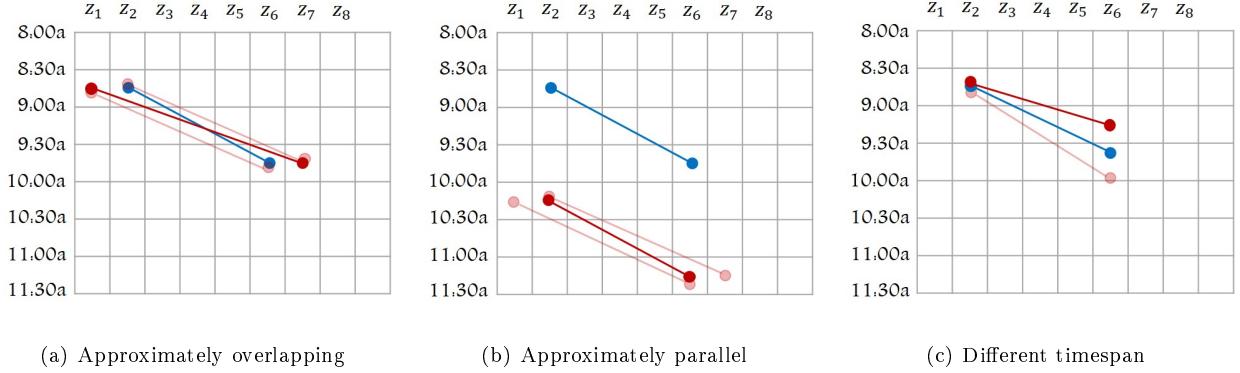


Figure 14: Habitual transitions on S-T matrix.

5.4.2. Integrated factor similarity

This layer quantifies the overall similarity of every two travelers and serves as the reference label in model optimization. Two such factors, S^I and S^{ST} , are formulated, among which the former is discrete while the latter is continuous. S^I is used to straightforwardly exhibit the consistency of two travelers' exhibited travel modes and is mathematically defined as

$$S_{ij}^I = \begin{cases} 1, & m_i = m_j \\ 0, & \text{otherwise} \end{cases}$$

In other words, S_{ij}^I essentially says that two travelers are highly similar (i.e., $S_{ij}^I = 1$) if they use the same travel mode and different (i.e., $S_{ij}^I = 0$) if they use different travel modes. For S^{ST} , the proposed convolution-based similarity measure is adopted, i.e.,

$$S_{ij}^{ST} = Sim_{ij}.$$

5.4.3. Parameter estimation

The results of parameter estimate for three inference models (see Figure (13)) are listed in Table (5).

Table 5: Parameter estimate

	α_S	α_T	α_R	α_F	α_P	β_I	β_{ST}
model 1	0.261	0.361	0.174	0.093	0.111	0.354	0.646
model 2	0.177	0.410	0.165	0.019	0.229	1	
model 3			0.294	0.099	0.607	0.330	0.670

There are several interesting observations worth mentioning. First of all, the importance of the temporal distribution similarity is outstanding, i.e., a large value of α_T , in both models 1 and 2, indicating a relatively strong correlation between the temporal distribution similarity and the travel mode and spatial-temporal similarities. Second, compared with model 2, the addition of the spatial-temporal similarity into model 1 largely increases the significance of the spatial distribution similarity (see α_S) and makes the spatial distribution similarity the second most important factor, indicating a relatively strong correlation between the spatial distribution similarity and the spatial-temporal similarity. Third, the effect of the travel frequency similarity (see α_F) is relatively minor, compared with other attribute-level similarities. Last but not least, compared with model 1, the removal of the spatial distribution similarity and the temporal distribution similarity in model 3 substantially increases the importance of the similarity of frequent patterns (see

α_P), indicating that the similarity of frequent patterns is believed to be correlated with the spatial and temporal distribution similarities and its importance was suppressed by the spatial and temporal distribution similarities in model 1.

5.5. Prediction

In addition to use the weighted average of the attribute-level similarities as the similarity between two travelers j and k , as shown in Equation (11), we can also directly use the spatial-temporal similarity S_{jk}^{ST} as the similarity between j and k . To be specific, we call the prediction model explicitly using spatial-temporal similarity the benchmark prediction model. We call the prediction model using attribute-level similarities based on the graph models 1, 2, and 3, as shown in Figure (13), prediction models 1, 2, and 3, respectively.

Leave-one-out method is applied to predict travelers' travel mode choices in the sample set. To predict Twitter user j 's most likely travel mode choice, we use all the rest of the users as training set and predict j 's mode choice based on their choices as well as the similarity scores between them and j . We apply the leave-one-out method to four prediction models, i.e., prediction models 1, 2, 3, and the benchmark prediction model.

