Regularity and Conformity: Location Prediction Using Heterogeneous Mobility Data

Yingzi Wang[⋆†], Nicholas Jing Yuan[†], Defu Lian[‡], Linli Xu[⋆], Xing Xie[†], Yong Rui[†]
[⋆]School of Computer Science, University of Science and Technology of China
[†]Microsoft Research
[‡]Big Data Research Center, University of Electronic Science and Technology of China
{v-ywang, nicholas.yuan, xingx, yongrui} at microsoft.com,
dove.ustc at gmail.com, linlixu at ustc.edu.cn

ABSTRACT

Mobility prediction enables appealing proactive experiences for location-aware services and offers essential intelligence to business and governments. Recent studies suggest that human mobility is highly regular and predictable. Additionally, social conformity theory indicates that people's movements are influenced by others. However, existing approaches for location prediction fail to organically combine both the *regularity* and *conformity* of human mobility in a unified model, and lack the capacity to incorporate *heterogeneous* mobility datasets to boost prediction performance. To address these challenges, in this paper we propose a hybrid predictive model integrating both the regularity and conformity of human mobility as well as their mutual reinforcement. In addition, we further elevate the predictive power of our model by learning *location profiles* from heterogeneous mobility datasets based on a gravity model. We evaluate the proposed model using several city-scale mobility datasets including location check-ins, GPS trajectories of taxis, and public transit data. The experimental results validate that our model significantly outperforms state-of-the-art approaches for mobility prediction in terms of multiple metrics such as accuracy and percentile rank. The results also suggest that the predictability of human mobility is time-varying, e.g., the overall predictability is higher on workdays than holidays while predicting users' unvisited locations is more challenging for workdays than holidays.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Data mining; H.2.8 [**Database Management**]: Spatial databases and GIS

General Terms

Algorithms, Experimentation, Performance

Keywords

location prediction, regularity, conformity, location profile, spatial influence, gravity model, collaborative filtering

KDD'15, Aug. 10–13, 2015, Sydney, Australia.

1. INTRODUCTION

Over the past decade, an overwhelming number of location-aware services and apps have profoundly changed the way people live, from route planning to dining and even social networking. Understanding user mobility thus becomes an essential factor for improving service quality and user engagement.

While sensing a user's current location provides the user with timely reactive experiences, e.g., "searching for the closest subway station", predicting users' future locations can enable appealing proactive experiences in various applications. For example, recently emerging digital assistants such as Microsoft Cortana[1] and Google Now[2] aim to push relevant information to users or help users accomplish tasks without their querying, e.g., pre-heating (or cooling) the house when the user is on the way home [30]. An accurate prediction of user mobility is hence crucial for such proactive services. As another example, mobility prediction brings business intelligence to advertising and marketing. Given potential high-value customers' future locations, advertisers/marketers can better choose locations for organizing promotion events or distributing advertisements and coupons. Predicting future mobility patterns of crowds can help governments deal with public emergencies such as stampede prevention[3].

The recent development of sensing technology and smart devices makes various types of mobility data available to the industry and researchers, such as GPS trajectories [23, 41], cellular tower data [13], WiFi signals [27, 34], smart card transactions [39], and location check-ins from online social networks [28, 4, 19], all of which facilitate the exploration of mobility understanding and prediction. For instance, using cellular tower data, Song et al. [33] show that the predictability of human mobility has a limit of 93%, which demonstrates that human mobility is highly regular and predictable. However, the actual prediction performance heavily depends on many aspects including data types, sampling frequency, and granularity of predictions [32]. Even the best results reported by state-of-the-art approaches are far below this limit [6, 19, 27].

To bridge the gap between actual prediction performance and theoretical limit, many challenges still remain to be addressed:

Regularity and Conformity. Several studies show that human mobility typically follows regular spatial-temporal patterns. In urban areas, people typically spend most of their time around several "major hubs", such as homes and workplaces [33], and periodically

[1] http://www.windowsphone.com/en-us/how-to/wp8/cortana/meet-cortana
[2] http://www.google.com/landing/now/
[3] http://en.wikipedia.org/wiki/2014_Shanghai_stampede

commute between them [18, 6]. Meanwhile, people frequently visit some "minor hubs" in a limited radius of their major hubs [9] at certain times, e.g., shopping malls, gyms, and restaurants. Nevertheless, human mobility is not only driven by *regularity*. People occasionally change their routines and visit some unfamiliar places, e.g., a bar recommended by friends or a popular restaurant on Yelp. Such irregular visits may be explicitly or implicitly influenced by others, usually a group of people who have similar social backgrounds, interests, and social statuses. This phenomenon is the so-called social *conformity* [7].

However, most existing approaches in location prediction typically fall into two categories: 1) developing individual mobility models, such as HMM [16], CRF [39], and periodic GMM [6], to capture users' regular behavioral patterns; 2) building collaborative models to leverage similar mobility patterns of different users [11, 26, 4]. Few studies have incorporated both the regularity and conformity of human mobility in predicting users' future locations. Although a few approaches have touched both factors to a certain extent, the main endeavor of these methods still focuses on a single factor, while the other one is typically used as side information or a constraint [19], Thus, the interdependency and mutual reinforcement of regularity and conformity are not fully exploited for location prediction.

Sparsity and Heterogeneity. Continuous and precise tracking of users' long-term movements (e.g., using GPS) is often energy-intensive and costly, while mobility data captured by low-energy sensing technologies is typically sparse in terms of either granularity (e.g., cellular tower data) or sampling frequency (e.g., location check-ins). Besides, a user's actual mobility is usually delineated in different forms of mobility data, where any single type only partially reveals a user's mobility patterns. However, existing models for location prediction lack the capacity to boost prediction accuracy with the help of heterogeneous mobility datasets. The difficulty lies in how to integrate the mobility patterns mined from heterogeneous mobility datasets into a unified prediction model .

In this paper, we tackle the above challenges by proposing a hybrid model called **RCH**, combining both **R**egularity and **C**onformity, and employing **H**eterogeneous mobility data for location prediction. Specifically, we introduce a mobility model containing a regularity term and a conformity term, where the conformity term is represented by a time-aware factorization model, and the regularity term is represented as interactions between users' hub visit patterns and spatial influence to users' visited venues (detailed in Sec. 3.2). The regularity and conformity terms interplay and reinforce each other. In particular, the spatial influence to venues are learned through a Gravity model (detailed in Sec. 3.3). Our main contributions are summarized as follows:

- We introduce a hybrid model for location prediction combing both the regularity and conformity of human mobility, which exploits the interdependent patterns of both routine visits and occasional visits.

- We develop a method to learn a location's profile from heterogeneous mobility datasets based on a gravity model, and integrate the learned location profiles into a time-aware prediction model.

- We evaluate our model for predicting location check-ins based on a large dataset containing 7,355,962 check-ins of 161,794 users, where the location profiles used in our model are learned from several extra city-scale heterogeneous mobility datasets, such as GPS trajectories of taxis and public transit data. The experimental results validate that our model significantly outperforms state-of-the-art methods in terms of multiple metrics such as prediction accuracy and percentile rank.

2. RELATED WORK

2.1 Predictability of Human Mobility

The increasing availability of mobility data provides marvelous potential to study human mobility patterns. A considerable number of works have shown that human mobility is regular, predictable and unique in both temporal and spatial spaces [35, 15, 8]. Observable regular movements among a few frequented locations, like home and work [9, 18], embody the regularity and predictability of human mobility. For example, using mobile phone logs of 100,000 users Gonzalez et al. [13] showed a high degree of mobility regularity among several highly visited haunts. Song et al. [33] demonstrated that a 93% potential predictability of mobility patterns of mobile phone users. Besides, de Montjoye et al. [8] quantified the mobility uniqueness and demonstrated four distinct points are enough to distinguish 95% of users.

In the past few years, several mobility prediction works have concentrated on human trajectory logs from personal mobile devices, smart cards, and vehicular digital records, like GPS data [23, 27, 1], wifi [23, 34] and bus-trip records [2, 39]. They have continual spatial and temporal mobility records and the conspicuous characteristic of periodical returning to some important places. Unlike these high frequency datasets, check-ins in LBSNs are usually sparse and sporadic [24]. Location prediction based on check-ins is more challenging than on dense datasets like GPS data [37].

2.2 Location Prediction Models

We summarize relevant mobility prediction models and their differences in Table 1 where we list the targeted mobility data (i.e., type of mobility to be predicted) and features incorporated in the models. According to whether the prediction model is trained independently among all users (i.e., whether a user's mobility model is learned from the user's own historical mobility alone), we categorize existing models into two types: individual models and collaborative models.

