A Privacy-Preserving Semantic Annotation Framework Using Online Social Media

Shuo Wang[1](✉), Richard Sinnott[2], and Surya Nepal[3]

[1] Monash University, Melbourne, Australia
shuo.wang@monash.edu
[2] University of Melbourne, Melbourne, Australia
rsinnott@unimelb.edu.au
[3] CSIRO, Sydney, Australia
Surya.Nepal@csiro.au

Abstract. Semantic annotation framework that allows enriching locations or trajectories with semantic abstractions of the raw spatiotemporal data benefits understanding the semantic behavior of moving objects. Existing semantic annotation approaches mainly analyze specific parts of a trajectory, e.g. stops, in association with data from 3rd party geographic sources, e.g. (POI) points-of-interest, road networks. However, these semantic resources are static thus miss important dynamic event information. Recent location-based social networking provides a new dynamic and prevalent source of human activity data that can be a potential semantic resource for annotation. However, using the large-scale spatiotemporal data from online social media gives rise to privacy concerns. This paper thus presents a privacy-preserving semantic annotation framework P-SAFE that (i) identifies dynamic region of interest (DRI) from large-scale data provided by location based social networks whilst labelling of DRI into appropriate categories derived from spatial and temporal features of geotags, (ii) aligns trajectories to a set of DRI and enriches trajectories with semantics annotation derived from aligned DRI via THMM model, and (iii) embeds robust privacy-preserving mechanisms under differential privacy in each stage that accesses to raw data. P-SAFE approach tackles the privacy and utility trade-offs for meaningful geographic regions identification and labeling as well as trajectory semantic annotation under differential privacy whilst combining them into a single task. We demonstrate the effectiveness of P-SAFE approach on a dataset of large-scale geotagged tweets and a benchmark trajectory dataset for DRI construction and trajectory semantic annotation evaluation. The experimental results illustrate that P-SAFE not only provides robust privacy guarantees but remains approximate 45–56% accuracy for meaningful geographic regions labelling and 62–76% accuracy for trajectory semantic annotation.

H. Jin et al. (Eds.): ICWS 2018, LNCS 10966, pp. 353–372, 2018.
https://doi.org/10.1007/978-3-319-94289-6_23

1 Introduction

Generally, existing works on trajectory data mainly focus on raw trajectories that are typically demonstrated as streams of spatiotemporal (x, y, t) points, for data management and mining applications, rarely considering the background semantic information that can provide a further understanding of human movements. Analysis on raw trajectories hardly answers questions like what is the purpose for this person to visit a certain location at a particular time. Some works have implemented a semantic interpretation, e.g., semantic annotation framework that allows enriching trajectories with semantic abstractions of the raw mobility data, known as semantic trajectory, to benefit understanding the semantic behavior of moving objects. An example of semantic trajectory annotation is shown in Fig. 1. Semantic enrichment of trajectory data materializes as annotations that attach additional knowledge to the spatiotemporal positions in the trajectory. Such semantics is commonly obtained from geometric properties, e.g., stops or moves, in the geography on which the trajectory passed, e.g., landmarks [1]. As detecting homologous daily activity categories for regular people using their large-scale spatiotemporal data is a tractable due to the spatial and temporal recurrence of these activities. Thus semantic information can also derive from large-scale spatiotemporal data that enables insights into patterns of people's mobility and activities, e.g., eating, leisure or at work, as illustrated in Fig. 1.

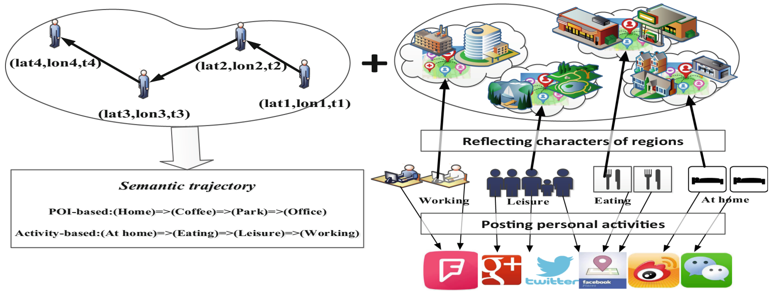
(lat4,lon4,t4)
(lat2,lon2,t2)
(lat3,lon3,t3)
(lat1,lon1,t1)
Semantic trajectory
POI-based:(Home)=>(Coffee)=>(Park)=>(Office)
Activity-based:(At home)=>(Eating)=>(Leisure)=>(Working)
Reflecting characters of regions
Working
Leisure
Eating
At home
Posting personal activities
twitter
facebook

Fig. 1. The semantic trajectory annotation.

However, there are several challenges when implementing a semantic annotation framework for generating semantic trajectories.

(1) Resource constraints. The commonly-adopted semantic annotation framework is based on 3^{rd}-party geographic information sources, e.g., POI. Densely populated urban areas may have many candidates POIs for annotation but it is not clear which POIs were visited. In contrast, there is a high number of venues not recorded by POI services, especially suburbs, rural areas or in the field. Therefore, it is intractable to infer the POI instance

for scenarios with restricted POI resources, e.g., annotation of residential places in rural area. On the other hand, semantics resources for location and trajectory semantic annotation, e.g., 3^{rd}-party geographic information sources, e.g., land-use categories [2], point of interest (POI)/landmarks [3] or landscapes [4], are generally assumed to be static regardless of the time. However, the category of functional regions will be dynamically changed due to different social activities performed there at the different time [5]. For example, City Conference Center, a common multi-purpose venue, could hold food festival (eating), business meeting (working) and art exhibition (leisure) as well as many other events. Therefore, such static semantic annotation will cause inappropriate annotating results.

(2) Periodic activity-based semantic annotation. Existing semantic annotation frameworks mainly ignore further insights on motivation or preferences of daily periodic activities reflected at the point of a trajectory. The study of underlying daily periodic activities performed at the position of trajectory that motivates the movements to conduct semantic enrichment is still on a less-explored stage. A periodic activity-based semantic annotation might benefit to understanding the impact of inhabiting dynamics in the city, better characterizing human mobility, and hence better urban planning and management [6]. Large-scale geo-referenced data from online social media offers a unique opportunity to gain insights into people's movements and activities that can be used in location and trajectory semantic annotation. Consequently, it is important to infer and implement the connection between spatiotemporal data and activity knowledge.

(3) Privacy concerns. The analysis of harvested (raw or processed) spatiotemporal data may compromise individuals' privacy (e.g., home location, political views or categories of disease based on their visited locations). The risk of revealing individual privacy are exacerbated when the adversaries possess background knowledge. It is essential that personal and sensitive information is not leaked when using individuals' spatiotemporal data, while the utility of the perturbed data should be maintained as far as possible.

Consequently, we propose P-SAFE, a semantic annotation framework that enables identifying dynamic and meaningful geographic regions and using them to turn raw trajectory data into semantic trajectories without breaching individuals' privacy. The core contributions of the article are summarized as follows.