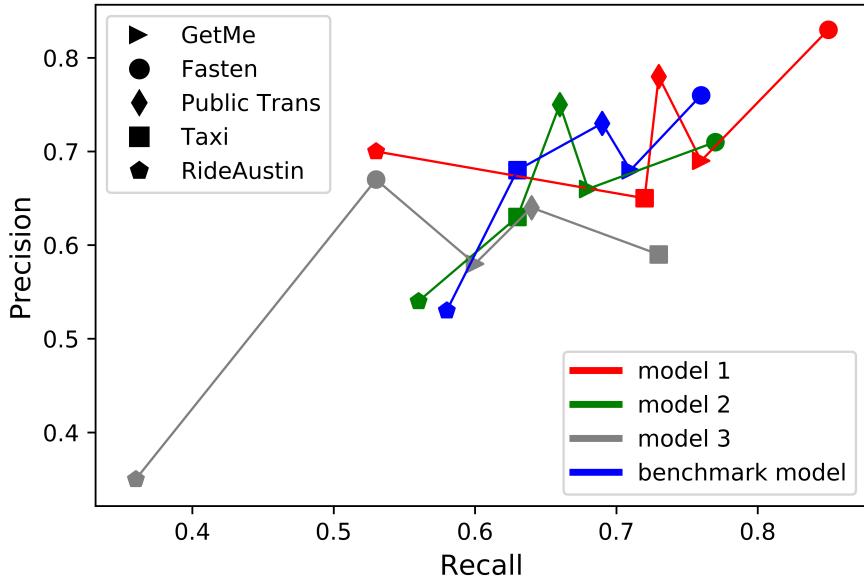


Figure 15: Performance of prediction models

The prediction results are presented in Tables (6), (7), (8), and (9) in Appendix B. The overall prediction accuracies of model 1, model 2, model 3, and the benchmark model are 74%, 67%, 58%, and 69%, respectively. Each row in the table presents the prediction results for one travel mode. For example, if we look at the first row in Table (6), there are in total 62 travelers using travel mode *GetMe*, and prediction model 1 successfully predicts 47 out of 62 using the mode, yielding a 76% recall for the mode *GetMe*. Figure (15) visualizes the performance in terms of recall and precision of all prediction models. Intuitively, the higher recall and precision the better the performance. The outperformance of model 1 is quite obvious. The prediction of the travel mode *Fasten* is generally the best, indicating the travel mode *Fasten* is quite characteristic and different from other modes. The prediction of travel mode *RideAustin* is again the worst, indicating the mode itself may not be characteristic enough.

The outperformance of model 1 over models 2 and 3 can be explained as follows. It is not surprising that the performance of model 3 is the worst since model 3 drops two important attribute-level factors, i.e.,

spatial distribution and temporal distribution. In other words, the remaining three attribute-level factors in model 3 are not sufficient to explain the integrated factors well. Compared with model 1, model 2 uses only the binary integrated factor, i.e., S^I . Considering the fact that the attribute-level features of two individuals are more continuous, it is as expected that a part of the correlation between the more continuous attribute-level features and the binary integrated factor is missed.

Independent of any inference models, the benchmark model outperforms prediction models 2 and 3 and can reach an overall acceptable prediction accuracy, i.e., 68%, indicating the effectiveness of the proposed convolution-based similarity measure. In particular, model 2 has already incorporated rich attribute-level information. Thus the outperformance of the benchmark prediction model over prediction model 2 indicates that the correlation between the travel mode and the spatial-temporal similarity is stronger than that between the travel mode similarity and the weighted average of five attribute-level similarities.

The outperformance of prediction model 1 over the benchmark prediction model validates the effectiveness of the proposed prediction framework. In other words, the addition of the spatial-temporal similarity as part of the integrated factor into inference model 1 can effectively adjust the weights of each attribute-level similarity and thus results in a better performance in prediction.

Based on prediction model 1, prediction results on *GetMe*, *Fasten*, *Taxi*, and *Public Transit* are relatively good, i.e., with a recall over 70%, while the recall of *RideAustin* is comparatively lower, indicating that *RideAustin* is less characteristic as a travel mode (or service type) compared with the other four modes. The values of entries for row *GetMe* and column *Taxi* and column *GetMe* and row *Taxi* are both relative high, indicating that the service modality of the two applications are essentially similar, and they are the direct alternative mode of one another from the perspective of travelers. Their similarity can also be found in the comparison shown in Figure (9).