Table 1: Comparison of location prediction methods
CI: check-in, SMP: spatial mobility pattern, TC: text content
IT: individual temporal patterns, SR: social relationship
CF: collaborative filtering, HT: heterogeneous mobility datasets

methods	target			feature					
	CI	GPS	Wifi	SMP	TC	IT	SR	CF	HT
PSMM [6]	√			√		√	√		
W^4 [40]	√			√	√	√			
M5Tree[25]	√					√	√		
CEPR [19]	√			√		√		√	
SHM [12]	√					√	√		
gSCorr [11]	√					√	√		
DBN [26]	√					√	√		
NextPlace [27]		√	√			√			
WhereNext [23]		√				√			
Markov [1]		√							
RCH(Our Model)	√			√		√		√	√

2.2.1 Individual Models

Historical spatial-temporal mobility patterns are fundamental factors for inferring users' future locations, given the regularity of human mobility [27, 23, 40, 10, 39, 16]. An approach based on nonlinear time series is applied for mobility prediction in [27], which focused on predicting most important places. Using a GPS trajectory dataset generated by 17,000 cars, Monreale et al. [23] built a

decision tree, named T-pattern tree, to find the best match path and predict future movements. Yuan et al. [40] proposed a probabilistic model W^4 (who, when, where, what) unifying spatial, temporal and activity topics to model users' behaviors. Using public transit records, Yuan et al. [39] provided a constraint Conditional Random Field model and successfully inferred unknown alighting/boarding stops given part of them.

The advantage of individual models is that the regularity of human mobility can be well captured. However, the similarity of mobility patterns between different users is not considered and utilized for predicting future locations. Instead, our hybrid approach excavates similar users' mobility patterns based on social conformity and collaborative filtering in addition to learning users' regular mobility patterns with a time-aware sparse group Lasso model. Furthermore, we collectively learn location profiles using several heterogeneous mobility datasets generated by city-scale populations and integrate the location profiles into our hybrid prediction model.

2.2.2 Collaborative Models

Different people may have similar location preferences. Social relationships of users have been taken into account for location prediction and recommendation to relieve data sparsity [6, 12, 25, 26, 4]. For example, Noulas et al. [25] developed a supervised learning model for next place prediction considering location histories of users' friends. Cho et al. [6] introduced a time-aware Gaussian Mixture model considering both users' periodic mobility and social activities. Sadilek et al. [26] provided a Dynamic Bayesian Network model, combining friends' temporal information for location prediction. Nevertheless, social relationships are reported to offer a limited predictive power for location check-ins due to the high sparsity [24].

Collaborative filtering methods are widely applied in recommendation systems including location recommendations, which assumes that similar users have similar behavioral patterns like rating or purchasing. This assumption is also in accordance with the conformity theory in social psychology [7]. For example, matrix factorization has evolved as a critical algorithm in location recommendation [5, 22, 20], where a user's preference of a venue is modeled as an inner product of latent factors. Lian et al. [20] introduced a location recommendation model considering both users' latent preferences and geographical influence of locations, however, the influence is empirically determined, instead of learned from the data. Recently, probabilistic non-negative matrix factorization has also been adopted for location recommendations [21, 29], where the users' visited venues are considered observations of a generative process. However, these recommendation models cannot be directly applied for mobility prediction. Lian et al. [19] first employed collaborative filtering approaches for location prediction. Users' location visits are separated to explorations of novel or regular places based on a binary classification. They proposed a hidden Markov model for capturing regular mobility patterns and social-based collaborative filtering with 2D kernel density estimation to excavate novel mobility patterns. However, the antecedent division of mobility types through exploration prediction confronts the risk of two-layer errors.

In contrast to the above collaborative filtering approaches for location recommendation and location prediction, our method 1) simultaneously incorporates both regularity and conformity of human mobility in a unified prediction model, and utilizes the interplay between these two factors; 2) provides a time-aware collaborative model considering users' preference drifting at different time slots so as to enable time-aware location predictions; 3) learns spatial influence on venues using heterogeneous mobility datasets based on a gravity model and feeds it into the prediction model.

Table 2: Important Notations

Notation	Size	Description
$\mathbf{R}(t)$	$M \times N$	user-venue preference matrix at time slot t
$\mathcal{T}$	$1 \times T$	time slot set, t is a time slot in $\mathcal{T}$
$\mathcal{C}$	$1 \times I$	grid set, d is a grid in $\mathcal{C}$ with length I
$\mathcal{G}$	$1 \times G$	group set, g is a group in $\mathcal{G}$
$\mathcal{P}$	1×3	mobility type ={B(bus), A(taxi), C(check-in)}
$\mathcal{O}^*$	$1 \times I$	out-going flows of all grids in $\mathcal{C}$ w.r.t. $*\in \mathcal{P}$
$\mathcal{D}^*$	$1 \times I$	in-coming flows of all grids in $\mathcal{C}$ w.r.t. $*\in \mathcal{P}$
$\mathbf{T}^*$	$I \times I$	transition matrix of grids in $\mathcal{C}$ w.r.t. $*\in \mathcal{P}$
$\mathbf{U}$	$M \times K$	user stationary latent factor
$\mathbf{U}(t)$	$M \times K$	user changing latent factor of time slot t
$\mathbf{V}$	$N \times K$	venue latent factor
$\mathbf{H}(t)$	$M \times I$	hub matrix at time slot t
$\mathbf{Q}^*$	$N \times I$	sptial influence matrix w.r.t. $*\in \mathcal{P}$
$\mathbf{H}^{(g)}(t)$	$M \times L^{(g)}$	$\mathbf{H}(t)$'s sub-matrix of group g
$\mathbf{Q}^{*(g)}$	$N \times L^{(g)}$	$\mathbf{Q}^*$'s sub-matrix of group g

3. MODEL

3.1 Overview

Given visited venues of a group of users, our goal is to predict their future locations at a certain time. Let $\mathcal{U} = \{u_1, u_2, ..., u_M\}$ be M users and $\mathcal{V} = \{v_1, v_2, ..., v_N\}$ denote N venues. Note that here $\mathcal{V}$ may contain unvisited venues of users in U. We categorize days into two types, workdays and holidays, and let $\mathcal{T} = \{t_1, t_2, ..., t_T\}$ represent the T time slots in the two classes of days. Given a specific time slot t, we predict user u_i's location by calculating the mobility preferences of u_i to v_j for $j = 1, 2, \ldots, N$ at t, and returning the v_j that has the maximum mobility preference. Let $\mathbf{R}(t) \in \mathbb{R}^{M \times N}$ be the preference matrix of $\mathcal{U}$ to $\mathcal{V}$ at time t, i.e., $R_{ij}(t)$ indicates u_i's preference to v_j at t. As mentioned earlier, a user u_i's visit to a venue v_j can be driven by either regularity or conformity, i.e.,

$$R_{ij}(t) = R_{ij}^{(r)}(t) + R_{ij}^{(c)}(t), \tag{1}$$

where $R_{ij}^{(r)}(t)$ is the regularity term, indicating that v_j is a regular venue of u_i at time t; and $R_{ij}^{(c)}(t)$ is the conformity term, indicating that v_j is frequently visited by users who are similar to u_i at time t. These two factors can interplay and reinforce each other to drive u_i's visit to v_j. Next, we introduce both terms respectively as follows.

3.1.1 Regularity Term $\mathbf{R}^{(r)}$

For simplicity, in the rest of Sec. 3.1.1, we restrict our notations and description of the model to a specific time (without considering the time varying effect), and later in Sec 3.3, we will formulate the time-aware model.

To learn users' regular mobility patterns, we map users' visited venues to the geospatial space. Let $\mathcal{C} = \{d_1, d_2, ..., d_I\}$ be the C geographical grid cells (e.g., 100m$\times$ 100m) discretizing the whole geospatial space of a city. Each venue v_j is associated with a geo-coordinate $\{lat_j, lon_j\}$ belonging to a certain grid d_{k_j}, where lat_j, lon_j are the latitude and longitude of v_j. As shown in Fig. 1 a), u_i's visited venues are mapped to the grids shown as the ones with plus signs.

