

Article

Mining Individual Similarity by Assessing Interactions with Personally Significant Places from GPS Trajectories

Mengke Yang, Chengqi Cheng and Bo Chen *

College of Engineering, Peking University, Beijing 100871, China; yangmk@pku.edu.cn (M.Y.); ccq@pku.edu.cn (C.C.)

* Correspondence: pkuchenbo@pku.edu.cn; Tel.: +86-010-6275-5390

Received: 29 January 2018; Accepted: 17 March 2018; Published: 19 March 2018

Abstract: Human mobility is closely associated with places. Due to advancements in GPS devices and related sensor technologies, an unprecedented amount of tracking data has been generated in recent years, thus providing a new way to investigate the interactions between individuals and places, which are vital for depicting individuals' characteristics. In this paper, we propose a framework for mining individual similarity based on long-term trajectory data. In contrast to most existing studies, which have focused on the sequential properties of individuals' visits to public places, this paper emphasizes the essential role of the spatio-temporal interactions between individuals and their personally significant places. Specifically, rather than merely using public geographic databases, which include only public places and lack personal meanings, we attempt to interpret the semantics of places that are significant to individuals from the perspectives of personal behavior. Next, we propose a new individual similarity measurement that incorporates both the spatio-temporal and semantic properties of individuals' visits to significant places. By experimenting on real-world GPS datasets, we demonstrate that our approach is more capable of distinguishing individuals and characterizing individual features than the previous methods. Additionally, we show that our approach can be used to effectively measure individual similarity and to aggregate individuals into meaningful subgroups.

Keywords: individual similarity measurement; trajectory; personally significant place; place semantics; human-place interactions

1. Introduction

Advancements in GPS devices and related sensor technologies have resulted in the generation of an unprecedented amount of tracking data, which has enabled investigations of human mobility across a wide range of disciplines, such as urban planning, traffic management, tourism, location-based services, and public health. Movement tracks are usually recorded as trajectories, which are temporal sequences of spatio-temporal points, such as (x, y, t) . Among the large number of studies on trajectory data, researchers have shown particular interest in the trajectories of moving individuals because of the latent social and commercial benefits of these trajectories. In particular, mining the similarity between individuals based on trajectory data, which plays an important role in inter-trajectory studies [1], is a major focus due to its potential for use in characterizing individual movement features [2–4], inferring personal preferences [5,6], and predicting individuals' future positions [7].

In this paper, individual similarity refers to the commonality that two individuals share in their interactions with places. Most existing works concerning measuring individual similarity have been proposed in the context of location-based social networks services (LBSNs), in which users' common interests are the main focus. Thus, previous methods of similarity measurement have been developed based on the sequential properties of visits to public places. For example, Zheng et al. [5] proposed

a hierarchical-graph-based similarity measurement (HGSM) framework to estimate user similarity based on the assumption that users with longer similar visit sequences at finer geographic granularities might be more similar. Xiao et al. [8] modeled users' trajectories using their location histories and then explored the user similarity based on the sequential properties, granularities, and popularities of the visited semantic locations (locations that are semantically meaningful to users). Although previous studies have placed strong emphasis on the sequential properties of trajectories, they have not focused on the trajectories' spatial and temporal properties, which are essential for depicting individuals' characteristics in the wider applications of geographic information science (GIScience).

Meanwhile, there is growing interest among researchers regarding the semantic aspects of trajectory data. Semantic information has been incorporated into numerous recent studies on individual similarity; in most, semantic information has been extracted using reverse geocoding technology or land use data [9–13]. For example, in [13], Comito et al. mined interesting semantic locations and frequent travel sequences among those locations in a given region from geo-tagged posts; they then used this information to estimate the similarity between users that was based on their location histories. However, when inferring semantic information about locations, only public places retrievable through the Foursquare API were considered, because the interesting locations considered in [13] were limited to culturally important places and commonly frequented public areas. Although querying the semantics of public places from geographic databases is adequate for some applications, we suggest that, in general, it is insufficient to ignore personally significant places and consider only public places that may have no personal meaning. Another limitation of previous studies is that they generally did not distinguish between the different personal semantics of two individuals visiting the same place. For instance, when two people go to the same restaurant, Person 1 may go for dinner, whereas Person 2 may work there. Existing similarity measurement methods commonly assume that Persons 1 and 2 are similar, rather than considering the different semantics of visits to the same restaurant with disparate personal meanings.

In this paper, we propose a framework for mining individual similarity that considers interactions with personally significant places. The main contributions of this paper can be summarized as follows.

- (1) We present a framework for mining individual similarity. The novelty of the proposed similarity measurement, called the ISM-PSP, lies in exploring both the spatial distribution and temporal signatures of individuals' significant visits. Unlike existing similarity approaches, we go beyond simply emphasizing the sequential aspects of individuals' movements.
- (2) We propose that two individuals' significant visits can be compared only when they are identical in their semantic aspects. In contrast to many previous works, which have addressed commonly interesting places and their functionality for the general public, we determine the semantics of individually significant places from the perspectives of personal behavior. Specifically, when extracting the semantics of significant places, we consider places of both personal and public interest. We interpret the semantics of personal places of interest based on the temporal distribution of a person's presence, and determine the semantics of public places of interest using public geographic databases.
- (3) We conduct several experiments using the real-world Geolife dataset. The results show that the proposed ISM-PSP outperforms previous works in its ability to differentiate between individuals. It presents a high ratio of finding identical individuals, while maintaining a low number of false identifications and can be used to generate meaningful groups of individuals. In addition, a comparison of the ISM-PSP results with and without personal meaning illustrates the insufficiency of many previous works, in which only visits to public places have been considered when mining individual similarity.

The remainder of this paper is organized as follows. Section 2 reviews the related studies. In Section 3, we propose our framework and present the details of the methods that are applied in each

step of the framework. Experiments using a real-world individual trajectory dataset are presented in Section 4. Finally, in Section 5, we draw conclusions, discuss limitations, and suggest future work.

2. Related Work

The framework proposed in this paper is closely related to previous studies of place semantics extraction and user similarity measurements.