6. Conclusion

This study provides insights into people's travel patterns by similarity analysis and introduces a novel prediction framework atop the proposed convolution-based similarity measure and graph inference models. Individual travel patterns are generalized into a spatial-temporal probabilistic distribution over given zones and specified temporal basis, i.e., the matrix representation of a traveler's S-T data. A modified convolution operation is then developed to calculate the similarity of two S-T matrices, with spatial and temporal lag effect well captured and incorporated into the similarity estimation. Entropy is borrowed from the information theory to account for distinct features of different zones and time intervals. A graph-based inference model is established to further analyze the relative importance of common factors like radius of gyration, trip frequency, and frequent travel patterns. To predict a traveler's mode choice, a weighted average method is employed atop the similarity values between the traveler and other travelers with known modes.

Based on spatial-temporal data crawled from one of the most prevalent location-based social media (LBSM), i.e., Twitter, a case study is conducted to validate the developed prediction framework. To the best of our knowledge, this is the first study on travel pattern similarity estimation that is compatible with social media data as well as ubiquitous data of any resolution. The similarity measure is applied to infer Twitter users' travel mode choices after Uber and Lyft's suspension in Austin, Texas in May 2016. The proposed prediction framework is capable of accurately predicting unreported travel modes. Prediction results validate not only the proposed convolution-based similarity measure but also the developed prediction framework. Here we stress that the prediction framework is general in the sense that it can be easily generalized to predict travelers' other attributes as long as those attributes are known among a small portion of the population. Thus, the generalized prediction framework can be utilized to uncover people's lifestyle, travel demand, social identity, and other attributes, indicating a great potential for creating value for city planners, policy makers, commercial corporations and other beneficiaries.

The current work can be extended in two-folds: (1) Instead of simply using keywords to extract the travel mode choice of each Twitter user, i.e., independent of context, a natural language processing based algorithm can be applied to refine the extraction of travel mode choices. (2) Each individual is assumed to

have only one travel mode, which may not be the case in practice. One potential way is to characterize each individual's mode choice as a probability distribution and will be left for future work.

Acknowledgments

This research was funded by the National Science Foundation (NSF) under Grant Number CMMI-1745708. The third author would also like to thank Prof. Yang Liu from University of California, Santa Cruz for his inspiring discussions in graph inference models.

Appendix A

Proof. First, we prove for any positive integer m and k we have $0 \leq \text{Sim}_{ij}^N(m, k) \leq 1$.

According to Equation (7) and (8), $\text{Sim}_{ij}^N(m, k) \geq 0$ is obvious. Given any m and k , let

$$\Delta = \sqrt{\sum_{x_1=m-T+1}^m \sum_{y_1=k-Z+1}^k \frac{p_{x_1 y_1}^i}{E_{x_1}^T \times E_{y_1}^Z} \times \sum_{x_2=1}^T \sum_{y_2=1}^N \frac{p_{x_2 y_2}^j}{E_{x_2}^T \times E_{y_2}^Z}},$$

that is, the denominator in Equation (8). Then from Equation (7),

$$\begin{aligned} \text{Sim}_{ij}^{Max}(m, k) &= \frac{1}{\Delta} \sum_{x_1=m-T+1}^m \sum_{y_1=k-Z+1}^k \frac{\sqrt{p_{x_1 y_1}^i \times p_{x_2 y_2}^j}}{\sqrt{E_{x_1}^T \times E_{y_1}^Z} \times \sqrt{E_{x_2}^T \times E_{y_2}^Z} \times e^{|E_{x_1}^T - E_{x_2}^T| + |E_{y_1}^Z - E_{y_2}^Z|}} \\ &\leq \frac{1}{\Delta} \sum_{x_1=m-T+1}^m \sum_{y_1=k-Z+1}^k \sqrt{\frac{p_{x_1 y_1}^i}{E_{x_1}^T \times E_{y_1}^Z}} \times \sqrt{\frac{p_{x_2 y_2}^j}{E_{x_2}^T \times E_{y_2}^Z}} \end{aligned}$$

where $x_2 = x_1 + T - m$ and $y_2 = y_1 + Z - k$. According to Cauchy-Schwarz inequality,

$$\begin{aligned} \left(\sum_{x_1}^m \sum_{y_1}^k \sqrt{\frac{p_{x_1 y_1}^i}{E_{x_1}^T \times E_{y_1}^Z}} \times \sqrt{\frac{p_{x_2 y_2}^j}{E_{x_2}^T \times E_{y_2}^Z}} \right)^2 &\leq \sum_{x_1}^m \sum_{y_1}^k \left(\sqrt{\frac{p_{x_1 y_1}^i}{E_{x_1}^T \times E_{y_1}^Z}} \right)^2 \times \sum_{x_2}^T \sum_{y_2}^Z \left(\sqrt{\frac{p_{x_2 y_2}^j}{E_{x_2}^T \times E_{y_2}^Z}} \right)^2 \\ &= \Delta^2 \end{aligned}$$

Therefore, for any given m and k

$$\text{Sim}_{ij}^{Max}(m, k) \leq \frac{\sqrt{\Delta^2}}{\Delta} = 1$$

which makes $\text{Sim}_{ij}^N(m, k) \leq 1$. Now we prove the four properties in Axiom 1 in order.