Consider the probability that u_i visits v_j in terms of regularity, denoted as $\Pr(v_j|u_i)$. We assume that v_j belongs to a grid d_{k_j}, and u_i travels from a grid d_k to v_j (note that it is possible that

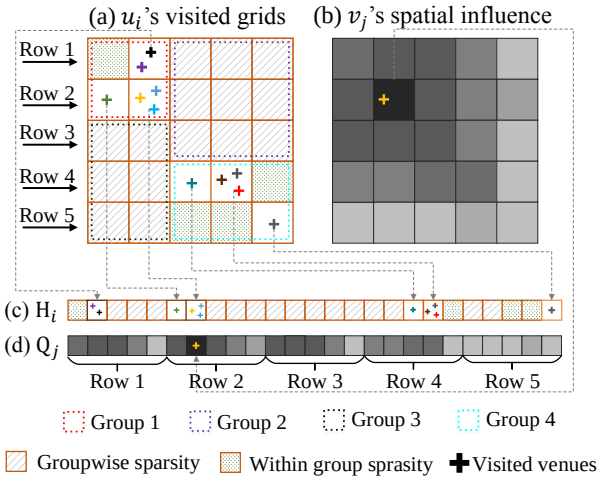


Figure 1: Learning the regularity term

$d_k = d_{k_j}$). By integrating over the geospatial grids and assuming the Markov property of users' transitions between grids, we have

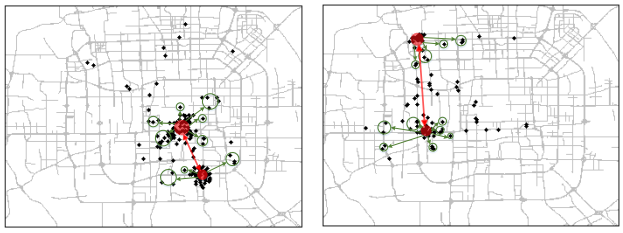
$$\Pr(v_j|u_i) \propto \sum_{k=1}^{I} \Pr(d_k|u_i)\Pr(v_j|d_k) \tag{2}$$

$$= \sum_{k=1}^{I} \Pr(d_k|u_i)\Pr(d_{k_j}|d_k)\Pr(v_j|d_{k_j}) \tag{3}$$

The first factor of the summation term in Eq. (3) can be estimated with the visiting frequency of a grid cell. As shown in Fig 1 a) and c), we flatten the 2D grid cells to a 1D vector $\mathbf{H}_i$, where H_{ik} is the visiting frequency of u_i to grid d_k, approximating $\Pr(d_k|u_i)$, and we term $\mathbf{H}$ the *hub matrix* of $\mathcal{U}$. The second factor $\Pr(d_{k_j}|d_k)$ is the transition probability from d_k to d_{k_j}, which is learned based on a gravity model using hetereneous mobility datasets, as detailed in Sec. 3.2. The third factor $\Pr(v_j|d_{k_j})$ can be estimated using the visiting frequency of v_j in grid d_{k_j}. We combine the second and third factor together as Q_{jk}, which represents the spatial influence of v_j to grid d_k as shown in Fig. 1 b) and d). Intuitively, Q_{jk} indicates the degree of influence that attracts users from grid d_k to v_j, and we refer to $\mathbf{Q}$ as the *spatial influence matrix*. Then, Eq. (3) can be re-written as $\mathbf{H}_i\mathbf{Q}_j^\top$.

Furthermore, in urban areas, the spatial distribution of users' visited venues often exhibits the hierarchical "multi-center" characteristics [6], which is an important aspect of regularity. For example, Fig. 2 plots the location check-ins of two users, where most check-ins gather around two major hubs, typically the work and home places. Meanwhile, the major hubs of a large population tend to form spatial agglomeration and foster different *functional zones* [38], such as residential areas and business districts. People actually commute between these functional zones to engage in various social-economic activities. As shown in Fig. 1 a), the agglomeration of major hubs can be represented by groups of grid cells in $\mathcal{C}$. Given heterogeneous mobility datasets such as public transit data and taxi GPS trajectories, Sec. 3.2 introduces a method to learn the group structure from commuting patterns. As a result, the I grid cells are clustered into G groups $\mathcal{G} = \{g_1, g_2, \ldots, g_G\}$. We note that users also frequently visit some minor hubs within the groups containing major hubs, e.g., shopping malls, restaurants, gyms, etc.

The regularity of human mobility indicates that both the major and minor hubs are sparsely distributed in the geospatial space, as exemplified in Fig. 2 where red circles are the major hubs and green



(a) check-ins of user A (b) check-ins of user B

Figure 2: Sparsity of major and minor hubs

ones are minor hubs. This leads to *groupwise sparsity* (i.e., only a few groups of grids have visited venues of u_i), and *within-group sparsity* (i.e., only a few grids are visited for the groups containing u_i' major hubs) of the vector $\mathbf{H}_i$, as illustrated in Fig. 1. To preserve both sparsities, Sec. 3.3 presents a sparse group Lasso-based model to constrain the regularity term, given by

$$R_{i,j}^{(r)} = \sum_g \mathbf{H}_i^{(g)} \left(\mathbf{Q}_j^{(g)}\right)^\top, \tag{4}$$

where g is a group in $\mathcal{G}$.

3.1.2 *Conformity Term* $\mathbf{R}^{(c)}$

In addition to the regular daily movements among several venues, people sometimes seek and visit novel locations. The social conformity theory [7] suggests that users who have similar backgrounds, interests, and social statuses often behave similar to each other, which is also the psychological root of collaborative filtering. A widely adopted approach [5, 17] is to factorize users' location preferences in terms of conformity $\mathbf{R}^{(c)}$ into two low dimensional latent matrices $\mathbf{U}$ and $\mathbf{V}$, where $\mathbf{U}_i$ and $\mathbf{V}_j$ are latent factors of user i and POI j, both with dimension K. Nevertheless, the preferences of people to venues may vary at different times. Sitting in a café enjoying coffee and sunshine is a wonderful choice for mornings but a bar is suitable for nights in most cases. We add the time changing part of user latent factor $\mathbf{U}_i(t)$ to describe changeable preferences, and $\mathbf{U}_i$ is the stationary part for unchangeable interests for venues. The conformity term is transformed to

$$R_{i,j}^{(c)}(t) = (\mathbf{U}_i + \mathbf{U}_i(t))\mathbf{V}_j^\top. \tag{5}$$

Next, we introduce how to learn the spatial influence matrix $\mathbf{Q}$ and the group structure $\mathcal{G}$ in Sec. 3.2, and formally propose the RCH model as well as the optimization algorithms in Sec. 3.3.

3.2 Location Profiling Based on Gravity Model

Location profiling consists of two tasks: 1) estimating the *spatial influence matrix* $\mathbf{Q} \in \mathbb{R}^{N\times I}$, where the element in the jth row and kth column Q_{jk} is the spatial influence of venue v_j to grid d_k; and 2) learning the group structure of the hub matrix $\mathbf{H}$.

The spatial influence of venue v_j to grid d_k is comprised of two factors (refer to Eq. (3)): the transition probability from grid d_{k_j} (the grid containing v_j) to grid d_k and the visiting frequency of v_j in d_{k_j}. The latter one is easier to obtain since we can estimate it by calculating the visiting frequency at v_j within grid d_k and conducting a kernel density estimation to obtain a smooth distribution, given a mobility dataset such as location check-ins and GPS trajectories. However, a precise estimation of the first factor can be tricky and subtle. To address this issue, we employ and adapt a gravity model [36][3], which is widely adopted in mobility analytics for a large population, to estimate transition probability using heterogeneous mobility datasets.

Let $\mathcal{O}_i^*$ be the number of individuals leaving grid d_i, for $i = 1, 2, ..., I$, and $\mathcal{D}_j^*$ be the number of people going towards grid d_j, for $j = 1, 2, \ldots, I$, where $* \in \mathcal{P}$ indicates a certain type of mobility. In this work, we consider three types of mobility: B(bus), C(check-in), and A(taxi), i.e., $\mathcal{P} = \{B, A, C\}$. The gravity model states that the commuting flows from grid i to grid j with respect to a certain mobility type, denoted as $T_{i,j}^*$, is determined by $\mathcal{O}_i^*$, $\mathcal{D}_j^*$, and the distance between the two grids, through a gravity-like law:

$$T_{i,j}^* = c\frac{(\mathcal{O}_i^*)^a(\mathcal{D}_j^*)^b}{\exp(r \cdot dis_{i,j})}, \quad * \in \{B, A, C\}, \tag{6}$$

where c is a constant, a and b are coefficients of $\mathcal{O}_i^*$ and $\mathcal{D}_j^*$ respectively, and r tunes the decay by distance. As a result, $(T_{i,j}^B, T_{i,j}^A, T_{i,j}^C)$ is a vector describing heterogeneous commuting flows from grid d_i to grid d_j.

Our goal is to estimate the coefficients a, b, r by fitting this model using observed mobility data. We achieve it using the multivariate regression method. Applying a logarithmic transformation to both sides of (6), we obtain the following expression:

$$\ln T_{i,j}^* = a \ln \mathcal{O}_i^* + b \ln \mathcal{D}_j^* - r \cdot dis_{i,j} + \ln c. \tag{7}$$

For readability, we flatten the *transition matrix* $\mathbf{T}^*$ as a vector $\mathbf{y} = (y_1, y_2, ..., y_{I^2})^\top$ where $y_{(i-1)I+j} = T_{i,j}^*$, and $\mathbf{X} \in \mathbb{R}^{I^2 \times 4}$ denotes the regressors, given by

$$\mathbf{X} = \begin{pmatrix} 1 & X_{11} & \cdots & X_{13} \\ 1 & X_{21} & \cdots & X_{23} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & \cdots & X_{n3} \end{pmatrix}, \tag{8}$$

where $n = I^2$, $X_{(i-1)I+j,2} = dis_{i,j}$, $X_{(i-1)I+j,3} = \ln \mathcal{D}_j^*$, and $X_{(i-1)I+j,4} = \ln \mathcal{O}_i^*$. Let $\boldsymbol{\beta}$ be a vector containing all parameters $(\ln c, -r, b, a)^\top$. We propose to estimate $\boldsymbol{\beta}$ by:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{9}$$

where $\boldsymbol{\epsilon}$ is the error vector.