2.1. Significant Place Semantics Extraction

Since Spaccapietra et al. first introduced the concept of semantic trajectory in [14], many researchers have attempted to enrich raw trajectory data with relevant semantic information. In the cited work, the authors proposed a well-known semantic trajectory model, known as stops and moves. In this model, stops are places where moving objects stay for a certain amount of time, and moves are the movements between any two stops. Motivated by this model, many studies on semantic trajectories have used stop detection as part of their semantic enrichment processes (e.g., [14–16]). Moreover, greater emphasis is placed on stops than on moves, because stops can be further clustered into visited places. Frequently visited places are termed “significant places” for the individuals, and they are believed to generally bear rich semantics, which is crucial for a better understanding of moving objects [17–20]. Methods of extracting the semantics of significant places can be divided into two main types: location-based and time-based methods.

Pioneering location-based methods were proposed by Alvares et al. [19] and Bogorny et al. [20], who integrated background geographical information into trajectories and then extracted the semantics of a potential stop, when that stop intersected with a given geographical object for some minimum amount of time. Their preprocessing step required all geographic places relevant to the application of interest to be defined a priori. More general location-based methods have attempted to infer significant place semantics by associating points of interest (POIs) with stops that are based on spatial proximity. Xiao et al. [21] transformed individuals’ location histories from the geographic space into a semantic space using a POI database. Considering the fact that the POI nearest to a stop may not be the place that was actually visited, they constructed a feature vector for each stay region that reflects the uncertainty of possible POI categories that are assigned to that region. In [22], Spinsanti et al. proposed a more sophisticated approach in which a probability-ranked list of possible POIs was generated for each visited place. After classifying the POIs into different categories, they computed the most likely POIs for each significant place by incorporating additional domain information about the POIs, such as their opening times. Next, they summed all of the probabilities for POIs belonging to the same category and used the aggregate probabilities to assign possible POI categories to each significant place to obtain the semantics. In general, the location-based studies discussed above failed to identify personally significant place semantics because they extracted places and their semantics solely from public geographic databases.

In contrast to location-based methods, which extract place semantics by comparing the spatial positions of visited places to those of predefined POIs, time-based methods extract the semantics of significant visited places based on the temporal signatures of those visits. Using the behavioristic assumption that “what you are can be determined by when you are”, Ye et al. [23] identified the semantics of various places by observing when large numbers of users interacted with those places. However, they attempted to assign category tags to untagged common public places, and thus were unable to address the problem of identifying place semantics with personal meanings. Shen et al. [24] implemented the ST-DBSCAN to detect spatio-temporal regions of interest (ST-ROIs) and then differentiated between individuals’ behaviors in the same ST-ROIs by assessing the differences between the individuals’ visit times. Because this framework was proposed to identify activity groups, the authors divided the ST-ROIs into generic region types and grouped individuals by comparing the time that was allocated to different generic ST-ROI types instead of to personally significant places. To address the personal semantics gap, Andrienko et al. [25] used a procedure to find individual POIs

and created “temporal signatures” that were characterizing the temporal distribution of a person’s presence at each POI. Their experiments demonstrated that there are only a small proportion of significant places (e.g., homes and workplaces) whose semantics can be derived from temporal and statistical information. Thus, their results suggested that the personal meanings of individual POIs could not be inferred solely from their temporal signatures.

2.2. User Similarity Measurement

Many existing methods of user similarity measurement have been proposed to provide recommendation services. Li et al. [26] proposed a framework for modeling users’ location histories and mining user similarity. They combined all users’ trajectory data and hierarchically clustered those data into geographic regions, which were then used to build individual hierarchical graphs. When measuring the similarity between users, they incorporated both the sequence of visited regions and the geographic granularity at which similar sequences were found. Zheng et al. [5] extended the work of Li et al. [26] by using a new sequence matching strategy and considering the popularity of the visited locations. Specifically, the enhanced framework considered three aspects of users’ location histories: the sequence of movements, the hierarchical properties of the geographic space, and the popularity of the visited places. The users’ location histories were represented by hierarchical graphs (HGs), and similar sequences shared between two users in each layer of the hierarchy were further matched and used to calculate the similarity between the users. Although these approaches measure user similarity in geographic space, they do not consider geographic properties, such as the distances between locations. Moreover, they do not take the semantics of locations into account. The idea that semantic meaning should be considered when measuring user similarity was presented in the work of Ying et al. [9], who proposed the Maximal Semantic Trajectory Pattern Similarity (MSTP-Similarity) to measure the similarity between two maximal semantic trajectory patterns based on the longest common sequence (LCS) of these two patterns. Next, they extended the MSTP-Similarity to explore the similarity between two users, measuring user similarity based on a weighted average obtained by incorporating all possible MSTP-Similarities between the patterns from the two pattern sets. However, Chen et al. [27] found that this weighted average of pattern similarities that is proposed in [9] is unsuitable for measuring user similarity and cannot guarantee the maximum similarity between two identical users. Rather than considering all of the maximal patterns of the other user, as in the MSTP, they proposed an MTP-Similarity measurement that considers only the most similar pattern of each maximal sequence pattern. Both the MSTP and MTP calculate the similarity between maximal patterns in the same way—based on the lengths of the LCSs; they then use different strategies to integrate the similarity values between maximal trajectory patterns.

We suggest that the existing approaches for mining individual similarity have the following drawbacks: (1) While emphasizing the sequential properties of individuals’ movements, most existing methods of similarity measurement have not considered the spatial and temporal aspects of those movements; thus, they cannot be used to assess the distinctive characteristics of individuals for many GIScience applications. (2) Although previous studies have incorporated semantic information in individual similarity measurements, many of them have focused on public places, while neglecting the essential role of personally significant places in characterizing individuals. In addition, most works have ignored the distinct semantics of different individuals’ visits to the same places.