(1) To solve the resource constraints of semantic resources and meet the dynamic requirement of external contextual information, we use the large-scale geotagged social media data as the sources for identifying the dynamic and meaningful geographic regions that reveal the spatial distribution of social media users at a different time is proposed. Specifically, a time-aware clustering approach is proposed to identify dynamic and meaningful geographic regions from a large-scale collection of geotagged social media data to explore the spatial distribution of geotags in a domain from the point view of individual social activity and mobility. The discovered clusters represent the dynamic and meaningful geographic regions.

(2) Daily activity categories are introduced as a new dimension for semantic enrichment of trajectories. To identify the dynamic region of interest (DRI) from meaningful geographic regions whilst labelling them using the daily-activity-based features, a location-entropy-based feature extraction model is proposed that allows accurately identifying and labelling of DRI into appropriate category derived from semantic analysis of inside geotags. Specifically, we identify DRI based on the distribution of users in the cluster reflecting in location entropy, as well as the temporal feature of the cluster. Then we label DRIs with semantics that reflects individuals' interests or purposes at a particular time, used to annotate human movement.

(3) Using the labelled DRI dataset, a semantic annotation model is proposed to infer the activity-based semantic trajectory behind raw trajectories, only using spatiotemporal information. There are two components: (i) a spatial alignment approach to discovery nearby DRI set for the position of a trajectory, and (ii) a Time-aware Hidden Markov Model (THMM) that intends to improve the inference accuracy, which is used to infer the daily activity type as the semantic annotation of the position of a trajectory from the nearby DRI set.

(4) Privacy. To address the privacy concerns, we embed robust privacy-preserving mechanisms under differential privacy in each stage that accesses to raw data. Specifically, there are two raw spatiotemporal data for protecting, i.e., the geotagged social media data for DIR discovery and labelling, as well as the raw trajectory as the input of semantic annotation. We apply a privacy-preserving mechanism to the results of time-aware clustering before releasing for the further use. In the location-entropy-based feature extraction model, we embed a differentially private mechanism into the location entropy calculation that accesses to raw spatiotemporal data. In the semantic annotation model for trajectory, we apply a privacy preserving mechanism to the spatial alignment to protect the privacy of trajectory. We conduct experiments on a dataset of large-scale geotagged tweets from Twitter and benchmark trajectory datasets for evaluation.

The rest of this paper is structured as follows. Background is presented in Sect. 2. Section 3 describes the ideas and mechanisms of privacy-preserving semantic annotation framework. Section 4 presents the evaluation metrics and the experimental performance of the approach. A survey of related work is given in Sect. 5. Section 6 draws conclusions on the work as a whole.

2 Problem Formulation

2.1 Preliminary Concepts

Definition 1 *(Dynamic Region of Interest - DRI). A set of meaningful geographic regions from spatiotemporal data of social media, associated with a dynamic semantics that reflects individuals' interests or purposes at a particular time and used for representing trajectory data. Each DRI is composed of* *geographic coordinate and semantic information to describe the location, denoted by* $D = \{d_1, d_2, \cdots, d_{|D|}\}$*, where* d_i *=(geographic value, semantic annotation). The DRI can be a spatial polygon or the geometric center of this polygon.*

Definition 2 *(Location Entropy). Given a location* l*, let the* GT_l *be the set of geotags (visits) to location* l *by all social media users. Let* $c_l = | GT_l |$ *represent the amount of visits at location* l*. Let* U_l *be the set of distinct individuals who visit at location* l*, and* $GT_{l,u}$ *be the set of visits that user* u *has made at location* l*. The* $s_{l,u} = | GT_{l,u} |$ *is represented as the amount of visits that user* u *made to location* l*. Thus* $p_{l,u} = \frac{| c_{l,u} |}{| c_l |}$ *denotes the portion of total visits that belongs to user* u*. The location entropy of* l *based Shannon entropy [7] is denoted by:*

$$H(l) = H(p_{l,u_1}, p_{l,u_2}, \cdots, p_{l,u_{|U_l|}}) = -\sum_{u \in U_l} p_{l,u} log_2 p_{l,u} \tag{1}$$

A higher location entropy value means the visits on this location are more evenly distributed among users that visited this location, i.e. it is a more popular location.

There are five generic daily activity labels defined in this article, denoted by $DAC = \{$*E: eating, H: at home, N: nightlife, R: recreation (leisure), W: working*$\}$, as well as five time slot categories.

Definition 3 *(Semantic trajectories-ST). A semantic trajectory* $st \in ST$ *is a structured trajectory where the spatial positions of a raw trajectory are replaced by geo-annotations and further semantic annotations, denoted by* $st = \{stp_1, stp_2, \cdots, stp_{|st|}\}$*. Each semantic annotation of a position is defined by* stp_i *=(semantic position, temporal value, semantic annotation). Here, the semantic position is a meaningful geographic region that represents the spatial location of the position at a semantic abstract level, e.g., the index of a nearby DRI. The semantic annotation is semantic enrichment about the position of a raw trajectory, e.g., the daily activity type performed at the position.*

The semantic annotation model aims to obtain an annotation sequence that can be potentially associated with a raw trajectory, based on the geographic sources DRI derived from geotagged social media data.

Definition 4 *(ϵ-Differential privacy* [8]*). Let* $\epsilon > 0$ *be a constant, a randomized function* F *satisfies* ϵ*-differential privacy if for all datasets* $D1$ *and* $D2$*, differing at most by one element from each other, all outcomes of the database* S *(*$S \subset$ *Range (F)) there is:*

$$Pr[F(D_1) \in S] \leq e^{\epsilon} Pr[F(D_2) \in S] \tag{2}$$

The parameter $\epsilon > 0$ is called the privacy budget and it allows users to control the level of privacy. A smaller privacy budget suggests more limits posed on the influence of an individual item, leading to a stronger privacy protection.

The basic idea to protect the privacy is to generate perturbed profiles or results such that the privacy leakage is minimized while the utility of the perturbed values can be still maintained.

2.2 Solution Overview

In this article, we aim to identify dynamic meaningful geographic region and label them with semantic information using the spatiotemporal information derived from social media, e.g., geotagged tweets, then use them to annotate the raw trajectories of moving objects. There are two types of input data: the raw trajectories, collected from a variety of platforms such as mobile devices and web services, denoted by $RT_u = \{p_1, p_2, \cdots, p_{|RT|}\}$, and spatiotemporal information from geotagged social media, i.e., geo tags, at a different time, denoted by $GT = \{g_1, g_2, \cdots, g_{|GT|}\}$, used as a prior knowledge of the individual activity and its transition pattern.

There are two main data processing components used to identify and label DRI and use them to conduct semantic annotation, based on the check-in data and raw trajectory data respectively, achieved in 5 steps, as shown in Fig. 2.