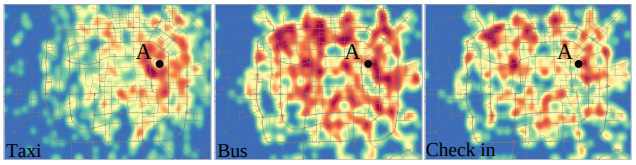
We employ least square estimation to estimate the parameter $\hat{\boldsymbol{\beta}}$, then $\hat{y_i} = \mathbf{X}_i\hat{\boldsymbol{\beta}}$ is the vector of the estimated transitions. Our target is now to find a $\hat{\boldsymbol{\beta}}$ to minimize the sum of squared residuals:

$$SSE = \sum_{i=1}^{n} \hat{\boldsymbol{\epsilon}}_i^2 = \sum_{i=1}^{n} (\mathbf{y}_i - \hat{\mathbf{y}}_i)^2 . \tag{10}$$

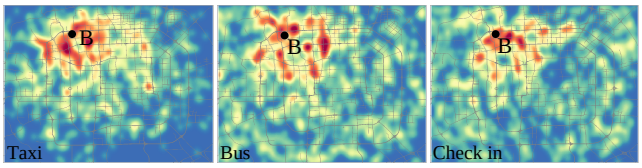
The value of $\hat{\boldsymbol{\beta}}$ that minimizes SSE in (10) is given by:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{y}. \tag{11}$$

As a result, we obtain the estimated and row-normalized [14] transition matrix $\hat{\mathbf{T}}^*$ with respect to a certain mobility type $* \in \mathcal{P}$. Fig. 3 a) and b) visualize the estimated transitions to two grids (A and B) using different mobility data sources, where grid A is in a bar district (Sanlitun) and grid B is in an IT district (Zhongguancun). We observe that transitions estimated using different mobility datasets show similar influence to certain areas (such as B) and discrepant influence on some other areas (such as A). We note that for the within-grid transitions, i.e., $d_k = d_{k_j}$, users' movements may not be observable from taxi and bus data when the grid size is too small (since people usually take a bus/taxi for commuting when the travel distance is longer than a walking distance). However, the check-in data can reflect such within-group transitions, which also suggests the necessity to use heterogeneous data for learning the



(a) spatial influence to grid A (a bar district) from other grids



(b) spatial influence to grid B (an IT district) from other grids

Figure 3: Spatial influence to a grid estimated based on a gravity model using taxi, bus, and check-in data

spatial influence on a venue. Combined with the estimated visiting frequency of v_j in d_{k_j}, denoted as ρ_j, we have the final spatial influence

$$Q_{jk}^* = \rho_j \hat{T}_{k,k_j}^*. \tag{12}$$

Next, we learn the group structure of users' major hubs. Given the estimated transition matrix $\hat{\mathbf{T}}^*$ for $* \in \mathcal{P}$, we employ the Dirichlet Multinomial Regression (DMR) based model [38] to learn the group structure (also known as the *functional zones*) of a city. Specifically, for grid d_i, $i = 1, \ldots, I$, we extract all out-going transitions $\hat{\mathbf{T}}_{i\cdot}^*$ and in-coming transitions $\hat{\mathbf{T}}_{\cdot i}$, which are considered "words" of a document, and learn the "topic" distribution through the DMR topic model, where the distribution of venues in grid d_i is deemed as meta-data of the document (such as author and email). Later, we cluster the different grids according to the learned topic distributions. As a result, the grids in a cluster are regarded as a group and we obtain a list of groups $\mathcal{G} = \{g_1, \ldots, g_G\}$. Since the spatial influence matrix $\mathbf{Q}^*$ is with dimension $N \times I$, we use the groups $\mathcal{G}$ to re-order the columns of $\mathbf{Q}^*$ so that the columns belonging to a group are put together. Then, for each group $g \in G$, we have a sub-matrix $\mathbf{Q}^{*(g)} \in \mathbb{R}^{N \times L^{(g)}}$, where $L^{(g)}$ is the length of group g.

3.3 The RCH Model

3.3.1 Model Formulation

We learn the predictor $R_{i,j}(t)$ using a supervised learning approach by constructing an optimization problem. As explained in Sec. 3.1.2, if we only consider the conformity term (refer to Eq. (5)), our problem can be solved using a time-aware matrix factorization model, given by

$$\min_{\mathbf{U}, \mathbf{U}(t), \mathbf{V}} \sum_{t \in \mathcal{T}} \|\mathbf{R}(t) - (\mathbf{U} + \mathbf{U}(t))\mathbf{V}^\top\|_F^2 + \gamma(\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2) + \beta \sum_{t \in \mathcal{T}} \|\mathbf{U}(t)\|_F^2, \tag{13}$$

where $\mathbf{U} \in \mathbb{R}^{M \times K}$ is the matrix of users' stationary latent factor and $\mathbf{U}(t)$ is the matrix of users' time-varying latent factor. The regularizations of $\mathbf{U}, \mathbf{U}(t), \mathbf{V}$ are added to reduce generalization errors and avoid over-fitting, and β balances the regularization terms between the stationary and time-varying latent factors.

To incorporate the regularity term, we need to learn a hub matrix $\mathbf{H}(t)$, which represents spatial distribution of users' regular venues at t. We group the columns of the hub matrix $\mathbf{H}$ in the same way as

$\mathbf{Q}^*$, based on the group structure learned from Sec 3.2 to obtain G sub-matrices. The inner product of $\mathbf{H}_i^{(g)}(t)$ and $\mathbf{Q}_j^{*(g)}$ is the regularity term of user u_i to venue v_j for grids in group g. The three spatial influence matrices $\mathbf{Q}^B$, $\mathbf{Q}^A$ and $\mathbf{Q}^C$ are linearly combined to affect the hub matrices with a parameter vector $\theta^B, \theta^A, \theta^C$. Combining both the regularity term and conformity term, our final objective function is

$$\begin{aligned}
&\Theta(\mathbf{U}, \mathbf{U}(t), \mathbf{V}, \mathbf{H}(t), \theta^B, \theta^A) \\
&= \sum_{t\in\mathcal{T}} \|\mathbf{R}(t) - \sum_{g\in\mathcal{G}} \mathbf{H}^{(g)}(t) \left(\sum_{*\in\mathcal{P}} \theta^* \mathbf{Q}^{*(g)}\right)^\top - (\mathbf{U} + \mathbf{U}(t))\,\mathbf{V}^\top\|_F^2 \\
&\quad + \sum_{t\in\mathcal{T}} ((1-\alpha)\sigma \sum_{j=1}^{M} \sum_{g\in\mathcal{G}} \|\mathbf{H}_j^{(g)}(t)\|_2 + \alpha\sigma \sum_{j=1}^{M} \|\mathbf{H}_j(t)\|_1) \\
&\quad + \gamma(\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2) + \beta \sum_{t\in\mathcal{T}} \|\mathbf{U}(t)\|_F^2, \qquad (14)
\end{aligned}$$

where $* \in \mathcal{P}$ and $\theta^C = 1$.

The l_2-norm of $\mathbf{H}^{(g)}(t)$ shrinks the number of groups, and offers *group-wise sparsity*. We use the l_1 norm of $\mathbf{H}(t)$, which preserves *within-group sparsity*, to reduce the number of nonzero elements in a group. In this way, we maintain the sparsity of major and minor hubs, as shown in Sec 3.1.1.

We note that in the above final model, the regularity term also implies certain conformity since heterogeneous mobility data contains mobility patterns of different groups of users; the conformity term also implies some regularity, since it captures regular temporal patterns of users' latent preferences. Therefore, regularity and conformity actually interplay and reinforce each other to co-influence the mobility model.

3.3.2 Optimization

We use alternative minimization to learn all parameters:

- **Optimization of U:**

When $\mathbf{U}(t), \mathbf{V}, \mathbf{H}(t), \theta^B$ and θ^A are fixed, the optimization problem is equivalent to the minimization of:

$$\Theta_{\mathbf{U}} = \sum_{t\in\mathcal{T}} \|\mathbf{R}(t) - (\mathbf{U} + \mathbf{U}(t))\,\mathbf{V}^\top - \mathbf{P}(t)\|_F^2 + \gamma\|\mathbf{U}\|_F^2, \qquad (15)$$

where $\mathbf{P}(t) = \sum_{g\in\mathcal{G}} \mathbf{H}^{(g)}(t) \left(\tilde{\mathbf{Q}}^{(g)}\right)^\top$ and $\tilde{\mathbf{Q}}^{(g)} = \sum_{*\in\mathcal{P}} \theta^* \mathbf{Q}^{*(g)}$
We update $\mathbf{U}$ by solving a least-square problem: the gradient of $\Theta_{\mathbf{U}}$ with respect to $\mathbf{U}$ is:

$$\nabla\Theta_{\mathbf{U}} = 2\sum_{t\in\mathcal{T}} \left(\mathbf{R}(t) - (\mathbf{U} + \mathbf{U}(t))\,\mathbf{V}^\top - \mathbf{P}(t)\right)(-\mathbf{V}) + 2\gamma\mathbf{U}. \qquad (16)$$

Let $\nabla\Theta_{\mathbf{U}} = 0$ and we get the update rule of $\mathbf{U}$:

$$\mathbf{U} = \sum_{t\in\mathcal{T}} \left(\mathbf{R}(t) - \mathbf{U}(t)\mathbf{V}^\top - \mathbf{P}(t)\right)\mathbf{V}\left(\gamma\mathbf{I} + T \times \mathbf{V}^\top\mathbf{V}\right)^{-1}, \qquad (17)$$

where T is the number of time slots.