3. Proposed Framework

3.1. Overview

Individual human mobility is closely associated with places. Each individual moves from place to place to perform various activities, driven by either daily routines or interests [28,29]. Studies have shown that individuals have a remarkable propensity to return to places that they frequently visit [30,31]. Hence, how individuals spatially and temporally interact with their frequently visited places

(i.e., personally significant places) shows promise for revealing individuals' characteristics [22,23]. By investigating the spatio-temporal distributions of individuals' visits to their personally significant places, we can mine similar individuals and aggregate them into meaningful subgroups. For example, consider the distinctive characteristics of young versus elderly people: from a spatial perspective, elderly people's significant places tend to be distributed within a smaller area, whereas young people's significant places may be far apart, and long-distance commuting to them is common. From a temporal perspective, young people may visit their various significant places for long intervals of time during the day; they also frequently visit public places of interest at night. By contrast, elderly people's nighttime visits are generally limited to their homes.

In addition to the spatial and temporal properties of individual movements, we believe that semantic information must be considered. In contrast to some existing methods (for example, [32]), in which semantics is treated as an independent dimension, similar to space and time, we consider semantics as a precondition. Specifically, we consider the spatial and temporal properties of two individuals' visits to their significant places to be comparable only when they are identical from the semantic perspective. In this way, we can mine individual similarity by separately measuring individuals' movements that are related to different semantics, and we can apply this approach in various fields by synthesizing individual similarity in the relevant semantics. For example, we can identify families and colleagues who share high spatio-temporal similarities in their homes and workplaces, respectively. Additionally, we can identify friends or potential friends from their high degree of similarity in visits to entertainment venues, even if they do not share other regular visits.

Motivated by these goals, we propose a framework for mining individual similarity that consists of two major phases (see Figure 1). In phase 1, we first detect stay points using raw GPS trajectories; we then identify each individual's significant places by separately clustering their stay points using a density-based clustering algorithm. Based on the temporal signatures of their visits, personal places of interest (homes and workplaces, in this paper) are identified. The remaining significant places are assigned to public places of interest, such as shopping malls and hotels, using additional geographical contextual information (i.e., a POI dataset). In phase 2, we mine individual similarity using a new measurement, the Individual Similarity Measurement considering interactions with Personally Significant Places (ISM-PSP). The ISM-PSP computes the spatial and temporal similarity of two individual's visits when their personal semantics match. Based on the similarity scores for visits to personally significant places, which are grouped by diverse semantic information, the ISM-PSP measures the similarity between individuals by computing the weighted sum of their similarity scores, based on different semantics.

Below, we clarify some basic concepts and notations that are used in this paper prior to detailing the methods used in each step of our framework.

Definition 1. (GPS point and GPS trajectory) A GPS point is a triple of the form $p = (\text{latitude}, \text{longitude}, t)$ that represents a latitude-longitude location and a timestamp. A GPS trajectory is sequence of triples $T = \langle p_1, p_2, \dots, p_n \rangle$, where p_i is a GPS point and $p_1.t < p_2.t < \dots < p_n.t$.

Definition 2. (Stay point) A stay point represents a geographic region in which an individual stays longer than a given time threshold θ_{time} within a distance threshold θ_{distance} . A stay point is denoted by a quadruple of the form $s = (\text{latitude}, \text{longitude}, t_{\text{arrive}}, t_{\text{leave}})$, which represents the latitude-longitude location of s , and the individual's arrival time at s and departure time from s .

Definition 3. (Individual significant place) An individual significant place is a collection of stay points denoted by $SP_k^i = \{s_{k1}^i, s_{k2}^i, \dots, s_{kn}^i\}$, where s_{kj}^i is the j th stay point corresponding to the k th significant place SP_k^i of a specific individual i . The coordinates of SP_k^i are represented by the average latitude and longitude of the constituent points. An individual significant place SP_k^i represents a region that is frequently visited by i ; this fact implies that the place possesses some personal meaning for i . Given the diversity in possible personal

meanings, the set of individual i 's significant places $SP^i = \{SP_1^i, SP_2^i, \dots, SP_K^i\}$ is divided into personal places of interest, $PeSP^i$ and public places of interest, $PuSP^i$.

Definition 4. (Personal place of interest and public place of interest) A personal place of interest $PeSP_k^i$ is a place that is frequently visited by an individual i due to its special personal meaning for i . A typical example of a $PeSP_k^i$ is an individual i 's home. By contrast, a public place of interest $PuSP_k^i$ is a place that is of interest to the individual i and has a personal meaning for i that is identical to its functionality for the general public. Typical examples of $PuSP_k^i$ are places where the individual i goes during his or her leisure time, such as a restaurant or a shopping mall.

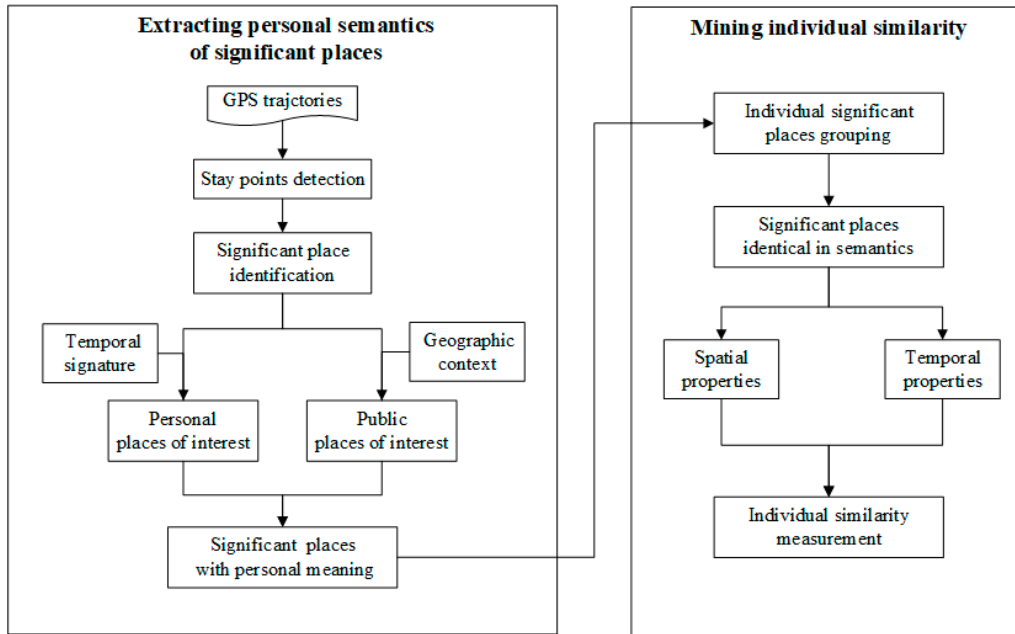


Figure 1. The proposed framework for mining individual similarity.