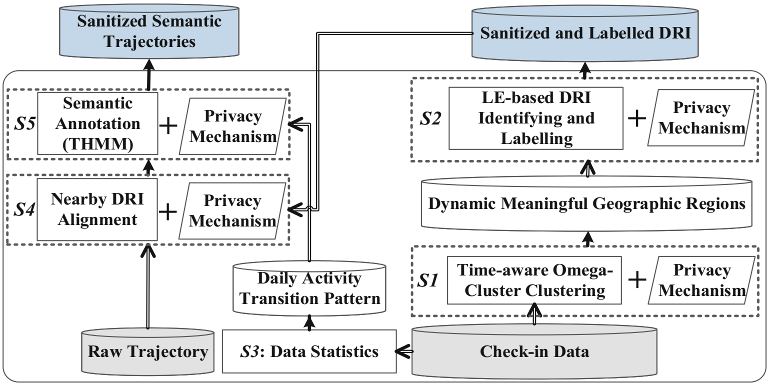
Sanitized Semantic Trajectories
Sanitized and Labelled DRI
S5
Semantic Annotation (THMM)
+
Privacy Mechanism
S2
LE-based DRI Identifying and Labelling
+
Privacy Mechanism
S4
Nearby DRI Alignment
+
Privacy Mechanism
Dynamic Meaningful Geographic Regions
Daily Activity Transition Pattern
S1
Time-aware Omega-Cluster Clustering
+
Privacy Mechanism
Raw Trajectory
S3: Data Statistics
Check-in Data

Fig. 2. The overview of P-SAFE framework.

The first component is *private DRI identifying and labelling* that extracts dynamic and meaningful geographic regions with semantic labels to annotate raw trajectory ($S1$–2). Specifically, a time-aware clustering approach is conducted on the large-scale geotags from Twitter to discover dynamic and meaningful clusters, followed by a privacy mechanism embedded into the clustering approach to protect the location privacy ($S1$). The output of $S1$ is an initially labelled sanitized clustering results. Then, the LE-based DRI identifying and labelling model is conducted, embedded with a privacy mechanism ($S2$). The output of $S2$ is the identified and labelled DRI dataset after sanitization for public use. In addition, a data statistics approach is used to generate the daily activity transition probability matrix ($S3$).

The other component is *private trajectory abstraction and annotation* to process raw trajectory data ($S4$–5). A private semantic annotation model is proposed to perform the semantic enrichments, achieved into: nearby DRI alignment

($S4$) that discover the nearby DRIs for the spatial positions of a raw trajectory, and semantic annotation ($S5$) that annotates the positions with DRI daily activity categories. The output of the second component is a sanitized semantic trajectory dataset for public use.

3 Private Semantic Annotation

3.1 Private DRI Identifying and Labelling

In this part, the input data is the geo-referenced social media data that enables users to share their activity-related choices, providing a new source of human activity data. The aim of the private DRI identifying and labelling is to identify and label dynamic and meaningful geographic regions associated characterization from geo-referenced posts in online social networks (specifically Twitter). Substantially, it is the task to conduct meaningful place recognition and annotation using the frequent temporal daily patterns and periodic occurrence of geotagged tweets for a geographic region. The dynamic human mobility patterns revealed by the check-ins over geo-referenced social media have been explored by existing works [9–12]. There are also some works that develop place classifier to conduct place annotation using machine learning based on the temporal and spatial features of geo-referenced social media data or using the 3^{rd} party resources, e.g., POI or Foursquare database. However, the prediction accuracy of machine learning approaches heavily depends on the number of instances in the training data and cannot well scale to the large-scale dataset. Further, many places cannot be mapped to 3^{rd} party resources, e.g., POI or land-use data, that are generally static. Therefore, in this part, we develop a fast and lightweight regions identifying and labelling framework to discover meaningful geographic regions (*time-aware clustering*) and conduct the DRI identifying and labeling based on the location entropy and the crowd behavioral patterns reflected in the temporal and spatial features of georeferenced tweets data (*LE-based DRI identifying and labelling*). Furthermore, we embed privacy-preserving mechanisms into these stages to handle the privacy concerns when accessing to raw spatiotemporal data.

Private Time-Aware Clustering. Existing works have demonstrated that there is a strong relationship between characteristics of a region and human daily mobility and activity patterns, reflected in geo-referenced social media data [9]. The collective dynamics of the greater metropolitan area are reflected in the geographic dynamics of Twitter usage [13]. Thus, the straightforward approach to identify the meaningful geographic regions relevant to dynamic activities is to perform density-based clustering algorithms on the various aggregated subsets of geotagged tweets in the term of the temporal features. For instance, a density-based clustering approach can be conducted on the geotagged tweets posted during the sleeping time period (e.g., 22:00 to 6:00 of weekdays) to identify potential residential regions. Therefore, this process can be divided into

Temporal-feature-based Aggregation and *Spatial-feature-based Clustering* as well as a *Clustering Results Sanitization* to sanitized clustering results for public use.

(1) Temporal-Feature-Based Aggregation. Based on the previous works and our observation of collected massive geotagged tweets (10 million for a year from users), we can see the periodic behavior of Twitter users reflected in the temporal patterns of geotagged tweets. The observation of the tweets frequency during the course of a week can confirm such periodic behaviors. We discover some temporal features that seem to mirror user daily behaviors, which is also confirmed in the previous works [6,13,14]. And users have very common temporal patterns of crowd daily activities, e.g., eating, working or leisure. For instance, during a week day, users always spend morning and afternoon on working while on leisure during weekends. As to exploit these patterns, we divided the day into five time slots with the difference between weekdays and weekends highlighted, according to the temporal patterns of five daily activity categories, formally specified by the following:

Definition 5 *(Time Slots Set-TSS)***.** *TSS is a set of time slots, where $ts \in TSS$ is a subset of varying time duration belonging to a day. $TSS = \{E_{ts}, H_{ts}, N_{ts}, R_{ts}, W_{ts}\}$, summarized in the following:*

E_{ts} *(Eating): [12:00, 15:00) of Weekdays;*
H_{ts} *(At Home): [17:00, 22:00), [22:00, 6:00) of Weekdays;*
N_{ts} *(Nightlife): [20:00, 6:00) of Weekends;*
R_{ts} *(Recreation): [6:00, 20:00) of Weekends;*
W_{ts} *(Working): [9:00, 12:00), [15:00, 17:00) of Weekdays.*

We aggregate the geotagged tweets into five subsets according to the TSS, and then conduct clustering algorithms on each subset to cluster coordinates into meaningful geographic regions (clusters).