- **Optimization of U(t):**

For a time changing part of user latent factor of time slot τ, when $\mathbf{U}, \mathbf{V}, \mathbf{H}(t)(t \in \mathcal{T})$, $\mathbf{U}(t)(t \in \mathcal{T}, t \neq \tau)$, θ^B and θ^A are fixed, the problem is transformed to the minimization of:

$$\Theta_{\mathbf{U}(\tau)} = \|\mathbf{R}(\tau) - (\mathbf{U} + \mathbf{U}(\tau))\,\mathbf{V}^\top - \mathbf{P}(t)\|_F^2 + \beta\|\mathbf{U}(\tau)\|_F^2. \qquad (18)$$

The gradient of $\Theta_{\mathbf{U}(\tau)}$ with respect to $\mathbf{U}(\tau)$ is:

$$\nabla\Theta_{\mathbf{U}(\tau)} = 2\left(\mathbf{R}(\tau) - (\mathbf{U} + \mathbf{U}(\tau))\,\mathbf{V}^\top - \mathbf{P}(t)\right)(-\mathbf{V}) + 2\beta\mathbf{U}(\tau). \qquad (19)$$

The changing part of user latent factor is updated as:

$$\mathbf{U}(\tau) = \left(\mathbf{R}(\tau) - \mathbf{U}\mathbf{V}^\top - \mathbf{P}(t)\right)\mathbf{V}\left(\beta\mathbf{I} + \mathbf{V}^\top\mathbf{V}\right)^{-1}. \qquad (20)$$

- **Optimization of V:**

Similarly, when $\mathbf{U}$, $\mathbf{U}(t)$, $\mathbf{H}(t)$, θ^B and θ^A are fixed, we set $(\mathbf{U} + \mathbf{U}(t))$ as $\tilde{\mathbf{U}}$, the latent factor is updated as:

$$\mathbf{V} = \sum_{t\in\mathcal{T}} \left(\mathbf{R}(t)^\top - \sum_{g\in\mathcal{G}} \tilde{\mathbf{Q}}^{(g)} \left(\mathbf{H}^{(g)}(t)\right)^\top\right)\tilde{\mathbf{U}}\left(\gamma\mathbf{I} + \tilde{\mathbf{U}}^\top\tilde{\mathbf{U}}\right)^{-1}. \qquad (21)$$

- **Optimization of θ^B and θ^A:**

The optimization problem is equivalent to minimizing:

$$\Theta(\theta^B, \theta^A) = \sum_{t\in\mathcal{T}} \|\tilde{\mathbf{Q}}^{(0)}(t) - \theta^B\tilde{\mathbf{Q}}^{(1)}(t) - \theta^A\tilde{\mathbf{Q}}^{(2)}(t)\|_F^2, \qquad (22)$$

where

$$\begin{cases} \tilde{\mathbf{Q}}^{(0)}(t) = & \mathbf{R}(t) - (\mathbf{U} + \mathbf{U}(t))\,\mathbf{V}^\top - \sum_{g\in\mathcal{G}} \mathbf{H}^{(g)}(t)(\mathbf{Q}^{C(g)})^\top, \\ \tilde{\mathbf{Q}}^{(1)}(t) = & \sum_{g\in\mathcal{G}} \mathbf{H}^{(g)}(t)(\mathbf{Q}^{B(g)})^\top, \\ \tilde{\mathbf{Q}}^{(2)}(t) = & \sum_{g\in\mathcal{G}} \mathbf{H}^{(g)}(t)(\mathbf{Q}^{A(g)})^\top. \end{cases} \qquad (23)$$

Hence, we obtain

$$\Theta(\theta^B, \theta^A) = \sum_{t\in\mathcal{T}} \sum_{i,j} (\tilde{Q}_{i,j}^{(0)}(t) - \theta^B\tilde{Q}_{i,j}^{(1)}(t) - \theta^A\tilde{Q}_{i,j}^{(2)}(t))^2. \qquad (24)$$

The Hessian matrix of $\Theta(\theta^B, \theta^A)$ is:

$$\nabla^2\Theta = \begin{pmatrix} \frac{\partial^2\Theta}{\partial^2\theta^B} & \frac{\partial^2\Theta}{\partial\theta^B\partial\theta^A} \\ \frac{\partial^2\Theta}{\partial\theta^A\partial\theta^B} & \frac{\partial^2\Theta}{\partial^2\theta^A} \end{pmatrix} = \begin{pmatrix} 2\sum_{(t)}^{1,1} & 2\sum_{(t)}^{1,2} \\ 2\sum_{(t)}^{1,2} & 2\sum_{(t)}^{2,2} \end{pmatrix}, \qquad (25)$$

where $\sum_{(t)}^{p,q} = \sum_{t\in\mathcal{T}} \sum_{i,j} \tilde{Q}_{i,j}^{(p)}(t)\tilde{Q}_{i,j}^{(q)}(t)$.
By applying *Cauchy-Schwarz* inequality, we have

$$\frac{\partial^2\Theta}{\partial^2\theta^B} \geq 0 \text{ and } |\nabla^2\Theta| \geq 0, \qquad (26)$$

implying that $\nabla^2\Theta$ is positive-definite and $\Theta(\theta^B, \theta^A)$ is convex.
Therefore, the estimations of θ^B and θ^A can be computed by:

$$\begin{cases} \frac{\partial\Theta}{\partial\theta^B} = 0 \\ \frac{\partial\Theta}{\partial\theta^A} = 0 \end{cases} \Rightarrow \begin{cases} \theta^B = \frac{\sum_{(t)}^{0,1}\sum_{(t)}^{2,2} - \sum_{(t)}^{0,2}\sum_{(t)}^{1,2}}{\sum_{(t)}^{1,1}\sum_{(t)}^{2,2} - \sum_{(t)}^{1,2}\sum_{(t)}^{1,2}} \\ \theta^A = \frac{\sum_{(t)}^{0,2}\sum_{(t)}^{1,1} - \sum_{(t)}^{0,1}\sum_{(t)}^{1,2}}{\sum_{(t)}^{1,1}\sum_{(t)}^{2,2} - \sum_{(t)}^{1,2}\sum_{(t)}^{1,2}} \end{cases} \qquad (27)$$

- **Optimization of H(t):**

Fixing $\mathbf{U}$, $\mathbf{U}(t)(t \in \mathcal{T})$, $\mathbf{V}$, θ^B and θ^A, we denote

$$\left(\mathbf{R}(t) - (\mathbf{U} + \mathbf{U}(t))\,\mathbf{V}^\top\right)$$

as $\tilde{\mathbf{R}}(t)$. For a given time slot τ, when $\mathbf{H}(t)$, $(t \in \mathcal{T}$ and $t \neq \tau)$ are fixed, Eq. (14) is equivalent to:

$$\begin{aligned}
\Theta_{\mathbf{H}(\tau)} =& \|\tilde{\mathbf{R}}(\tau) - \sum_{g\in\mathcal{G}} \mathbf{H}^{(g)}(\tau) \left(\tilde{\mathbf{Q}}^{(g)}\right)^\top\|_F^2 \\
&+ (1-\alpha)\,\sigma \sum_{j=1}^{M} \sum_{g\in\mathcal{G}} \|\mathbf{H}_j^{(g)}(\tau)\|_2 + \alpha\sigma \sum_{j=1}^{M} \|\mathbf{H}_j^{(g)}(\tau)\|_1. \qquad (28)
\end{aligned}$$

For a $M \times N$ matrix $\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_M \end{pmatrix}$ ($\mathbf{X}_j$ is the jth row of $\mathbf{X}$), we have

$$\|\mathbf{X}\|_F^2 = \sum_{j=1}^{M} \mathbf{X}_j^2 = \sum_{j=1}^{M} \|\mathbf{X}_j\|_F^2. \quad (29)$$

Hense, we can rewrite Eq. (28) as:

$$\Theta_{\mathbf{H}(\tau)} = \sum_{j=1}^{M} (\|\tilde{\mathbf{R}}_j(\tau) - \sum_{g\in\mathcal{G}} \mathbf{H}_j^{(g)}(\tau) \left(\tilde{\mathbf{Q}}^{(g)}\right)^\top \|_F^2 + (1-\alpha)\,\sigma \sum_{g\in\mathcal{G}} \|\mathbf{H}_j^{(g)}(\tau)\|_2 + \alpha\sigma \|\mathbf{H}_j(\tau)\|_1). \quad (30)$$