3.2. Extracting the Semantics of Personally Significant Places

In this section, we describe the extraction of the semantics of individual significant places when considering the problem from the perspective of personal behavior. The details of the process are given in Algorithm 1.

Algorithm 1. PersonalSemanticExtraction (TH, STI, POI).

Input: TH: The set of individuals' trajectories $TH = \{TH^i \mid 1 \leq i \leq |I|\}$
 STI: The set of standard time intervals $STI = \{STI_{ps}\}$
 POI: The set of points of interest

Output: SP_{ps} : The set of individuals' significant places with personal meaning
 $SP_{ps} = \{SP_{ps}^i \mid 1 \leq i \leq |I|\}$

Foreach $i \in I$ do
 $s^i = \emptyset$; // stay points of i
 Foreach $T^i \in TH^i$ do
 s^i .Add(StayPointDetection($T^i, \theta_{time}, \theta_{distance}$));
 $SP^i = OPTICS(s^i, r, MinPts)$; // obtain significant places from stay points
 $PeSP^i = MatchPe(SP^i, STI, \epsilon)$; // identify semantics of personal places of interest
 $SP^i = SP^i - PeSP^i$;
 $PuSP^i = MatchPu(SP^i, POI)$; // identify semantics of public places of interest
 $SP_{ps}^i = PeSP^i \cup PuSP^i$;
 SP_{ps} .Add(SP_{ps}^i);
 Return SP_{ps} ;

3.2.1. Identification of Individual Significant Places

As illustrated in Figure 2, a four-layered model is applied to identify the significant places with personal meaning for an individual. The lowest layer of the model consists of the raw historical GPS trajectory data of individual i , which are semantically poor. In the second layer, stay points are detected from every GPS trajectory in i 's historical trajectory data. In the third layer, these stay points are clustered in order to identify the individual i 's significant places. These clusters bear rich semantic information and are used to further extract personal places of interest and public places of interest. Finally, the extracted significant places with personal meaning constitute the top layer of the model.

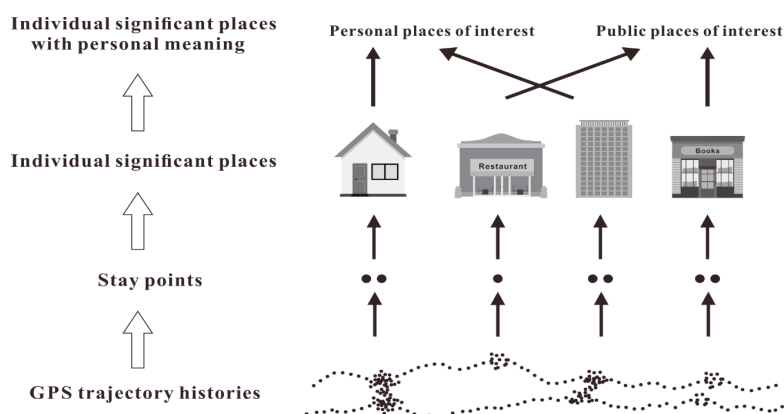


Figure 2. Identifying individual significant places with personal meaning using a four-layered model.

Stay point detection is a fundamental problem in trajectory studies and has been addressed by numerous researchers. Common solutions to the problem include (1) density-based methods [33,34] that are derived from the well-known density-based clustering algorithm DBSCAN, which incorporates physical parameters of trajectories, such as speed, acceleration and changes in direction; (2) spatio-temporal constraint-based methods [26,35], in which a stay point is detected when a sub-trajectory remains within a spatial region for longer than a certain time threshold and within a certain distance threshold; and, (3) index-based methods [36], in which customized indices are used to measure the status of each trajectory point. In this paper, we use the intuitive concept of stay points as a starting point and then apply the most popular approach, the spatio-temporal constraint-based method.

After detecting stay points from raw GPS trajectories, in the third layer, we cluster these points separately for each individual to find the individual significant places. Specifically, for each individual, we cluster all of the stay points detected from that individual's trajectories by applying the density-based clustering algorithm OPTICS; this procedure identifies places that are frequently visited by that individual. The OPTICS algorithm was selected from among the several available clustering methods because it is rather insensitive to the input parameters; thus, a broad range of parameter settings can produce results of similar quality [37].

3.2.2. Semantic Interpretation of Individual Significant Places

At the identified significant places, individuals frequently participate in various activities. Hence, individual significant places generally bear rich personal meanings. In this step, the semantics of such places are extracted for each individual from that individual's perspectives. The results of this step constitute the fourth layer of our model, as shown in Figure 2.

Most previous studies [9–13] have interpreted the semantics of individual significant places by means of reverse geocoding (i.e., comparing the locations of significant places to those of predefined POIs). As mentioned earlier, the two major drawbacks of these approaches are that (1) they consider only public places, and (2) they do not consider the personal meanings of those places. Inspired by the

idea that semantic information about public places can be derived from mobility data at the collective level [23,38], our solution is based on the assumption that at the individual level, the semantics of personally significant places can be derived from a person's long-term trajectory data. To avoid the problems that are associated with existing studies, the inherent subjectiveness of individuals is considered. Therefore, individual significant places are divided into two types: personal places of interest and public places of interest. The semantics of each type are interpreted separately.