(2) Spatial-Feature-Based Clustering. Common density-based clustering algorithms such as DBSCAN [15], aggregate many locations within the density definition, and each of these can be used to discover further density-reachable locations. Based on DBSCAN clustering algorithm, we can find dense regions on each subset from the temporal feature based aggregation and filter out any non-clustered locations as noise. Each cluster is considered to be a potential DRI candidate, represented by the geographic centroid of the cluster. Accordingly, we discover five clustering subsets (potential DRI subsets) using the five subsets from temporal feature based aggregation, denoted by $PD = \{PD_E, PD_H, PD_N, PD_R, PD_W\}$. Intuitively, we can associate the category of a potential DRI subsets to the accordingly activity category at that time slot. For instance, we can associate the category of PD_E to "Eating".

(3) Clustering Results Sanitization. We adopt the geo-indistinguishable mechanism [16] as the planar geographic centroid perturbation approach. The basic idea is to generate a new sanitized location with some privacy budget to protect the secret location. Given a centroid of a cluster that requires to be perturbed and the privacy budget allocated in this step, the sanitized result is generated

Algorithm 1. Clustering Results Perturbation Algorithm

```
Input: $M$, $\gamma_c$, $\epsilon_{ct}$, $\epsilon_{cc}$.
Output: Perturbed centroid.
$CT'_i = |Pmc_i| +$ Lap($\frac{\Delta f^i_{ct}}{\epsilon_{ct}}$);
if $CT'_i > \gamma_c$ then
    Figure out centroid $CentroidCC_i$ ;
    Transform $CC_i$ from (x,y) to polar coordinate $(\theta, r)$;
    Drawing $\theta$ uniformly from $[0, 2\pi]$;
    Drawing $p$ uniformly from $[0, 1]$;
    $r = -\frac{1}{\epsilon_{cc}}(W_{-1}(\frac{p-1}{e}) + 1)$;
    $CC'_i = CC_i + < rcos(\theta), rsing(\theta) >$;
```

by the Algorithm 1. The clustering results from a subset of PD can be denoted by $M = \{mc_1, \cdots, mc_{|M|}\}$, where Pmc_i is a set of points in the i^{th} cluster of M. The count and centroid of the cluster mc_i is denoted by CT_i and $CC_i = (x, y)$ respectively, and the perturbed values are denoted by CT'_i and CC'_i. The input parameters are threshold for counts γ_c, and privacy budgets $\epsilon_{ct}, \epsilon_{cc}$ used for count and centroid perturbation respectively. Here, count sensitivity for the cluster mc_i is given as $\Delta f^i_{ct} = MAX(NUM_{individual}(points)), \forall user\ u \in Pmc_i$. And the W_{-1} is the Lambert W function (−1 branch). We release the sanitized clustering results for the further use under the robust guarantees of differential privacy.

Private LE-Based DRI Identifying and Labelling. After discovering potential DRIs from the previous steps, the next step is to identify the actual DRIs based on the location entropy value. [17] shows that locations with high entropy are more likely to be shared (visited) than places with low entropy. Therefore, a higher location entropy value means the visits on this location are more evenly distributed among users that visited this location, i.e. it is a more popular location while the regions with small location entropy might be personal/residential regions. Thus, we can use location entropy to identify real DRI and label it using the category of activity, if the location entropy values meet the threshold of different activity scenarios. For instance, in the clustering results PD_{ts} derived from H_{ts} subset, we can infer the cluster with low location entropy as a DRI associated to "at home" activity category. In contrast, for the N_{ts} scenario, we can infer the cluster with higher location entropy as the DRI region where the type of user's behavior is typical in "nightlife", since a higher entropy implies that many users tweet from this region but they have few tweets. Note that, the threshold of different activity scenarios, denoted by $\theta = \{\theta_E, \theta_H, \theta_N, \theta_R, \theta_W\}$, might be different as well as the identifying rules. E.g., the rule in "at home" is $\leq \theta_H$ while it is $\geq \theta_i$ in others scenarios.

However, there are privacy concerns when directly using true values of location entropy. For instance, location entropy can reveal whether users visit a location or not by potential adversaries. Therefore, perturbation mechanisms should be implemented for true values of location entropy to ensure that any given adversary cannot determine whether or not a particular user visited a location.

More specifically, random noise is added to the real location entropy value to achieve differential privacy guarantees via Laplace mechanism. According to the definition of differential privacy, the first step is to decide the global sensitivity, denoted by ΔH, which controls the magnitude of the random noise. ΔH reveals the maximum change of location entropy of all locations when adding (or deleting) a single user from the dataset. Let C represent the maximum amount of visits a user contributes to a location, thus the function of C can be used as the sensitivity bound. Global sensitivity is decided as follows [18]:

$$\Delta H = max\{log2, logC - log(logC) - 1)\} \tag{3}$$

Then, the bounded change of adding (or deleting) a single user from the dataset on the entropy of all visited locations can be obtained by $\Delta f = \Delta H \times M_{max}$, where M_{max} is the maximum number of locations visited by a user. Thus, the random Laplace distribution is generated with mean 0 and scale $\sigma = \dfrac{\Delta f \times M_{max}}{\epsilon}$. Algorithm 2 illustrates the private LE-based DRI Identifying and Labelling.

Algorithm 2. Private LE-based DRI Identifying and Labelling

```
Input: A subset of potential DRIs $PD_x = \{sp_1, \cdots, sp_{|PD_x|}\}$, a set of users
       $U_{sp} = \{u_1, \cdots, u_{|U_{sp}|}\}$ visited cluster $sp$, privacy budget $\epsilon_{wp}$, thresholds set $\theta$.
Output: Sanitized DRI dataset $DD$.
for $sp \in PD_x$ do
    Calculate $\Delta H$ based on Equation(3) using C=$C_{max}$;
    for $i$ = 1 to $|PD_x|$ do
        for $j$ = 1 to $|U_{sp_i}|$ do
            Calculate $c_{sp_i,u_j}$ and $p_{sp_i,u_j}$;
        Calculate $H(sp_i) = -\sum_{u \in U_{sp_i}} p_{sp_i,u} log p_{sp_i,u}$;
        $\widehat{H}(sp_i)$=$H(sp_i)$+ $Lap(\frac{\Delta H \times M_{max}}{\epsilon_{wp}})$;
    if $\widehat{H}(sp_i)$ meets the rule associated with $\theta$ then
        $sp.label = PD_x.catagory$;
        $DD$.add(sp);
```

3.2 Private Semantic Annotation

Once the DRIs are identified and labelled, we use them to associate each position location of a raw trajectory with the most-likely activity type that represents the semantic enrichment. This semantic annotation can be achieved by two steps: select the nearby DRIs for each position and associate a most-likely activity category inferred from the activity labels of nearby DRIs. Furthermore, we embed privacy-preserving mechanisms into these steps to handle the privacy concerns when accessing to raw spatiotemporal data.