1. According to Equation (10), $S_{ij} \geq 0$ is obvious. Since it has been proven that $\text{Sim}_{ij}^N(m, k) \leq 1$,

$$S_{ij} = \sum_{x=1}^{2N-1} \sum_{y=1}^{2N-1} w_{xy} \times \text{Sim}_{ij}^N(x, y) \leq \sum_{x=1}^{2N-1} \sum_{y=1}^{2N-1} w_{xy} = 1.$$

2. When $ST^i = ST^j$,

$$\begin{aligned} \text{Sim}_{ij}^{Max}(T, Z) &= \sum_{x=1}^T \sum_{y=1}^Z \frac{\sqrt{p_{x y}^i \times p_{x y}^j}}{\sqrt{\sum_{x_1=1}^T \sum_{y_1=1}^N \frac{p_{x_1 y_1}^i}{E_{x_1}^T \times E_{y_1}^Z} \times \sum_{x_2=1}^T \sum_{y_2=1}^N \frac{p_{x_2 y_2}^j}{E_{x_2}^T \times E_{y_2}^Z}}} \\ &= \frac{\sum_{x=1}^T \sum_{y=1}^Z \frac{p_{x y}^i}{E_x^T \times E_y^Z}}{\sum_{x=1}^T \sum_{y=1}^Z \frac{p_{x_1 y_1}^i}{E_{x_1}^T \times E_{y_1}^Z}} = \frac{\sum_{x=1}^T \sum_{y=1}^Z \frac{p_{x y}^j}{E_x^T \times E_y^Z}}{\sum_{x=1}^T \sum_{y=1}^Z \frac{p_{x_2 y_2}^j}{E_{x_2}^T \times E_{y_2}^Z}} \\ &= 1 \end{aligned}$$

Then $\text{Sim}_{ij}^N(N, N) = \text{Sim}_{ij}^{Max}(T, Z) = 1$, meaning that there exists at least one entry with the value 1 in matrix Sim_{ij}^N . Suppose there are another $(a-1)$ entries whose value equal 1. We index those entries from 1 to a and the rest from $a+1$ to N^2 , and then rearrange all entries into set $\{d_i\}$ ($i = 1, 2, \dots, N^2$). Since function $e^{\frac{1}{1-x}}$ is not defined at $x = 1$, we denote $d_i = \lim_{\delta \rightarrow 0} (1 - \delta)$ for $i = 1, 2, \dots, a$ and calculate the limit of S_{ij} . If $d_i \neq 0$ for any i , from Equation (10),

$$S_{ij} = \sum_{x=1}^{2N-1} \sum_{y=1}^{2N-1} w_{xy} \times \text{Sim}_{ij}^N(x, y) = \sum_{i=1}^{N^2} w_i \times d_i \sum_{i=1}^{N^2} \frac{e^{\frac{1}{1-d_i}}}{\sum_{j=1}^{N^2} e^{\frac{1}{1-d_j}}} \times d_i$$

Now that $d_i = \lim_{\delta \rightarrow 0} (1 - \delta)$ for $i = 1, 2, \dots, a$ then

$$\begin{aligned} \lim_{\delta \rightarrow 0} S_{ij} &= \lim_{\delta \rightarrow 0} \left[\sum_{i=1}^a \frac{e^{\frac{1}{\delta}}}{\sum_{j=1}^a e^{\frac{1}{\delta}} + \sum_{j=a+1}^{N^2} e^{\frac{1}{1-d_j}}} \times (1 - \delta) + \sum_{i=a+1}^{N^2} \frac{e^{\frac{1}{1-d_i}}}{\sum_{j=1}^a e^{\frac{1}{\delta}} + \sum_{j=a+1}^{N^2} e^{\frac{1}{1-d_j}}} \times d_i \right] \\ &= a \times \lim_{\delta \rightarrow 0} \frac{e^{\frac{1}{\delta}}}{a \times e^{\frac{1}{\delta}} + \text{const}} \times (1 - \delta) + 0 = a \times \frac{1}{a} \times 1 = 1 \end{aligned}$$