Personal places of interest, such as homes and workplaces, normally exhibit high levels of visit frequency and temporal regularity. Generally speaking, individuals spend the most time at their homes in the evening and at their workplaces during the daytime. Accordingly, we estimate the semantics of personal places of interest, as follows:

We define a set of standard time intervals $\{STI_{ps}\}$, in which $STI_{ps} = [t_{arrive}, t_{leave}]$ is the typical temporal signature of a visit based on its personal semantics ps . For example, $STI_{home} = [00:00, 07:00] \cup [19:00, 24:00]$, and $STI_{work} = [08:00, 17:00]_{workday}$. Given the set of an individual i 's significant places $SP^i = \{SP_1^i, SP_2^i, \dots, SP_K^i\}$, as identified from i 's historical trajectories $TH^i = \{T_1^i, T_2^i, \dots, T_M^i\}$, for each SP_k^i in $SP^i = \{SP_1^i, SP_2^i, \dots, SP_K^i\}$, we calculate its matching score for the personal semantics ps . For each ps , the significant place with the highest matching score, calculated as shown below, is assigned the corresponding personal meaning.

$$match_{ps}(SP_k^i) = \frac{1}{|SP_k^i|} \sum_{s_{kj}^i \in SP_k^i} \frac{[s_{kj}^i.t_{arrive}, s_{kj}^i.t_{leave}] \cap [STI_p.t_{arrive}, STI_p.t_{leave}]}{[s_{kj}^i.t_{arrive}, s_{kj}^i.t_{leave}] \cup [STI_p.t_{arrive}, STI_p.t_{leave}]}, \quad (1)$$

Most individuals have only one personal place of interest with respect to a given personal meaning—the one that shows the greatest similarity to the corresponding temporal signature. However, it is unlikely that a similar one-to-one mapping can be established between temporal signatures and the visits to public places of interest. For instance, at night, people may frequently go shopping, or to a gym, a park, or a bar. Such visits have undistinguishable temporal signatures, even though their semantics are disparate. Thus, the geographic contextual information is used instead of temporal signatures to interpret the semantics of public places of interest.

Specifically, for a set of an individual i 's significant places $SP^i = \{SP_1^i, SP_2^i, \dots, SP_K^i\}$, the identified personal places of interest $PeSP^i$ are first filtered out, and the remaining places are then regarded as public places of interest $PuSP^i$. Next, the semantics of the public places of interest are extracted by associating each of these places with a spatial context (e.g., a POI). Given the set of remaining public places of interest $PuSP^i$, for each $PuSP_k^i$, we compute the distance r between the center coordinates of $PuSP_k^i$ and the farthest stay point in $PuSP_k^i$, and we construct a searching circle c of radius r (see Figure 3). Next, c is used to associate POIs with $PuSP_k^i$ and to interpret the corresponding semantics. If at least one POI is contained in c , then we annotate $PuSP_k^i$ with the category having the greatest numbers of POIs in c ; otherwise, we find the nearest POI and annotate $PuSP_k^i$ with its category. After extracting the semantics of $PuSP_k^i$, the temporal signatures of visits to $PuSP_k^i$ can be used for verification.



Figure 3. An example of extracting the semantics of public places of interest.

3.3. Mining Individual Similarity

Intuitively, the more commonality that two moving objects share, the more similar they are [39]. As a result, a universal measurement suitable for more general GIScience applications should be

proposed. Here, we present such a new measurement, the Individual Similarity Measurement considering interactions with Personally Significant Places (ISM-PSP). The basic assumptions of the ISM-PSP are that (1) the spatial and temporal interactions of individuals with their significant places can be used to mine the similarity among individuals, which can help to characterize individuals' features, and (2) the significant places of two individuals can be compared only when they are semantically identical from each individual's own perspective. The details of the individual similarity measurement process are shown in Algorithm 2.

Algorithm 2. IndividualSimilarityMeasurement (TH, SP_{ps}).

Input: TH: The set of individuals' trajectories TH = {THⁱ | 1 ≤ i ≤ |I|}

SP_{ps}: The set of individuals' significant places with personal meaning SP_{ps} = {SP_{ps}ⁱ | 1 ≤ i ≤ |I|}

Output: SimMatrix: Individual similarity matrix

Foreach a ∈ I do

 PS^a = GetPersonalSemantics(SP_{ps}^a);

 PS = PS^a; // personal semantics

 Foreach b ∈ I do

 PS^b = GetPersonalSemantics(SP_{ps}^b);

 PS = PS ∪ PS^b;

 SimMatrix(a,b) = 0;

 Foreach ps ∈ PS do

 Sim_{spatial}(a,b) = CalSpatialSim(a,b);

 Sim_{temporal}(a,b) = CalTemporalSim(a,b);

 Sim_{ps}(a,b) = w₁* Sim_{spatial}(a,b) + w₂* Sim_{temporal}(a,b);

 SimMatrix(a,b) += w_{ps}* Sim_{ps}(a,b);

 Return SimMatrix;

3.3.1. Grouping Individual Significant Places

Based on our assumptions, in the ISM-PSP calculation, the individual significant places are first grouped by those that share the same personal meaning. Given two individuals *a* and *b*, there are two sets of personal semantics, $PS^a = \{ps_1^a, ps_2^a, \dots, ps_n^a\}$ and $PS^b = \{ps_1^b, ps_2^b, \dots, ps_n^b\}$; these sets are extracted from *a*'s significant places SP^a and from *b*'s significant places SP^b , respectively. We group all of *a*'s significant places into n_a groups, each with identical personal semantics among all the members of that group: $SP_{ps}^a = \{SP_k^a | SP_k \text{ belongs to } a \text{ and } SP_k.\text{semantics} = ps\}$. Next, we similarly group *b*'s significant places into n_b groups: $SP_{ps}^b = \{SP_k^b | SP_k \text{ belongs to } b \text{ and } SP_k.\text{semantics} = ps\}$.