Private Nearby DRI Alignment. The basic idea of nearby DRI (from sanitized DRI dataset DD) alignment is that the locations in a raw trajectory are aligned to specific DRI set according to a geometric strategy. The strategy used in this method is to discover nearby DRIs, for every location in the trajectory within a specific domain. The maximum distance threshold η is used to restrict the maximum distance to search nearby DRIs for a position location of a raw trajectory, e.g., p_i, denoted by $NDRI_\eta^i$. Thus if there is no DRI within η for one location in a given trajectory, this location will be skipped.

Then we embed a privacy-preserving into the nearby DRI searching. Exponential mechanisms can be adopted to privately select nearby DRIs. Specifically, the chosen probability of a nearby DRI is calculated based on sensitivity and the score function. Using the chosen probability, the nearby DRI searching can be executed. The Euclidean distance between a position location of a raw trajectory, e.g., p_i, and a nearby DRI, e.g., dri_j, is used as the score function q that is defined as:

$$q(dri_j \in NDRI_\eta^i, pi) = (GS - dist(p_i, dri_j)), \tag{4}$$

$$Pr_{dri_j} = \frac{exp(\dfrac{\epsilon_{rs} \times q(dri_j, p_i)}{2 \times GS})}{\sum_{dri_j} exp(\dfrac{\epsilon_{rs} \times q(dri_j, p_i)}{2 \times GS})}, \tag{5}$$

The maximal change in the distance between p_i, dri_j can be used for the sensitivity GS. Note that the GS can be set as η in this case, since it is the maximum distance between p_i and nearby DRIs within η. Then the probability arranged for each $dri_j \in NDRI_\eta^i, p_i$ is calculated as Eq. (5). Here the ϵ_{rs} is the privacy budget used in the randomization. Based on the chosen probability, k most possible nearby DRIs can be chosen to generate a nearby DRI set, denoted by $\widehat{NDRI}_\eta^i$, $\widehat{NDRI}$ for short.

THMM-Based Semantic Annotation. This layer annotates each position of a trajectory with semantic information derived from its nearby DRI set $\widehat{NDRI}$. Specifically, we infer the most likely type of daily activities for each point of an individual trajectory from the types of its nearby DRI set. The Hidden Markov Model (HMM) [19] is improved for conducting this task, named Time-aware Hidden Markov Model (THMM). Generally, the probability of transition between two states at different time slots is different, e.g., the probability of transition from eating to working is different at daytime and night period respectively due to the essence of daily activity patterns. In THMM, a time dimension is used to the state transition along with the observation sequence advancing.

In this case, the observed states correspond to the sequence of positions in a raw trajectory, as the initial input. The DIR dataset is the superficial hidden states, which is identified as the nearby IDR set for each point of a raw trajectory. In this work, the hidden states, i.e., $DAC = \{C_1 : E, C_2 : H, C_3 : N, C_4 :, C_5 : W\}$, are the semantic enrichment for positions of a raw trajectory. Our goal is to identify the hidden states to annotate the points of a trajectory, e.g., the types of activities (eating, working or at home) performed by an individual at certain times, derived from the types of nearby DRIs.

The THMM, i.e., $\lambda = (\pi, A, B)$, is defined as follows. $\pi = \{\pi_i\}$ is the initial state matrix, where $\pi_i = Pr(C_i)$. $A = \{a_{i_p}^{j_q}\}$ is the state transition probability matrix, where $a_{i_p}^{j_q} = Pr(C_j^q|C_i^p)$. It represents the probability to engage in type of activity C_i at time slot p after the previous activity with type C_j at time slot q. Here, the p, q refer to the index of time slot in previous definition. $B = \{b_i^p\}$ is the observation probability matrix, where $b_i^p = Pr(o|C_i^p)$. It reveals the probability that an individual is observed at the location o at the time slot p to engage in type of activity C_i.

It is a sophisticated work to estimate these three components. As a possible approach, we approximate these values of THMM based on the activity patterns derived from the statistical analysis of geotagged social media information as a prior knowledge. The initial state matrix π can be estimated as the percentage of geotags belonging to each category of activity. The state transition probability matrix A can be approximated using the transition pattern in geotags among different categories of activity. $b_i^p = Pr(o|C_i^p)$ in observation probability matrix B can be approximated as $\sum_{dri_j \in \widehat{NDRI}} Pr(o|dri_j^{C_i})$ [1], where only the DRIs in the nearby DRI set are considered. In this paper, we assume a location in space o follows a normal distribution with $dri_j^{C_i}$ as the mean and a constant σ as the variance, i.e., $Pr(o|dri_j^{C_i}) = N(dri_j^{C_i}, \sigma^2)$.

Given the estimated values of three components in THMM, $\lambda = (\pi, A, B)$ can be used to annotate the hidden states (categories of activity behind the point of a trajectory) $HS = \{HS_1, HS_2, \cdots, HS_n\}$ from the observed points sequence $O = \{o_1, o_2, \cdots\}$ from stop/move identified or raw trajectory. Here, the HS_i is a category of activity from DAC. This problem can be treated as a dynamic programming problem, i.e., identifying the optimal state sequence associated with given observation sequence, known as the decoding of THMM. The input of THMM is a trajectory with each timestamp abstracted to the defined five time slots, and then let $\delta_d(i)$ denote the highest probability of the d^{th} point of a trajectory caused due to engaging in type of activity C_i at the time slot, defined in Eq. (6).

$$\delta_d(i) = \max_i Pr(HS_1, \cdots, HS_d = C_i, o_1, \cdots, o_d|\lambda) \tag{6}$$

$$\delta_{d+1}(j) = \max_i \{\delta_d(i) A_{ij}\} \times B_j(o_{t+1}) \tag{7}$$

$$\psi_{d+1}(j) = arg \max_i \delta_d(i) A_{ij} \tag{8}$$

Then, the corresponding induced the form of the highest probability of the $(d+1)^{th}$ point of a trajectory caused due to engaging in the type of activity C_j at time slot t is given in Eq. (7), via state transition probability. Further, the previous state C_i that give the highest probability to current state C_j is recorded as ψ_{d+1}, i.e., maximized Eq. (7), given in Eq. (8). The Viterbi algorithm [20] is used to conduct this dynamic programming problem to obtain the hidden state sequence for an input trajectory. Specifically, the first step is to recursively compute $\delta_d(i)$, following by deducing the state of the final point with the highest probability in the last point. Then return to the state of previous point via $HS^*_{d-1} = \psi_d(HS^*_d)$.