3. If we consider spatial-temporal lag effect in travelers' spatial-temporal distribution, two travelers i and j are regarded totally different only if their footprints have no overlapping throughout the lag scope, i.e., for any m and k , once $p_{mk}^i > 0$ then $\sum_{x=m-N+1}^{m+N-1} \sum_{y=k-N+1}^{k+N-1} p_{xy}^j = 0$. This makes the interaction function $\Phi(\cdot) = 0$ for any entry $\text{Sim}_{ij}^{Max}(m, k)$ falling into submatrix Sim_{ij}^N , and thus makes Sim_{ij}^N a zero matrix. Then according to Equation (10), we have $S_{ij} = 0$.
4. We first show $\text{Sim}_{ij}^{Max}(m, k) = \text{Sim}_{ji}^{Max}(2T - m, 2Z - k)$, that is, if we transpose Sim_{ij}^{Max} and then flip it over its anti-diagonal we get Sim_{ji}^{Max} . Assume $m < T$ and $k < Z$.

$$\begin{aligned} \text{Sim}_{ij}^{Max}(m, k) &= \sum_{x=m-T+1}^m \sum_{y=k-Z+1}^k \Phi(ST^i(x, y), ST^j(x + T - m, y + Z - k)) \\ &= \frac{1}{\Delta(m, k)} \sum_{x=m-T+1}^m \sum_{y=k-Z+1}^k \frac{\sqrt{p_{xy}^i p_{x+T-m, y+Z-k}^j}}{E} \\ &= \frac{1}{\Delta(m, k)} \left[\frac{\sqrt{p_{m-T+1, k-Z+1}^i p_{11}^j}}{E} + \frac{\sqrt{p_{m-T+1, k-Z+2}^i p_{12}^j}}{E} + \dots + \frac{\sqrt{p_{mk}^i p_{TZ}^j}}{E} \right] \end{aligned}$$

According to margin condition described by Equation (9), when $m < T$ and $k < Z$, for any $x \in \{m - T + 1, m - T + 2, \dots, 0\}$ or $y \in \{k - Z + 1, k - Z + 2, \dots, 0\}$ we have $p_{xy}^i = 0$. Then

$$\text{Sim}_{ij}^{Max}(m, k) = \frac{1}{\Delta(m, k)} \left[\frac{\sqrt{p_{11}^i p_{T-m+1, Z-k+1}^j}}{E} + \frac{\sqrt{p_{11}^i p_{T-m+1, Z-k+2}^j}}{E} + \dots + \frac{\sqrt{p_{mk}^i p_{TZ}^j}}{E} \right] \quad (15)$$

And for matrix Sim_{ji}^{Max} , bring the entry $(2T - m, 2Z - k)$ into Equation (7)

$$\begin{aligned} \text{Sim}_{ji}^{Max}(2T - m, 2Z - k) &= \sum_{x=T-m+1}^{2T-m} \sum_{y=Z-k+1}^{2Z-k} \Phi(ST^j(x, y), ST^i(x - T + m, y - Z + k)) \\ &= \frac{1}{\Delta(2T - m, 2Z - k)} \sum_{x=T-m+1}^{2T-m} \sum_{y=Z-k+1}^{2Z-k} \frac{\sqrt{p_{xy}^j p_{x-T+m, y-Z+k}^i}}{E} \\ &= \frac{1}{\Delta(2T - m, 2Z - k)} \left[\frac{\sqrt{p_{T-m+1, Z-k+1}^j p_{11}^i}}{E} + \frac{\sqrt{p_{T-m+1, Z-k+2}^j p_{12}^i}}{E} + \dots + \frac{\sqrt{p_{2T-m, 2Z-k}^j p_{TZ}^i}}{E} \right] \end{aligned}$$

Similarly, by considering the margin condition, for any $x \in \{T + 1, T + 2, \dots, 2T - m\}$ or $y \in \{Z + 1, Z + 2, \dots, 2Z - k\}$ we have $p_{xy} = 0$. Then

$$\begin{aligned} Sim_{ji}^{Max}(2T - m, 2Z - k) &= \frac{1}{\Delta(2T - m, 2Z - k)} \left[\frac{\sqrt{p_{T-m+1,Z-k+1}^j p_{11}^i}}{E} + \right. \\ &\quad \left. \frac{\sqrt{p_{T-m+1,Z-k+2}^j p_{12}^i}}{E} + \dots + \frac{\sqrt{p_{T,Z}^j p_{mk}^i}}{E} \right] \end{aligned} \quad (16)$$

By comparing Equation (15) and (16), one should notice that the numerators are equivalent. The denominator $\Delta(m, k)$ and $\Delta(2T - m, 2Z - k)$ can be proven equal similarly using the margin condition. Hence, when $m < T$ and $k < Z$

$$Sim_{ij}^{Max}(m, k) = Sim_{ji}^{Max}(2T - m, 2T - k)$$

and for any N , Sim_{ij}^N share entry values with Sim_{ji}^N . Based on Formula (10), we know $S_{ij} = S_{ji}$. Under the conditions $\{m < T, k \geq Z\}, \{m \geq T, k < Z\}$ and $\{m \geq T, k \geq Z\}$ the proofs are similar.