Example 1: Grouping the significant places of two individuals, *a* and *b* (Table 1). Five significant places $\{SP_1^a, SP_2^a, SP_3^a, SP_4^a, SP_5^a\}$ are identified from *a*'s historical trajectories. In accordance with the extracted semantics $PS^a = \{Home, Workplace, Restaurant, Bookstore\}$, *a*'s significant places are grouped into $Home = \{SP_2^a\}$, $Workplace = \{SP_3^a\}$, $Restaurant = \{SP_1^a, SP_5^a\}$ and $Bookstore = \{SP_4^a\}$. Seven significant places $\{SP_1^b, SP_2^b, SP_3^b, SP_4^b, SP_5^b, SP_6^b, SP_7^b\}$ are identified from *b*'s historical trajectories. In accordance with the corresponding extracted semantics $PS^b = \{Home, Workplace, Restaurant, Bookstore, Shopping mall\}$, *b*'s significant places are grouped into $Home = \{SP_3^b\}$, $Workplace = \{SP_1^b\}$, $Restaurant = \{SP_2^b\}$, $Book store = \{SP_4^b, SP_7^b\}$ and $Shopping mall = \{SP_5^b\}$.

Table 1. An example of grouping the significant places of two individuals, *a* and *b*.

Place Semantics	<i>a</i> 's Significant Places	<i>b</i> 's Significant Places
Home	SP_2^a	SP_3^b
Workplace	SP_3^a	SP_1^b
Restaurant	SP_1^a, SP_5^a	SP_2^b
Bookstore	SP_4^a	SP_4^b, SP_7^b
Shopping mall		SP_5^b

3.3.2. Measuring Individual Similarity

After the significant places have been grouped based on the diverse personal place semantics of each individual, the ISM-PSP measures the similarity between individual a and b , as follows:

$$Sim(a, b) = \sum_{ps \in PS^a \cup PS^b} w_{ps} * Sim_{ps}(a, b), \quad (2)$$

where w_{ps} is the weight of personal semantics ps and is determined by the level of the importance of ps to the corresponding applications, and $Sim_{ps}(a, b)$ is the similarity score of a and b for their specific personal semantics ps , and is calculated as follows:

$$Sim_{ps}(a, b) = w_1 * Sim_{spatial}(\{SP_{ps}^a\}, \{SP_{ps}^b\}) + w_2 * Sim_{temporal}(\{SP_{ps}^a\}, \{SP_{ps}^b\}), \quad (3)$$

In this way, we convert the individual similarity measurements into a sum over the set of spatial and temporal similarities between two sets. In the ISM-PSP approach, the similarity between two individuals is determined by measuring their spatio-temporal similarity for every type of personal semantics $ps \in PS^a \cup PS^b$. For a given $ps \in PS^a \cup PS^b$, we compute the spatial and temporal similarity between SP_{ps}^a and SP_{ps}^b using the following equation:

$$Sim_{spatial|temporal}(\{SP_{ps}^a\}, \{SP_{ps}^b\}) = \frac{\sum_{SP_k^a \in SP_{ps}^a, SP_k^b \in SP_{ps}^b} I(SP_k^a, SP_k^b)}{|\{SP_{ps}^a\}| + |\{SP_{ps}^b\}|}, \quad (4)$$

where $I(SP_k^a, SP_k^b)$ is an indicator function that is defined as follows:

$$I(SP_k^a, SP_k^b) = \begin{cases} 1 & \text{if } dist(SP_k^a, SP_k^b) \leq distThreshold \\ 0 & \text{otherwise} \end{cases}, \quad (5)$$

Regarding spatial distance, as described in Definition 3, a significant place SP_k consists of a collection of stay points. The coordinates of SP_k are represented by the average latitude and longitude of the constituent points. The Euclidean distance between the coordinates of SP_k^a and SP_k^b is used as the spatial distance measurement.

Regarding temporal distance, we measure the temporal difference between a 's visit to SP_k^a and b 's visit to SP_k^b by dividing the day into 24 h and constructing an hourly distribution of each individual's visits to his or her significant places. Given the probability distributions $pd_1(t)$ and $pd_2(t)$ of two individuals' visits to their respective significant places, when the two places are semantically identical, their temporal distance is measured by the Kullback-Leibler divergence of pd_1 and pd_2 , as follows:

$$dist_{temporal}(SP_k^a, SP_k^b) = \frac{1}{2} (D_{KL}(pd_1(t)||pd_2(t)) + D_{KL}(pd_2(t)||pd_1(t))), \quad (6)$$

$$D_{KL}(pd_1(t)||pd_2(t)) = \sum pd_1(t) \log \frac{1}{pd_2(t)}, \quad (7)$$

To address the case in which $D_{KL}(pd_1(t)||pd_2(t))$ becomes infinite when $pd_1 \neq 0$ but $pd_2 = 0$, a small constant C is introduced, and the Kullback-Leibler divergence is computed using a smoothing method.

4. Experiments

In this section, several experiments with the real-world Geolife dataset are performed to evaluate our proposed framework. The datasets and their preparation are described in Section 4.1. Section 4.2 corresponds to phase 1 of our proposed framework and presents the results of semantics extraction of personally significant places. In Section 4.3, to assess the performance of the proposed ISM-PSP

in phase 2 of our proposed framework, we perform comparative experiments with two previous approaches and a modified version of the ISM-PSP. To illustrate the possible applications of our framework, we also use the proposed method to generate individual groups in Section 4.4.

4.1. Dataset

4.1.1. GPS Trajectory Dataset

The GPS trajectory dataset was collected in the Geolife project by 182 users over a period of more than five years (from April 2007 to August 2012) [35]. This dataset covers widely distributed areas, including over 30 cities in China and several cities that are located in the United States of America (USA) and Europe. In our experiments, we used only those trajectories from Beijing, which constitute the majority of the Geolife dataset.

The Geolife dataset contains records for a broad range of individuals' outdoor movements, covering daily routines such as going home and going to work as well as leisure activities, such as dining and shopping. In other words, this dataset includes visits to both personal and public places of interest.

4.1.2. POI Dataset

To interpret the semantics of public places of interest in Beijing, we obtained a dataset of public POIs in Beijing from Dianping. As shown in Table 2, this Beijing POI dataset includes 181,924 POIs, corresponding to eight major types of individual daily activities.

Table 2. Beijing points of interest (POI) dataset.