3.3 Private Analysis

The P-SAFE can be achieved into *private DRI identifying and labelling* and *private semantic annotation*. For *private DRI identifying and labelling*, there are two steps that access to the original geotags data: time-aware clustering and LE-based DIR identifying and labelling. To handle the privacy, privacy mechanisms are embedded into the first step to sanitize the clustering results (count and centroid of each cluster) for the further use and the second step for location entropy calculation of each cluster, which are the only two processes that access to the original data. For the perturbation of location entropy, [8,18] have proved that it is sufficient to achieve ϵ_{wp}-differential privacy using Laplace mechanisms with global sensitivity. The private perturbation of anchor point is divided into two step: planar centroid perturbation and counting perturbation with privacy budget ϵ_{cc} and ϵ_{ct} respectively. For counting perturbation, it satisfies ϵ_{ct}-differential privacy using a Laplacian mechanism scaled by ϵ_{ct} [8]. For centroid perturbation, given privacy budget ϵ_{cc} to perturb each centroid, this satisfied $\epsilon_{cc}d_{\chi}$ -privacy, which is a generalized variant of differential privacy under the metric d_{χ} [16]. For *private semantic annotation*, it can be divided into private nearby DRI alignment and THMM-based semantic annotation. We use an exponential mechanism to perturb the nearby DRI searching with given privacy budget ϵ_{rs} and global sensitivity GS. It is proved to satisfy ϵ_{rs}-differential privacy in [21]. As THMM-based semantic annotation is considered as post-processing on differentially private sanitized nearby DRI set without assessing user private data, there is no privacy loss in this phase. Based on the composition property of differential privacy, *private DRI identifying and labelling* and *private semantic annotation* satisfy differential privacy and then P-SAFE satisfies differential privacy.

4 Experiments

4.1 Experimental Data and Setting

In this article, experiments on real-world database and benchmark database to compare the efficiency and utility of our proposed approach were conducted. When permissions are given, each of their tweets can be attached with a corresponding geolocation. Experiments on one real-world dataset are conducted using geotagged Twitter data, specifically, we use a large-scale tweets data collected from Feb 25, 2015 to January 25, 2016, within Melbourne City. We select geotagged tweets to be GT. In addition, an individual's mobility traces can be constructed using the geotagged tweets of users who have more than 50 check-ins. Specifically, the traces are created by aggregating two or more geo-located tweets produced by the same user in locations more than $d = 100\,\mathrm{m}$ far away from one to another, within a time interval less than $t = 75\,\mathrm{min}$. This trace dataset can be used to infer the transition patterns among different categories of activity. Further, we use the land-use data and Foursquare venues database to be evaluation database, i.e., the prior knowledge used to evaluate the accuracy of labelled DRIs. For instance, if there are venues with the type of Pub, Nightclub, Bar, Entertainment, Theatre, Dance Studio, Opera House, etc., within the DRI associated to the type of "Nightlife", then we can say this DRI is properly labelled. The benchmark database, i.e., RT, is a labelled raw trajectory data that consists of 10 volunteers and 3568 geo-referenced records collected in Melbourne from August 2016 to September 2016 (30 days). In addition, volunteers defined their own activity type at the logged location. RT is used to evaluate the accuracy of outputs from P-SAFE.

We conducted the experiments on a personal computer equipped with 2.5 GHz CPU and 16 GB RAM. Each experiment was tested 10 times and the average result reported. The default threshold value parameters for different time slot scenarios were set experimentally, and default privacy budget set is $\epsilon = \{\epsilon_{cc} = 5, \epsilon_{ct} = 5, \epsilon_{wp} = 5, \epsilon_{rs} = 5\}$. Note that the experiments were performed on six classical differential privacy leakage levels, i.e., $\epsilon =$ Strong: 0.1 and 0.5; Normal: 1 and 5; Weak: 10 and 20. There are also some default parameters derived from experiments: $\gamma_c = 10, \theta = (0.5, 0.3, 0.5, 0.5, 0.5), k = 5, \eta = 100, Eps = 50, MinPts = 0.1$.

4.2 Experimental Evaluation

In order to estimate the accuracy of labelled DRIs and semantic annotation of raw trajectories, we use a set of metrics typically used in classification problems: precision, recall, and f-measure. On the other hand, we use Average Distance Error (ADE) as the utility metric for the sanitized DRIs and Mean Absolute Error (MAE) is used for the sanitized location-entropy values. These are denoted as follows.

$$ADE = \frac{\sum\limits_{idr_i \in RI} dist(idr_i - idr_i')}{|RI|}, MSE = \frac{\sum\limits_{idr_i \in RI} |H(i) - \widehat{H(i)}|}{|RI|} \tag{9}$$

Here, RI is the set of original DRIs that have sanitized values. $dist(dri_i - dri_i')$ is the Euclidean distance between them. $H(i)$ and $\widehat{H(i)}$ are the actual and sanitized weights respectively.

Evaluation on DRI. First, we evaluated the average accuracy of identified DRIs using the F-measure metric, as shown in Fig. 3. The F-measure was adopted as the accuracy metric, which was the harmonic mean of precision and recall. We used the land-use category data to evaluate the accuracy of "At home" labels associated to DRIs and Foursquare venues data for other scenarios. We compared the labelling accuracy of DRI labelling approach with and without the privacy-preserving mechanism (denoted by ON and OFF). As we can see, the highest average accuracy was obtained for characters "At home" and "Working". The DRI labelling approach labelled the "At home" relevant regions, i.e., residential regions, with an accuracy 91% referring to the land-use data, even achieving an accuracy 76% with perturbation mechanisms. "Working" relevant regions had average accuracy at 90% using OFF and 74% using ON. For the labelled "Recreation" and "Eating" regions, we observed that the average accuracy was approximately than 85% with OFF and 70% with ON, 87% and 84% using OFF while 71% and 65% using ON respectively. "Eating" relevant regions had average accuracy at 77% using OFF and 62% using ON. Most of the labelling results had accuracy more than 70% using the private DRI labelling approach, which was a satisfying result for the semantic annotation.

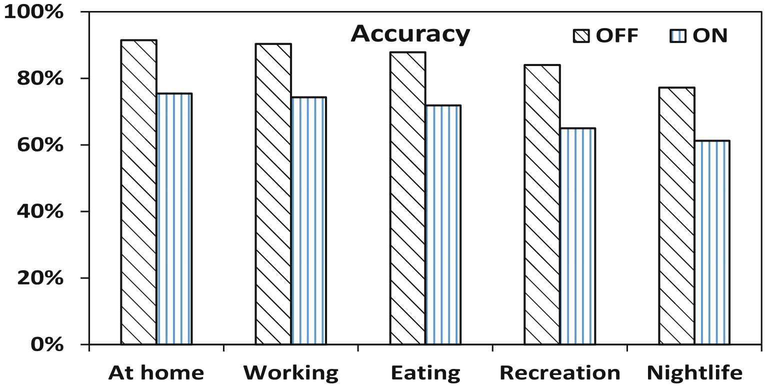

Fig. 3. The accuracy evaluation of DRI.