□

Appendix B

Table 6: Confusion matrix of travel mode prediction (prediction model 1). Overall accuracy: 74%

		Predicted					
Actual		GetMe	Fasten	RideAustin	Taxi	Public Trans	Recall
	GetMe	47	3	0	11	1	47/62=0.76
	Fasten	3	63	1	2	5	63/74=0.85
	RideAustin	3	1	19	5	8	19/36=0.53
	Taxi	11	3	1	43	2	43/60=0.72
	Public Trans	4	6	6	5	56	56/77=0.73
	Precision	47/68=0.69	63/76=0.83	19/27=0.70	43/66=0.65	56/72=0.78	

Table 7: Confusion matrix of travel mode prediction (prediction model 2). Overall accuracy: 67%

		Predicted					
Actual		GetMe	Fasten	RideAustin	Taxi	Public Trans	Recall
	GetMe	42	6	2	9	3	42/62=0.68
	Fasten	6	57	1	7	3	57/74=0.77
	RideAustin	4	1	20	3	8	20/36=0.56
	Taxi	10	7	2	38	3	38/60=0.63
	Public Trans	2	9	12	3	51	51/77=0.66
	Precision	42/64=0.66	57/80=0.71	20/37=0.54	38/60=0.63	51/68=0.75	

Table 8: Confusion matrix of travel mode prediction (prediction model 3). Overall accuracy: 58%

		Predicted					
Actual		GetMe	Fasten	RideAustin	Taxi	Public Trans	Recall
	GetMe	37	4	1	14	6	$37/62=0.60$
	Fasten	6	39	11	8	10	$39/74=0.53$
	RideAustin	5	6	13	3	9	$13/36=0.36$
	Taxi	12	2	0	44	2	$44/60=0.73$
	Public Trans	4	7	12	5	49	$49/77=0.64$
Precision		$37/64=0.58$	$39/58=0.67$	$13/37=0.35$	$44/74=0.59$	$49/76=0.64$	

Table 9: Confusion matrix of travel mode prediction (the benchmark prediction model). Overall accuracy: 69%

		Predicted					
Actual		GetMe	Fasten	RideAustin	Taxi	Public Trans	Recall
	GetMe	44	3	4	8	3	$44/62=0.71$
	Fasten	5	56	0	6	7	$56/74=0.76$
	RideAustin	3	2	21	1	9	$21/36=0.58$
	Taxi	11	6	3	38	2	$38/60=0.63$
	Public Trans	2	7	12	3	53	$53/77=0.69$
Precision		$44/65=0.68$	$56/74=0.76$	$21/40=0.53$	$38/56=0.68$	$53/74=0.72$	

References

- Allahviranloo, M., Recker, W., 2013. Daily activity pattern recognition by using support vector machines with multiple classes. *Transportation Research Part B: Methodological* 58, 16–43.
- Allahviranloo, M., Recker, W., 2015. Mining activity pattern trajectories and allocating activities in the network. *Transportation* 42 (4), 561–579.
- Allahviranloo, M., Regue, R., Recker, W., 2014. Pattern Clustering and Activity Inference.
- Atluri, G., Karpatne, A., Kumar, V., Aug. 2018. Spatio-Temporal Data Mining: A Survey of Problems and Methods. *ACM Comput. Surv.* 51 (4), 83:1–83:41.
- Bhat, C. R., 1995. A heteroscedastic extreme value model of intercity travel mode choice. *Transportation Research Part B: Methodological* 29 (6), 471–483.
- Bhat, C. R., Dubey, S. K., 2014. A new estimation approach to integrate latent psychological constructs in choice modeling. *Transportation Research Part B: Methodological* 67, 68–85.
- Brown, S. J., Goetzmann, W., Ibbotson, R. G., Ross, S. A., Oct. 1992. Survivorship Bias in Performance Studies. *The Review of Financial Studies* 5 (4), 553–580.
- Calastri, C., Hess, S., Choudhury, C., Daly, A., Gabrielli, L., 2017. Mode choice with latent availability and consideration: theory and a case study. *Transportation Research Part B: Methodological*.
- Candia, J., González, M. C., Wang, P., Schoenharl, T., Maday, G., Barabási, A.-L., 2008. Uncovering individual and collective human dynamics from mobile phone records. *Journal of physics A: mathematical and theoretical* 41 (22), 224015.
- Cantillo, V., Heydecker, B., de Dios Ortúzar, J., 2006. A discrete choice model incorporating thresholds for perception in attribute values. *Transportation Research Part B: Methodological* 40 (9), 807–825.
- Cantillo, V., Ortúzar, J. d. D., Williams, H. C., 2007. Modeling discrete choices in the presence of inertia and serial correlation. *Transportation Science* 41 (2), 195–205.
- Chen, S.-W., Fang, C.-Y., Tien, C.-T., 2013. Driving behaviour modelling system based on graph construction. *Transportation research part C: emerging technologies* 26, 314–330.
- Cliff, A., 1981. Ord, jk,(1981), spatial processes: models and applications. Pion, London.
- Corman, F., D'Ariano, A., Pacciarelli, D., Pranzo, M., 2012. Bi-objective conflict detection and resolution in railway traffic management. *Transportation Research Part C: Emerging Technologies* 20 (1), 79–94.