Type ID	Type	Count
1	Shopping	72,370
2	Restaurant	52,567
3	Education	21,224
4	Hotel	15,272
5	Sports	10,557
6	Life service	9035
7	Entertainment	784
8	Healthcare	115
Total		181,924

4.2. Semantics Extraction of Significant Places

This experiment was designed to interpret the semantics of individual significant places extracted from the Geolife historical trajectory data.

First, we used each individual's GPS trajectory histories to detect stay points by applying the spatio-temporal constraint-based method with the *distance threshold* set to 30 m and the *time threshold* set to 30 min. In total, 19,374 stay points were detected from the raw GPS trajectories of 182 individuals in this stage. Next, the OPTICS algorithm was applied to each individual's collected stay point data to identify that individual's significant places. During this process, we set the *reachability-distance* to 100 m and the value of *MinPts* to 10; consequently, in our experiment, a place was considered to be significant to an individual only when it was visited more than 10 times by that individual. A total of 154 significant places were discovered, and 63 of the 182 individuals had at least one significant place. Among these 63 individuals, the median number of significant places possessed by each individual was 2, and the maximum number was 10 (corresponding to individual #4). Since it can be inferred that the spatio-temporal interactions between individuals and places are generally similar in the Geolife dataset, we set the parameters relatively strictly to improve the accuracy of significant places identification. As a result, there were many individuals for whom no significant places were captured.

Then, we interpreted the semantics of the identified significant places. The semantics of two types of personal places of interest were extracted in this experiment: homes and workplaces. We set

$STI_{home} = [00:00, 07:00] \cup [19:00, 24:00]$ and $STI_{work} = [08:00, 17:00]_{workday}$. For each identified significant place, the home matching score was calculated first. If the highest home matching score for an individual exceeded the threshold value ϵ (ϵ was set to 0.3 in our experiment), the significant place corresponding to that matching score was interpreted as that individual's home. After filtering out the significant places identified as individuals' homes, we computed the workplace matching scores for the remaining significant places. Similarly, when the highest workplace matching score for an individual exceeded ϵ (ϵ was set to 0.3 in our experiment), the significant place corresponding to that matching score was interpreted as that individual's workplace. Figure 4 shows the results of semantics extraction. Regarding the personal places of interest, the identified homes were all located in the northern part of Beijing, whereas the identified workplaces were mostly located in the northwestern part of the city. This is because most of the individuals who participated in the Geolife program came from academic institutions in northwestern Beijing, such as Tsinghua University or Microsoft Research Asia. In addition, nine homes and 52 workplaces were discovered. The reason that more personal places of interest were identified as workplaces than as homes may be because the participating individuals tended to record their trajectories more often during the daytime than at night. Because we considered a place to be an individual's home only when his or her stay duration in that place sufficiently overlapped with STI_{home} , sparse records may have caused failures in home identification.

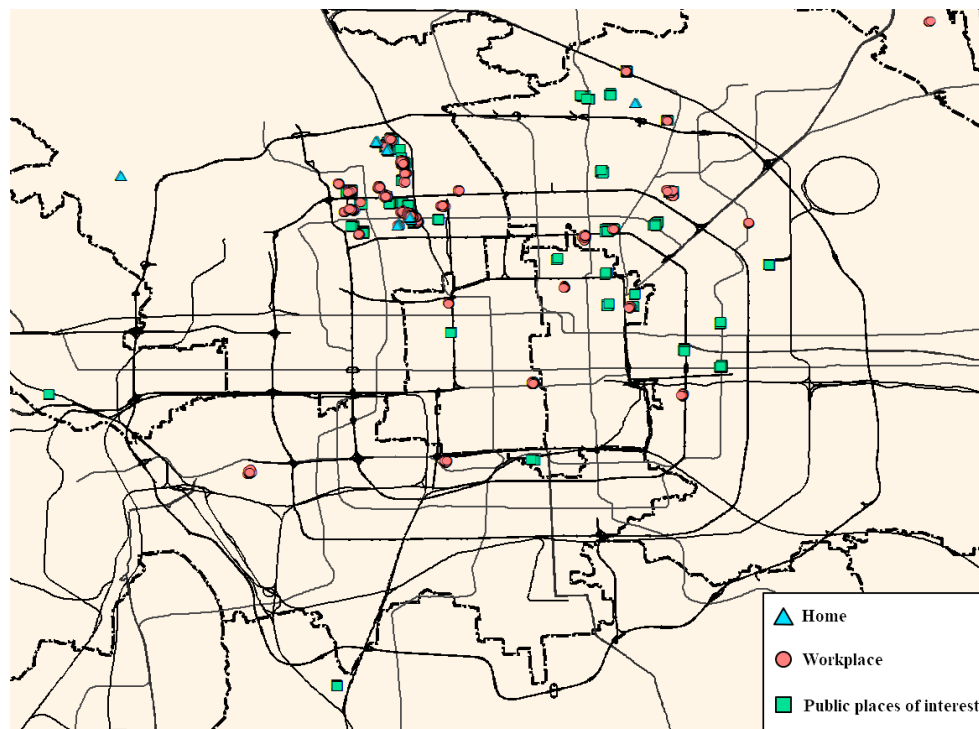


Figure 4. Semantic interpretation of significant places.

After the personal places of interest were excluded, the remaining significant places were compared to the Beijing POI dataset using the method that is described in Section 3.2.2. As shown in Table 3, 50 public places of interest were identified in total, and the majority of which were restaurants, education facilities and shopping locations. These public places of interest are distributed, as shown in Figure 4. We found that (1) many public places of interest were in close proximity to individuals' personal places of interest, implying that most individuals' leisure activities occurred near their homes and workplaces, and (2) other public places of interest were generally located in commercial centers, such as the eastern part of Beijing, which contains several major super-malls that are capable of satisfying individuals' higher entertainment needs.

Table 3. Public places of interest identified.