Assuming that the regular patterns of users are independent, every row in $\mathbf{H}(\tau)$ is uncorrelated with the other rows. Therefore,Eq. (30) is equivalent to:

$$\Theta_{\mathbf{H}(\tau)} = \sum_{j=1}^{M} \min_{\mathbf{H}(\tau)} (\|\tilde{\mathbf{R}}_j(\tau) - \sum_{g\in\mathcal{G}} \mathbf{H}_j^{(g)}(\tau) \left(\tilde{\mathbf{Q}}^{(g)}\right)^\top \|_F^2 + (1-\alpha)\,\sigma \sum_{g\in\mathcal{G}} \|\mathbf{H}_j^{(g)}(\tau)\|_2 + \alpha\sigma \|\mathbf{H}_j(\tau)\|_1). \quad (31)$$

Our goal now is to find an $\mathbf{H}_j(\tau)$ to minimize:

$$\Theta_{\mathbf{H}_j(\tau)} = \|\tilde{\mathbf{R}}_j(\tau) - \sum_{g\in\mathcal{G}} \mathbf{H}_j^{(g)}(\tau) \left(\tilde{\mathbf{Q}}^{(g)}\right)^\top \|_F^2 + (1-\alpha)\,\sigma \sum_{g\in\mathcal{G}} \|\mathbf{H}_j^{(g)}(\tau)\|_2 + \alpha\sigma \|\mathbf{H}_j(\tau)\|_1, \quad (32)$$

for $j = 1, \ldots, M$.

This problem is similar to the sparse group Lasso problem [31]. We leave the detailed derivation to the appendix[4] and give the update rules as follows:

1. Check whether $\|\mathscr{F}(\tilde{\mathbf{R}}_j^{(-g)}(\tau)\tilde{\mathbf{Q}}^{(g)}, \alpha\sigma)\|_2 \le (1-\alpha)\,\sigma$. Here,

$$\tilde{\mathbf{R}}_j^{(-g)}(\tau) = \tilde{\mathbf{R}}_j(\tau) - \sum_{l\neq g} \mathbf{H}_j^{(l)}(\tau)(\tilde{\mathbf{Q}}^{(l)})^\top, \quad (33)$$

and $\mathcal{F}$ is a soft thresholding operator, defined as:

$$(\mathscr{F}(\mathbf{x}, y))_i = \text{sign}(\mathbf{x}_i)(|\mathbf{x}_i| - y)_+ \quad (34)$$

for $i = 1, 2, \ldots, N$, where $z_+ = \max(z, 0)$. If so, assign $\mathbf{H}_j^{(g)}(\tau) = \mathbf{0}$.

2. If not, iteratively update $\mathbf{H}_j^{(g)}(\tau)$ by

$$\mathscr{H}(\mathbf{H}_j^{(g)}(\tau)) = \left(1 + \frac{(1-\alpha)\sigma}{\|\mathscr{F}\left(\tilde{\mathbf{H}}_j^{(g)}(\tau), \alpha\sigma\right)\|_2}\right) \mathscr{F}\left(\tilde{\mathbf{H}}_j^{(g)}(\tau), \alpha\sigma\right). \quad (35)$$

Here,

$$\tilde{\mathbf{H}}_j^{(g)}(\tau) = \phi \mathbf{H}_j^{(g)}(\tau) - \nabla\mathscr{G}(\mathbf{H}_j^{(g)}(\tau)), \quad (36)$$

[4] http://bit.ly/1yYaD4o or http://t.cn/RwSiej5 (for Chinese users)

Algorithm 1: Optimization of RCH Model

Input: $\alpha, \beta, \gamma, \sigma, \tilde{\mathbf{R}}(t)(t \in \mathcal{T}), \mathbf{Q}^*(* \in \mathcal{P})$
Output: $\mathbf{U}, \mathbf{U}(t), \mathbf{V}, \mathbf{H}(t), \theta^B$ and θ^A minimizing Θ in (32)

```
U, U(t), V, H(t), θ^B, θ^A ⇐ U_0, U_0(t), V_0, H_0(t), θ_0^B, θ_0^A ;
repeat
    update U with (17);
    update V with (21);
    update θ^B and θ^A with (27);
    for τ = 1, 2, ..., T do
        update U(t) with (20);
        for j = 1, 2, ..., M do
            for g = 1, 2, 3, ..., G do
                if ||F(R̃_j^(-g)(τ) Q̃^(g), ασ)||_2 ≤ (1-α) σ then
                    H_j^(g)(τ) = 0;
                else
                    H_j^(g)(τ) ⇐ H(H_j^(g)(τ))
until convergence;
return U, U(t), V, H(t), θ^B and θ^A
```

where ϕ is a sufficiently large number, $\nabla\mathscr{G}(\mathbf{H}_j^{(g)}(\tau)')$ is the gradient of $\mathscr{G}(\mathbf{H}_j^{(g)}(\tau))$ with respect to $\mathbf{H}_j^{(g)}(\tau)$, and

$$\mathscr{G}(\mathbf{H}_j^{(g)}(\tau)) = \|\tilde{\mathbf{R}}_j(\tau) - \sum_{g\in\mathcal{G}} \mathbf{H}_j^{(g)}(\tau) \left(\tilde{\mathbf{Q}}^{(g)}\right)^\top \|_2. \quad (37)$$

Finally, we provide the pseudo-code of the whole optimization procedure in Algorithm 1. The time complexities of Eq. (17)(20) and (21) are $O(TMN(\max\{K, I\}))$. The time complexities of Eq. (27) and (28) are $O(TMN)$ and $O(M\sum_{g\in\mathcal{G}} NL^{(g)})$ (which equals to $O(MNI)$) respectively. If the number of iterations is S, the time complexity of the whole optimization for Eq. (14) is

$$O(STMN(\max\{K, I\})).$$

4. EXPERIMENT

4.1 Settings

4.1.1 Dataset

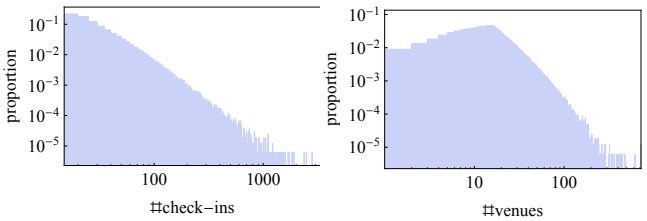
We utilized three real-world mobility datasets in our experiments, where the first one is the targeted mobility to be predicted, and the latter two are used together to learn location profiles (refer to Sec. 3.2).

• **Check-in Dataset.** We crawled 12,133,504 location check-ins at 146,962 venues in Beijing from 1,765,633 users through Sina Weibo (China's Twitter) API[5] from March. 2011 to Sep. 2013. After removing users who have less than 15 check-ins, we obtain 7,355,962 check-ins at 118,534 venues from 161,794 users where each user has 45 check-ins and visits 20 venues on average. We plot the Log-Log histograms in Fig. 4, where the number of check-ins shows a power-law distribution. Each check-in contains the user ID, check-in time, venue ID and the venue's geo-coordinates.

• **Bus Dataset.** This dataset contains 3 million bus-trip records from August 2012 to May 2013 in Beijing. Every trip includes the card ID, alighting time, boarding and alighting stops of the trip.

• **Taxi Dataset.** This data set contains GPS trajectories of Beijing taxis from March 2011 to August 2011. We segmented the trajectories into 19.4 million taxi transitions. Each transition includes the

[5] http://bit.ly/1rgzRch



(a) distibution of #check-in (b) distribution of #venues

Figure 4: Statistics of location check-in data

boarding and alighting times as well as the geo-coordinates of the origin and destination.

4.1.2 Metrics

In our experiments, we learn the models and obtain the matrix $\mathbf{R}(t)$ for time slot t. Each row of the matrix is the predicted scores for venues of a user. We use two metrics to measure the performance of location prediction: prediction accuracy (Acc@top$P$) and the average percentile rank (APR) of the actually visited venues.

Acc@top$P$ is the percentage of accurate predictions for a list of predictions with length P. The percentile rank of prediction for venue v_j [25] is defined as:

$$PR = \frac{N - rank(v_j) + 1}{N},$$

where $rank(v_j)$ is the position of venue v_j in the predicted list and N is the number of venues. It is clear that PR is 1 if the true venue is ranked as the first. Average Percentile Rank (APR) is the average of PR over the testing set.

4.1.3 Baselines

We compare the RCH model with 4 baselines, CEPR [19], W^3 (W^4 in [40] without "what"), PMM [6] and MF (Most Frequent).

• MF. Most Frequent Model calculates the frequencies of user u_i's checking in at different venues at time t based on u_i's location history, and assigns the most frequently visited venue as the predicted venue given time t.