Furthermore, in order to show the utility of the perturbed DRI and perturbed location entropy over varying ϵ, a range of experiments were conducted as follows. Figure 4 showed the ADE metric with varying ϵ on the each subset of the GT dataset. As ϵ increased, i.e. reducing privacy preserving level, the ADE decreased while the recall increased in all scenarios, as less noise was needed. The ADE was similarly in each privacy budget set for the subset associated to a different time slot with various amount of geotags, which revealed the amount of noise is not related to the size of the dataset. As shown, the distortion of perturbed cluster centroids under differential privacy with normal privacy preserving level was approximately 400 m in subset associated to "At home" and 400 m in other scenarios, which was acceptable for general location-based services, e.g., POI recommendation. In addition, Fig. 4 also gave the MAE metric with varying ϵ on the each subset of the GT dataset. Note that, as the value of location entropy can be negative and/or huge, we normalized the original values of location entropy for the MAE evaluation. The same trend appeared again, as ϵ increases, the MAE

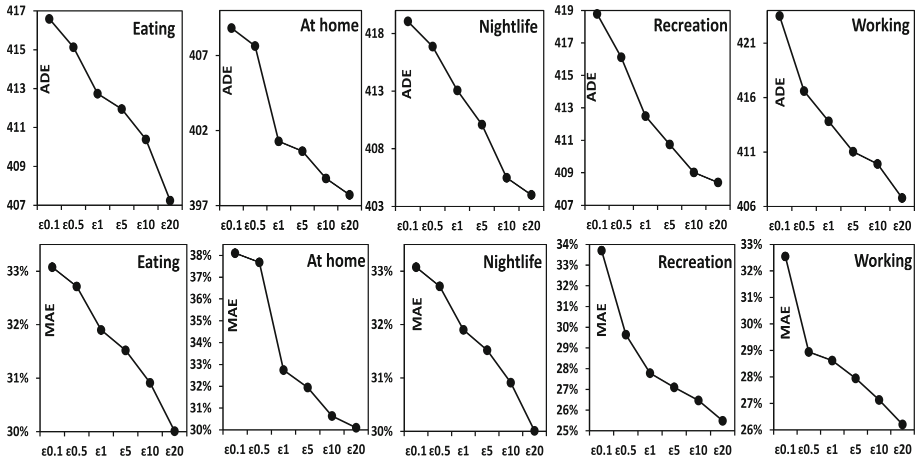

Fig. 4. The effect of varying ϵ on ADE evaluation of DRI.

and KL-divergence decrease while the recall increases, as less noise was needed. As illustrated, the distortion of perturbed location entropy under differential privacy with normal privacy preserving level was approximate 28% to 31% in all subsets compared to the true value of location entropy, which slightly impacted the DRI identification when appropriate thresholds were chosen.

Evaluation on Semantic Trajectory. The following experiments were conducted to evaluate the utility and accuracy of semantic trajectory generated by private semantic annotation using the labelled DRIs from previous steps. First, we evaluated the accuracy of semantic annotation using P-SAFE with and without the privacy-preserving mechanism (denoted by ON and OFF), compared with two competitors: P-SAFE without the THMM mechanism, denoted by $Near$, and POI-based annotation, denoted by POI. For competitor $Near$, we changed the THMM mechanism to a nearest DRI searching mechanism that adopted the label of the nearest DRI to be the semantic annotation for the position of a raw trajectory. For competitor POI, we used the 3^{rd} party POI data, e.g., Foursquare venues, to be the semantics source instead of DRIs. Note that, the daily activity transition patterns can be derived from geotags data after constructing human mobility traces. We used the Variable Order Mobility Markov Models [22] to infer the daily activity transition probability matrix.

The first sub-figure in Fig. 5 illustrated the values of *F-measure* of P-SAFE and its competitors. As shown in this figure, the OFF was the most performing approach with the highest accuracy in all five characters annotation. Generally, the accuracy obtained by OFF was 10% more than the $Near$ approach, which demonstrated that the accuracy of semantic annotation using P-SAFE without privacy-preserving mechanisms was improved by adopting the THMM. The $Near$ approach achieved the second highest accuracy of semantic annotation,

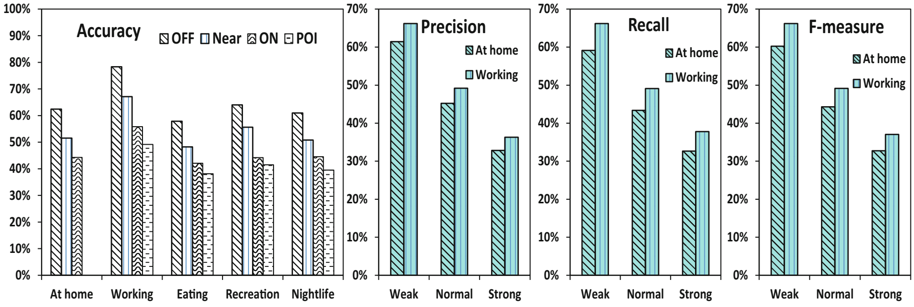

Fig. 5. The accuracy evaluation of semantic trajectory.

generally, 10–18% more than *POI*, which illustrated the accuracy of semantic annotation was improved by using DRIs than using POIs. In addition, the accuracy obtained by P-SAFE, even with privacy-preserving mechanisms, was better than the *POI* approach. Furthermore, this figure also demonstrated the ability to annotate residential activity of P-SAFE with and without the privacy-preserving mechanism. As evidenced by these comparisons, P-SAFE achieved an overall better performance in trajectory semantic annotation. The accuracy evaluation on evaluation dataset with varying ϵ are illustrated in Fig. 5. Due to the similar trends, we only compared the *Recall*, *Precision* and *F-measure* of "At home" and "Working" scenarios for example. As ϵ decreased, i.e. improving privacy preserving level, the *Recall*, *Precision* and *F-measure* decreased in both scenarios, as more noise was added. This confirmed the trade-off between privacy and utility again. Based on these experiments, it demonstrated that P-SAFE enabled to achieve two contradictory goals: provable robust privacy protection and efficient meaningful geographic regions labelling and trajectory semantic annotation.

5 Related Work

The semantic information used for trajectory semantic annotation can be derived from 3^{rd} party. E.g., comparing the positions with locations of predefined places of interest, land-use as well as from large-scale spatiotemporal data that enables insights into patterns of people's mobility and activities. Many existing works have illustrated the relation between geo-referenced social networks data and potential semantic enrichment resources. [13] developed insight into both geographic social dynamics and attention through social media analysis. [13] developed insight into both geographic social dynamics and attention through social media analysis. [12] used the spatiotemporal data from online social media check-in data to characterize urban human activity and mobility patterns. Specifically, they characterized individual activity patterns by finding the timing distribution of visiting different places depending on activity category. [23] proposed an approach that uses unsupervised learning and automatically determines land uses in urban areas by clustering geographical regions with similar tweeting activity patterns. [11] proposed an approach that learns how to classify personal places and trips while a human analyst visually analyzes and semantically annotates selected subsets of movement data using Twitter data. While we still have to be aware of the privacy concerns in exploring individual data. Many of these issues are discussed in [24]. Differential privacy presented in [8] has been broadly adopted to protect sensitive data in geospatial locations [16] and trajectory-based data [25].