- Di, X., Fabusuyi, T., Simek, C., Chen, X., Hampshire, R. C., 2019. Inferred Switching Behavior in Response to Re-entry of Uber and Lyft: A Revealed Study in Austin, TX. *Transport Findings*.
- Garrido, R. A., Mahmassani, H. S., 2000. Forecasting freight transportation demand with the space-time multinomial probit model. *Transportation Research Part B: Methodological* 34 (5), 403–418.
- Goulet-Langlois, G., Koutsopoulos, H. N., Zhao, Z., Zhao, J., 2018. Measuring regularity of individual travel patterns. *IEEE Transactions on Intelligent Transportation Systems* 19 (5), 1583–1592.
- Goulias, K. G., 1999. Longitudinal analysis of activity and travel pattern dynamics using generalized mixed markov latent class models. *Transportation Research Part B: Methodological* 33 (8), 535–558.
- Hampshire, R., Simek, C., Fabusuyi, T., Di, X., Chen, X., May 2017. Measuring the Impact of an Unanticipated Disruption of Uber/Lyft in Austin, TX. SSRN Scholarly Paper ID 2977969, Social Science Research Network, Rochester, NY.
- Hasnat, M. M., Hasan, S., 2018. Identifying tourists and analyzing spatial patterns of their destinations from location-based social media data. *Transportation Research Part C: Emerging Technologies* 96, 38–54.
- Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., Ratti, C., 2014. Geo-located twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science* 41 (3), 260–271.
- Hu, W., Jin, P. J., 2017. An adaptive hawkes process formulation for estimating time-of-day zonal trip arrivals with location-based social networking check-in data. *Transportation Research Part C: Emerging Technologies* 79, 136–155.
- Jiang, S., Fiore, G. A., Yang, Y., Ferreira, Jr., J., Fazzoli, E., González, M. C., 2013. A Review of Urban Computing for Mobile Phone Traces: Current Methods, Challenges and Opportunities. In: Proceedings of the 2Nd ACM SIGKDD International Workshop on Urban Computing. UrbComp '13. ACM, New York, NY, USA, pp. 2:1–2:9, event-place: Chicago, Illinois.
- Joh, C.-H., Arentze, T., Hofman, F., Timmermans, H., 2002. Activity pattern similarity: a multidimensional sequence alignment method. *Transportation Research Part B: Methodological* 36 (5), 385–403.
- Joh, C.-H., Arentze, T., Timmermans, H., Jan. 2001. Pattern Recognition in Complex Activity Travel Patterns: Comparison of Euclidean Distance, Signal-Processing Theoretical, and Multidimensional Sequence Alignment Methods. *Transportation Research Record* 1752 (1), 16–22.
- Krause, C. M., Zhang, L., 2018. Short-term travel behavior prediction with gps, land use, and point of interest data. *Transportation Research Part B: Methodological*.
- Kruskal, J. B., 1983. An overview of sequence comparison: Time warps, string edits, and macromolecules. *SIAM review* 25 (2), 201–237.
- Kurkcu, A., Ozbay, K., Morgul, E., 2016. Evaluating the usability of geo-located twitter as a tool for human activity and mobility patterns: A case study for nyc. In: Transportation Research Board's 95th Annual Meeting. pp. 1–20.
- Lavieri, P. S., Bhat, C. R., 2019a. A joint model of experience and frequency of use of solo and pooled ride-hailing services in dallas, texas. Tech. rep.
- Lavieri, P. S., Bhat, C. R., 2019b. A multivariate model of ride-hailing trip characteristics in dallas, tx. Tech. rep.
- Lehouillier, T., Omer, J., Soumis, F., Desaulniers, G., 2017. Two decomposition algorithms for solving a minimum weight maximum clique model for the air conflict resolution problem. *European Journal of Operational Research* 256 (3), 696–712.
- Ma, X., Wu, Y.-J., Wang, Y., Chen, F., Liu, J., 2013. Mining smart card data for transit riders' travel patterns. *Transportation Research Part C: Emerging Technologies* 36, 1–12.
- Martin, N. F. G., England, J. W., Jun. 2011. Mathematical Theory of Entropy. Cambridge University Press.
- McFadden, D., Train, K., 2000. Mixed mnl models for discrete response. *Journal of applied Econometrics* 15 (5), 447–470.
- Morikawa, T., Ben-Akiva, M., McFadden, D., 2002. Discrete choice models incorporating revealed preferences and psychometric data. In: Advances in Econometrics. Emerald Group Publishing Limited, pp. 29–55.
- Needleman, S. B., Wunsch, C. D., Mar. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology* 48 (3), 443–453.
- Ortega-Tong, M. A., 2013. Classification of london's public transport users using smart card data. Ph.D. thesis, Massachusetts Institute of Technology.
- Osorio, B., Kulesza, A., Hero, A. O., 2014. Multi-layer graph analysis for dynamic social networks. *IEEE Journal of Selected Topics in Signal Processing* 8 (4), 514–523.
- Pei, J., Han, J., Mortazavi-Asl, B., Pinto, H., Chen, Q., Dayal, U., Hsu, M.-C., 2001. Prefixspan: Mining sequential patterns efficiently by prefix-projected pattern growth. In: icccn. IEEE, p. 0215.
- Rashidi, T. H., Abbasi, A., Maghrebi, M., Hasan, S., Waller, T. S., Feb. 2017. Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges. *Transportation Research Part C: Emerging Technologies* 75, 197–211.
- Recker, W. W., 1995. The household activity pattern problem: general formulation and solution. *Transportation Research Part B: Methodological* 29 (1), 61–77.
- Shou, Z., Di, X., 2018. Similarity analysis of frequent sequential activity pattern mining. *Transportation Research Part C: Emerging Technologies* 96, 122–143.
- Song, C., Qu, Z., Blumm, N., Barabási, A.-L., Feb. 2010. Limits of Predictability in Human Mobility. *Science* 327 (5968), 1018–1021.
- Sun, L., Axhausen, K. W., 2016. Understanding urban mobility patterns with a probabilistic tensor factorization framework. *Transportation Research Part B: Methodological* 91, 511–524.
- Sun, L., Lee, D.-H., Erath, A., Huang, X., 2012. Using smart card data to extract passenger's spatio-temporal density and train's trajectory of mrt system. In: Proceedings of the ACM SIGKDD international workshop on urban computing. ACM, pp. 142–148.
- Tang, J., Song, Y., Miller, H. J., Zhou, X., 2016. Estimating the most likely space-time paths, dwell times and path uncertainties