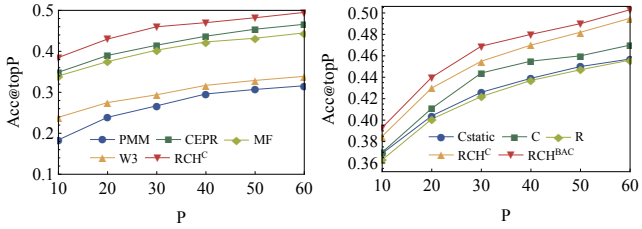
• PMM. Periodic Mobility Model employs a Gaussian mixture model centered at "home" and "work" to learn user locations, which are divided into two states, home state and work state, modeled by a 2-dimensional time-independent spatial Gaussian distribution. PMM randomly initializes the state of each location and the parameters are estimated using the EM algorithm.

• W^3. W^4 (who, when, where, and what) is a probabilistic mobility model based on interactions between user regions, locations, personal topics etc. It can be generally applied to various applications such as location prediction of a tweet, location prediction of a user and time. We only make comparisons with respect to the location prediction for a user, which is calculated by:

$$P(v_j|u_i, s, t) = \frac{\sum_z \sum_r P(u_i, s, t, r, z, v_j)}{\sum_z \sum_r \sum_k P(u_i, s, t, r, z, v_k)},$$

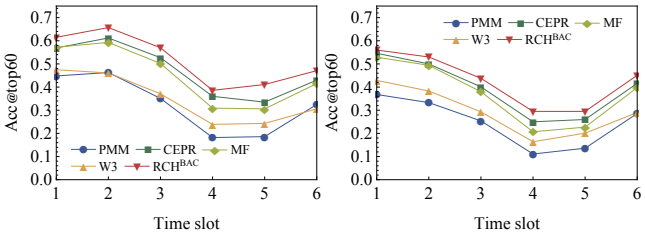
where $s \in \{0, 1\}$ denotes workdays or holidays, t is time of a day, r is the latent geographical region and z is the latent topic of a tweet. Since our dataset does not include tweets, the W^4 is de-generalized to W^3, i.e., who, where, and when.

• CEPR. This model divides human mobility into two types, regular and novel ones. It first classifies the next check-in as novel or



(a) different models (b) different variations of RCH

Figure 5: Acc@top$P$



(a) workdays (b) holidays

Figure 6: Acc@top$P$ for different type of days and time slots (12am-4am, 4am-8am, 8am-12pm, 12pm-4pm, 4pm-8pm, 8pm-12am)

regular by taking 5 different features into a supervised classification model. If the next venue is classified as regular, it employs the hidden markov model to predict the location. Otherwise, a social-based collaborative filtering algorithm is applied to explore suitable candidates.

4.1.4 Parameter Setting

Matrix $\mathbf{U}$, $\mathbf{V}$, $\mathbf{U}(t)$ and $\mathbf{H}(t)$ were initialized with the random assignments in $[0, 10]$. The default values of θ^A and θ^C are both 1. We set the four parameters in our model, γ, β, α and σ, as Ψ. The default value of Ψ is $(0.005, 0.005, 0.95, 10^{-5})$. They were tuned one by one and the results are reported in Sec. 4.2. We categorize days into two classes, workdays and holidays, and split each day into 6 time slots equally, i.e., 12am–4am, 4am-8am, etc. Therefore we have 12 time slots in total. We performed 6-fold cross validation to evaluate all methods.

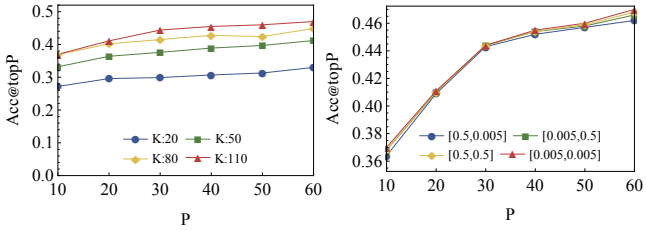
We use C/C$_{static}$ to represent our RCH model considering only the conformity term with/without time-varying user latent factors respectively. R is denoted as our RCH model that only contains the regularity term. RCH^{BAC}/RCH^C are the RCH model with/without heterogeneous datasets for learning location profiles 3.2, where RCH^C use only check-ins and RCH^{BAC} use all three mobility datasets.

4.2 Results

We first investigate the performance of all compared methods without considering heterogeneous mobility datasets. The average Acc@top$P$ with respect to P are illustrated in Fig. 5(a). The performances of individual models such as PMM and W^3 are not as good as the MF model, probably due to the sparsity of check-in data, and high visiting frequencies of partial venues. Cho et al. [6] used a dataset with users having at least 10 check-ins *every day*. In our experiment, we only filtered users who have less than 15 total check-ins, which further aggravates data sparsity and weakens the advantage of individual models. RCH^C and CEPR perform much better than the other 3 models, outperforming MF by 13.3% and 2.6% respectively when $P = 10$. The use of collaborative filtering improves the performance of CEPR and our models. How-

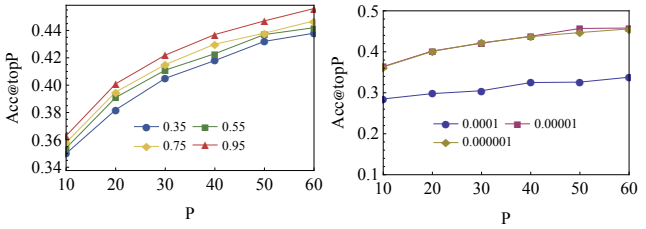
Models	Workdays						Holidays					
t	12-4am	4-8am	8am-12pm	12-4pm	4-8pm	8pm-12am	12-4am	4-8am	8am-12pm	12-4pm	4-8pm	8pm-12am
C_{static}	0.884	0.899	0.865	0.799	0.801	0.832	0.872	0.848	0.807	0.753	0.761	0.840
C	0.885	0.908	0.869	0.820	0.825	0.848	0.868	0.854	0.818	0.781	0.787	0.844
R	0.887	0.884	0.873	0.826	0.831	0.860	0.859	0.843	0.814	0.768	0.790	0.837
RCH^C	0.896	0.911	0.880	**0.835**	0.838	0.870	0.881	0.859	0.823	0.781	0.793	0.849
RCH^{BAC}	**0.899**	**0.912**	**0.883**	**0.835**	**0.840**	**0.871**	**0.890**	**0.863**	**0.829**	**0.786**	**0.795**	**0.850**

Table 3: APR of our models in different time slots.



(a) different K (b) different (γ, β)

Figure 7: Acc@top$P$ for different parameters in C



(a) different α (b) different σ

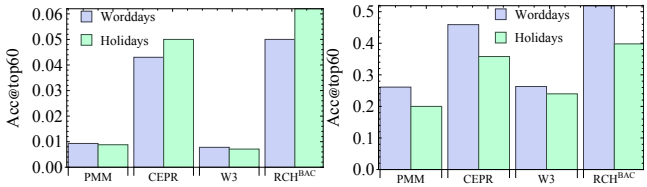
Figure 8: Acc@top$P$ for different parameters in R

ever, RCH^C shows an extra 10.7% improvement over CEPR when $P = 10$ due to the incorporation of both regularity and conformity.

Fig. 5(b) shows average Acc@top$P$ of the 5 variational RCH models. The Acc@top10 of the C$_{static}$ model is 0.06 higher than the R model, demonstrating that more check-ins are driven by conformity than regularity in our dataset. Besides, the introduction of time-varying latent factors in C improves the result of C$_{static}$ by 4.5% when $P = 30$. The combination of regularity and conformity of human mobility, RCH^C, exceeds C and C$_{static}$ by 6.5% and 8.9% respectively when $P = 50$. Compared with the C model, the improvement of RCH^{BAC} is 33% higher than that of RCH^C when $P = 60$, which indicates the predictive power of location profiles learned from heterogeneous mobility datasets. Our experiments demonstrate: 1) the effectiveness of combining regularity and conformity, 2) the superiority of time-aware collaborative model, and 3) the advantage of using heterogeneous mobility datasets.

Table 3 presents the APR and Figure 6 shows the Acc@topP of compared models in terms of different time slots and type of day. Similar to the results of accuracy, time slot 2 of workdays has the highest APR in most cases. Although the C model has greater accuracy than R shown in Fig. 5(b), its APR is less than that of R for 6 time slots. One possible reason is that users may have certain occasional movements that are not influenced by others and not highly regular for themselves (i.e., they visit such venues less regularly than their major/minor hubs). The APR of RCH^C consistently outperforms C and R, which suggests that regularity and conformity are complementary drivers of human mobility.

We studied the effects of different parameters in $\Psi = \{\gamma, \beta, \alpha, \sigma\}$. Fig. 7(a) reports the accuracy of the C model for varying K from 20



(a) unvisited venues (b) visited venues

Figure 9: Acc@top60 of visited and unvisited venues

to 110 when $\gamma = 0.5$ and $\beta = 0.5$. It is clear that a larger K results in higher accuracy for different P. This demonstrates that the more dimensions into which we segment the users, the more accurate predictions we have. However, when K increases, the growth on accuracy becomes slow. Fig. 7(b) shows that the change of γ and β has little impact on model C. Fig. 8(a) plots the accuracy of the R model for different α when $\sigma = 0.00001$, and Fig. 8(b) plots the accuracy of the R model for different σ when $\alpha = 0.95$. α is the proportion of l_1 norm in the regularization of $\mathbf{H}(t)$. Fig. 8(a) shows that the larger ratio of within-group sparsity tends to slightly increase prediction performance. This confirms our assumption that a user is active only in a few grids of a group. We observe from Fig. 8(b) that when σ is less than 10^{-5}, its effect on the performance remains stable. The sudden decrease happens when σ is larger than 10^{-5}.