To the best of our knowledge, there is no research on combining dynamic geographic regions identifying and labeling, trajectory semantic annotation and privacy preserving into single task.

6 Conclusions

In this paper, we proposed the P-SAFE approach to infer meaningful geographic regions that can reflect the periodic human activity using dynamic spatiotemporal data from online social media and then use them for trajectory semantic annotation, embedded with robust privacy guarantees. The approach improves the accuracy of DRI identifying and labelling as well as trajectory semantic annotation with robust privacy guarantees from three aspects. (1) A time-aware clustering approach is proposed to discover activity-related clusters whilst identifying dynamic region of interest (DRI) and labelling such regions via a LE-based DRI identifying and labelling approach; (2) A THMM model is used to infer most-likely activity category of each position of a raw trajectory from nearby DRIs that it passed through, as the semantic enrichment, improving the accuracy of semantic annotation; (3) Robust privacy preserving mechanisms are embedded into clustering, labelling and spatial queries perturbation under differential privacy. The P-SAFE approach tackles the privacy and utility trade-offs for meaningful geographic regions identifying and labelling as well as trajectory semantic annotation under differential privacy, combining them into a single task. Extensive experiments illustrate that it not only provides robust privacy guarantees but remains approximate 45–56% accuracy for meaningful geographic regions labelling and 62–76% accuracy for trajectory semantic annotation compared to competitors. On the future, we will extend P-SAFE into more comprehensive and practical cases from two aspects: (1) we will develop a utility-enhanced perturbation embedded into semantic labelling and annotating, e.g., location entropy calculation; (2) we will consider the content of tweets both in privacy and semantic aspects.

References

1. Yan, Z., Chakraborty, D., Parent, C., Spaccapietra, S., Aberer, K.: Semantic trajectories: mobility data computation and annotation. ACM Trans. Intell. Syst. Technol. (TIST) **4**(3), 49 (2013)
2. Yan, Z., Chakraborty, D., Parent, C., Spaccapietra, S., Aberer, K.: SeMiTri: a framework for semantic annotation of heterogeneous trajectories. In: Proceedings of the 14th International Conference on Extending Database Technology, pp. 259–270. ACM (2011)

3. Ashbrook, D., Starner, T.: Using GPS to learn significant locations and predict movement across multiple users. Pers. Ubiquitous Comput. **7**(5), 275–286 (2003)
4. Spaccapietra, S., Parent, C., Damiani, M.L., de Macedo, J.A., Porto, F., Vangenot, C.: A conceptual view on trajectories. Data Knowl. Eng. **65**(1), 126–146 (2008)
5. Rodrigue, J.P., Comtois, C., Slack, B.: The Geography of Transport Systems. Taylor & Francis, Abingdon (2016)
6. Phithakkitnukoon, S., Horanont, T., Di Lorenzo, G., Shibasaki, R., Ratti, C.: Activity-aware map: identifying human daily activity pattern using mobile phone data. In: Salah, A.A., Gevers, T., Sebe, N., Vinciarelli, A. (eds.) HBU 2010. LNCS, vol. 6219, pp. 14–25. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-14715-9_3
7. Shannon, C.E.: A mathematical theory of communication. ACM SIGMOBILE Mob. Comput. Commun. Rev. **5**(1), 3–55 (2001)
8. Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating noise to sensitivity in private data analysis. In: Halevi, S., Rabin, T. (eds.) TCC 2006. LNCS, vol. 3876, pp. 265–284. Springer, Heidelberg (2006). https://doi.org/10.1007/11681878_14
9. Lee, R., Wakamiya, S., Sumiya, K.: Urban area characterization based on crowd behavioral lifelogs over Twitter. Pers. Ubiquitous Comput. **17**(4), 605–620 (2013)
10. Cheng, Z., Caverlee, J., Lee, K., Sui, D.Z.: Exploring millions of footprints in location sharing services. In: ICWSM 2011, pp. 81–88 (2011)
11. Andrienko, G.L., Andrienko, N.V., Fuchs, G., Raimond, A.M.O., Symanzik, J., Ziemlicki, C.: Extracting semantics of individual places from movement data by analyzing temporal patterns of visits. In: COMP@ SIGSPATIAL, pp. 9–15 (2013)
12. Hasan, S., Zhan, X., Ukkusuri, S.V.: Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. In: Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing, p. 6. ACM (2013)
13. França, U., Sayama, H., McSwiggen, C., Daneshvar, R., Bar-Yam, Y.: Visualizing the "heartbeat" of a city with tweets. Complexity **21**(6), 280–287 (2016)
14. Li, L., Goodchild, M.F., Xu, B.: Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. Cartogr. Geogr. Inf. Sci. **40**(2), 61–77 (2013)
15. Ester, M., Kriegel, H.P., Sander, J., Xu, X., et al.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: KDD 1996, pp. 226–231 (1996)
16. Andrés, M.E., Bordenabe, N.E., Chatzikokolakis, K., Palamidessi, C.: Geo-indistinguishability: differential privacy for location-based systems. In: Proceedings of the 2013 ACM SIGSAC Conference on Computer and Communications Security, pp. 901–914. ACM (2013)
17. Toch, E., Cranshaw, J., Drielsma, P.H., Tsai, J.Y., Kelley, P.G., Springfield, J., Cranor, L., Hong, J., Sadeh, N.: Empirical models of privacy in location sharing. In: Proceedings of the 12th ACM International Conference on Ubiquitous Computing, pp. 129–138. ACM (2010)
18. To, H., Nguyen, K., Shahabi, C.: Differentially private publication of location entropy. In: Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, p. 35. ACM (2016)
19. Rabiner, L.R.: A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE **77**(2), 257–286 (1989)
20. Forney, G.D.: The Viterbi algorithm. Proc. IEEE **61**(3), 268–278 (1973)
21. McSherry, F., Talwar, K.: Mechanism design via differential privacy. In: 48th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2007, pp. 94–103. IEEE (2007)

22. Begleiter, R., El-Yaniv, R., Yona, G.: On prediction using variable order Markov models. J. Artif. Intell. Res. **22**, 385–421 (2004)
23. Frias-Martinez, V., Frias-Martinez, E.: Spectral clustering for sensing urban land use using Twitter activity. Eng. Appl. Artif. Intell. **35**, 237–245 (2014)
24. Xue, M., Kalnis, P., Pung, H.K.: Location diversity: enhanced privacy protection in location based services. In: Choudhury, T., Quigley, A., Strang, T., Suginuma, K. (eds.) LoCA 2009. LNCS, vol. 5561, pp. 70–87. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-01721-6_5
25. He, X., Cormode, G., Machanavajjhala, A., Procopiuc, C.M., Srivastava, D.: DPT: differentially private trajectory synthesis using hierarchical reference systems. Proc. VLDB Endow. **8**(11), 1154–1165 (2015)