- from vehicle trajectory data: A time geographic method. *Transportation Research Part C: Emerging Technologies* 66, 176–194.
- Toole Jameson L., Herrera-YaqÃije Carlos, Schneider Christian M., GonzÃ¡lez Marta C., Apr. 2015. Coupling human mobility and social ties. *Journal of The Royal Society Interface* 12 (105), 20141128.
- Wesolowski, A., Eagle, N., Noor, A. M., Snow, R. W., Buckee, C. O., 2013. The impact of biases in mobile phone ownership on estimates of human mobility. *Journal of the Royal Society Interface* 10 (81), 20120986.
- Wilson, W. C., 1998. Activity pattern analysis by means of sequence-alignment methods. *Environment and Planning A* 30 (6), 1017–1038.
- Wolfson, O., Sistla, A. P., Xu, B., 2012. The tranquyl language for data management in intelligent transportation. *Transportation Research Part C: Emerging Technologies* 23, 3–13.
- Wu, R.-C., Chen, R.-S., Chen, C.-C., 2005. Data mining application in customer relationship management of credit card business. In: Computer Software and Applications Conference, 2005. COMPSAC 2005. 29th Annual International. Vol. 2. IEEE, pp. 39–40.
- Xiao, Y., Lo, H. K., 2016. Day-to-day departure time modeling under social network influence. *Transportation Research Part B: Methodological* 92, 54–72.
- Ying, J. J.-C., Lu, E. H.-C., Lee, W.-C., Weng, T.-C., Tseng, V. S., 2010. Mining user similarity from semantic trajectories. In: Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks. ACM, pp. 19–26.
- Zhao, Z., Koutsopoulos, H. N., Zhao, J., 2018. Detecting pattern changes in individual travel behavior: A bayesian approach. *Transportation Research Part B: Methodological* 112, 73–88.