Fig. 9(a) and 9(b) report the average Acc@top60 of visited venues and unvisited venues respectively. Because the MF model is not able to predict a use's unvisited venuess, we compared the other 4 models. RCH^{BAC} and CEPR benefit from the collaborative filtering approaches and outperform PMM and W^3 apparently for the prediction of unvisited venues. RCH^{BAC} outperforms CEPR by 24% when $P = 60$ in holidays. The Acc@top$P$ is 7.38 times as PMM and 8.7 times as W^3. For visited venues, the advantages of RCH^{BAC} and CEPR compared to individual models are not as obvious as unvisited venues, while the improvement of RCH^{BAC} is still higher than that of CEPR. An interesting observation is that the prediction accuracy of *unvisited* venues on holidays is higher than workdays for both CEPR and RCH, in contrast to the previous observation of higher *overall* predictability on workdays. This might be because people's movements to new places are more alike to each other during holidays than workdays, thus conformity shows a stronger predictive power for unvisited places.

5. CONCLUSION

In this paper, we have proposed a hybrid location prediction model called RCH integrating both the regularity and conformity of human mobility, capturing users' regular movement patterns and their occasional visits influenced by others. In particular, we incorporate heterogeneous mobility data to learn the spatial influence to venues and the group structure of different regions based on a gravity model. We have conducted extensive experiments using several large-scale mobility datasets, such as location check-ins, taxi tra-

jectories, and public transit data, to validate the effectiveness of our model. Based on the evaluation results, our model significantly outperforms existing approaches.

There are several interesting directions that could be explored in the future. For example, given semantic information of users' mobility such as geo-tweets and location tags, the location profiles can be further enriched, which should be beneficial for predicting the semantic mobility of users.

References

[1] D. Ashbrook and T. Starner. Using GPS to learn significant locations and predict movement across multiple users. *Personal and Ubiquitous Computing*, 7(5):275–286, 2003.

[2] J. P. Attanucci and N. H. Wilson. Bus passenger origin-destination estimation and related analyses using automated data collection systems. *Journal of Public Transportation*, 14(4):131, 2011.

[3] D. Balcan, V. Colizza, B. Gonçalves, H. Hu, J. J. Ramasco, and A. Vespignani. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences*, 106(51):21484–21489, 2009.

[4] J. Chang and E. Sun. Location 3: How users share and respond to location-based data on social networking sites. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, pages 74–80, 2011.

[5] C. Cheng, H. Yang, I. King, and M. R. Lyu. Fused matrix factorization with geographical and social influence in location-based social networks. In *AAAI*, volume 12, page 1, 2012.

[6] E. Cho, S. A. Myers, and J. Leskovec. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1082–1090. ACM, 2011.

[7] R. B. Cialdini and N. J. Goldstein. Social influence: Compliance and conformity. *Annual review of psychology*, 55:591–621, 2004.

[8] Y.-A. de Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel. Unique in the crowd: The privacy bounds of human mobility. *Scientific reports*, 3, 2013.

[9] N. Eagle and A. S. Pentland. Eigenbehaviors: Identifying structure in routine. *Behavioral Ecology and Sociobiology*, 63(7):1057–1066, 2009.

[10] H. Gao, J. Tang, and H. Liu. Mobile location prediction in spatio-temporal context. In *Nokia mobile data challenge workshop*. Citeseer, 2012.

[11] H. Gao, J. Tang, and H. Liu. gSCorr: modeling geo-social correlations for new check-ins on location-based social networks. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 1582–1586. ACM, 2012.

[12] H. Gao, J. Tang, and H. Liu. Exploring social-historical ties on location-based social networks. In *ICWSM*, 2012.

[13] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.

[14] H. H. Kelejian and I. R. Prucha. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, 157(1):53–67, 2010.

[15] M. Kim and D. Kotz. Periodic properties of user mobility and access-point popularity. *Personal and Ubiquitous Computing*, 11(6):465–479, 2007.

[16] Y.-J. Kim and S.-B. Cho. A HMM-based location prediction framework with location recognizer combining k-nearest neighbor and multiple decision trees. In *Hybrid Artificial Intelligent Systems*, pages 618–628. Springer, 2013.

[17] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.

[18] Z. Li, B. Ding, J. Han, R. Kays, and P. Nye. Mining periodic behaviors for moving objects. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1099–1108. ACM, 2010.

[19] D. Lian, X. Xie, V. W. Zheng, N. J. Yuan, F. Zhang, and E. Chen. CEPR: A collaborative exploration and periodically returning model for location prediction. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 6(1), 2014.

[20] D. Lian, C. Zhao, X. Xie, G. Sun, E. Chen, and Y. Rui. Geomf: joint geographical modeling and matrix factorization for point-of-interest recommendation. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 831–840. ACM, 2014.

[21] B. Liu, Y. Fu, Z. Yao, and H. Xiong. Learning geographical preferences for point-of-interest recommendation. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1043–1051. ACM, 2013.

[22] Y. Liu, W. Wei, A. Sun, and C. Miao. Exploiting geographical neighborhood characteristics for location recommendation. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 739–748. ACM, 2014.

[23] A. Monreale, F. Pinelli, R. Trasarti, and F. Giannotti. Wherenext: a location predictor on trajectory pattern mining. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 637–646. ACM, 2009.

[24] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil. An empirical study of geographic user activity patterns in foursquare. *ICWSM*, 11:70–573, 2011.

[25] A. Noulas, S. Scellato, N. Lathia, and C. Mascolo. Mining user mobility features for next place prediction in location-based services. In *ICDM*, pages 1038–1043. Citeseer, 2012.

[26] A. Sadilek, H. Kautz, and J. P. Bigham. Finding your friends and following them to where you are. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 723–732. ACM, 2012.

[27] S. Scellato, M. Musolesi, C. Mascolo, V. Latora, and A. T. Campbell. Nextplace: a spatio-temporal prediction framework for pervasive systems. In *Pervasive Computing*, pages 152–169. Springer, 2011.

[28] S. Scellato, A. Noulas, R. Lambiotte, and C. Mascolo. Socio-spatial properties of online location-based social networks. *ICWSM*, 11:329–336, 2011.

[29] M. N. Schmidt, O. Winther, and L. K. Hansen. Bayesian non-negative matrix factorization. In *Independent Component Analysis and Signal Separation*, pages 540–547. Springer, 2009.

[30] J. Scott, A. Bernheim Brush, J. Krumm, B. Meyers, M. Hazas, S. Hodges, and N. Villar. Preheat: controlling home heating using occupancy prediction. In *Proceedings of the 13th international conference on Ubiquitous computing*, pages 281–290. ACM, 2011.

[31] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani. A sparse-group lasso. *Journal of Computational and Graphical Statistics*, 22(2):231–245, 2013.

[32] C. Song, T. Koren, P. Wang, and A.-L. Barabási. Modelling the scaling properties of human mobility. *Nature Physics*, 6(10):818–823, 2010.

[33] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.

[34] L. Song, D. Kotz, R. Jain, and X. He. Evaluating location predictors with extensive wi-fi mobility data. In *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, volume 2, pages 1414–1424. IEEE, 2004.

[35] D. Wang, D. Pedreschi, C. Song, F. Giannotti, and A.-L. Barabasi. Human mobility, social ties, and link prediction. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1100–1108. ACM, 2011.

[36] A. G. Wilson. A statistical theory of spatial distribution models. *Transportation research*, 1(3):253–269, 1967.

[37] J. Ye, Z. Zhu, and H. Cheng. WhatâĂŹs your next move: User activity prediction in location-based social networks. In *Proc. of SIAM International Conference on Data Mining (SDM)*, 2013.

[38] N. J. Yuan, Y. Zheng, X. Xie, Y. Wang, K. Zheng, and H. Xiong. Discovering urban functional zones using latent activity trajectories. *IEEE Transactions on Knowledge & Data Engineering*, 27(3):712–725.

[39] N. J. Yuan, Y. Wang, F. Zhang, X. Xie, and G. Sun. Reconstructing individual mobility from smart card transactions: A space alignment approach. In *Data Mining (ICDM), 2013 IEEE 13th International Conference on*, pages 877–886. IEEE, 2013.

[40] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. M. Thalmann. Who, where, when and what: discover spatio-temporal topics for twitter users. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 605–613. ACM, 2013.

[41] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma. Understanding mobility based on gps data. In *Proceedings of the 10th international conference on Ubiquitous computing*, pages 312–321. ACM, 2008.