Type ID	Type	Count
1	Shopping	11
2	Restaurant	14
3	Education	13
4	Hotel	5
5	Sports	5
6	Life service	2
Total		50

4.3. Comparative Analysis of Individual Similarity Measurements

To evaluate the performance of the proposed framework for measuring individual similarity and to demonstrate its ability to discern individuals, a comparative analysis was performed between our method and the existing MSTP [9] and MTP [27] methods. In this experiment, each individual's trajectory data were divided into two parts, which were then treated as trajectories from two different people. Specifically, for each individual i 's trajectory data, trajectories from odd weeks were assigned to u^i , and trajectories from even weeks were assigned to v^i . In this way, we generated two datasets, called Dataset #1 and Dataset #2. To implement the MSTP and MTP methods, the Beijing POI dataset was used to transform raw GPS trajectories into semantic trajectories. In addition, to verify the necessity of considering personal meaning when extracting the semantics of significant places, phase 1 of our proposed framework was modified to consider only public places.

For each individual u^i in Dataset #1, four different individual similarity measurements were applied to find u^i 's neighbors in Dataset #2 and the neighbors were sorted in descending order by their similarity scores. Here, the ground truth concerning the nearest neighbors was known, because for every individual, u^i and v^i were artificially generated from i . Thus, ideally, v^i should appear in the first position of the neighbors list for u^i .

To evaluate the performance of the individual similarity measurements, we defined a metric called the *average rank*, as follows. Given an individual u^i , suppose that the obtained similarity score between u^i and its neighbor nb_l is denoted by $sim(u^i, nb_l)$ and that the ordered neighbors list is denoted by N^{u^i} , where $N^{u^i} = \{(nb_1, sim(u^i, nb_1)), (nb_2, sim(u^i, nb_2)), \dots, (nb_m, sim(u^i, nb_m))\}$. Then, the average rank of the neighbor nb_l , denoted by $\bar{R}(nb_l)$, is defined as the average of nb_l 's best and worst possible ranks in the ordered neighbors list, and is calculated as follows:

$$\bar{R}(nb_l) = \frac{1}{2}(|\{nb | sim(u^i, nb) > sim(u^i, nb_l)\}| + |\{nb | sim(u^i, nb) \geq sim(u^i, nb_l)\}| + 1), \quad (8)$$

Example 2: Calculating the average rank. For the individual u^{000} , three different similarity measurements are applied to calculate similarity scores and identify the neighbors. The results of the three similarity measurements are $\{(v^{001}, 1), (v^{000}, 0.8), (v^{002}, 0.3)\}$ (case #1), $\{(v^{000}, 0.9), (v^{001}, 0.9), (v^{002}, 0.5)\}$ (case #2) and $\{(v^{000}, 0.9), (v^{001}, 0.5), (v^{002}, 0.5)\}$ (case #3). Next, we calculate the average rank of v^{000} (which is the true nearest neighbor of u^{000}) for the three different cases, as follows:

$$\text{Case \#1 : } \bar{R}(v^{000}) = \frac{1}{2}(|\{v^{001}\}| + |\{v^{001}, v^{000}\}| + 1) = \frac{1}{2}(1 + 2 + 1) = 2,$$

$$\text{Case \#2 : } \bar{R}(v^{000}) = \frac{1}{2}(|\emptyset| + |\{v^{000}, v^{001}\}| + 1) = \frac{1}{2}(0 + 2 + 1) = 1.5,$$

$$\text{Case \#3 : } \bar{R}(v^{000}) = \frac{1}{2}(|\emptyset| + |\{v^{000}\}| + 1) = \frac{1}{2}(0 + 1 + 1) = 1,$$

Using the average rank, we can evaluate the results of the similarity measurements. We choose to use the average rank instead of the absolute rank because this metric represents the ability of the similarity measurements to find identical individuals, as well as to discern different individuals. In example 2, although v^{000} was found to be the most similar to u^{000} in both case #2 and case #3, $\bar{R}(v^{000})$ in case #3 is smaller because this was the only nearest neighbor obtained, whereas in case #2, v^{001} was also considered to be the most similar to u^{000} . For each individual u^i in Dataset #1, we calculated the average rank of its corresponding ground-truth identical individual v^i . A smaller $\bar{R}(v^i)$ value indicates a better similarity result. In the ideal case, $\bar{R}(v^i)$ should be equal to 1, meaning that v^i was identified as the only nearest neighbor of u^i . Therefore, we defined the average rank score to evaluate the performance of an individual similarity measurement by incorporating its ability to find the true identical individual v^i for each u^i . A better individual similarity measurement will have a higher average rank score. If for every u^i , a similarity measurement could always achieve $\bar{R}(v^i) = 1$, its average rank score would be equal to 1.

$$\text{Average rank score} = \frac{1}{|U|} \sum_{u^i \in U} \frac{1}{\bar{R}(v^i)}, \quad (9)$$

Figure 5 shows the total numbers of v^i whose average rank is less than or equal to various \bar{R} values using the MSTP approach, the MTP approach, the ISM-PSP approach without personal semantics (ISM-PSP without ps) and the proposed ISM-PSP approach (which considers the perspective of personal behavior). As mentioned earlier, a better similarity measurement should identify the true nearest neighbors at smaller \bar{R} values more often. The results show that the ISM-PSP method yields the highest number of v^i with $\bar{R} = 1$. Thus, the ISM-PSP results are the closest to the ground truth, indicating that the ISM-PSP is more capable of finding identical individuals than the other measurements are. In addition, both the ISM-PSP and the ISM-PSP without ps achieve better performance than the MSTP and the MTP approaches at smaller \bar{R} values, because using the spatio-temporal properties of visits to significant places is more accurate than using the sequential properties. We note that in this case, the MSTP approach performs better than the MTP approach because when comparing two individuals u^i and v^i , for each LCS of u^i , the MSTP compares all of the LCSs of v^i , while the MTP compares only the nearest LCS of v^i . Therefore, when the LCSs are generally short, as in our dataset, the MTP approach fails to obtain good results because of the lack of information. As \bar{R} increases, the ISM-PSP still achieves the best results. However, for large \bar{R} values, a large number of v^i , for which the average rank is less than or equal to \bar{R} is not necessarily a guarantee of good performance. An unsatisfactory similarity measurement could also generate many v^i of average rank \leq a large \bar{R} value by increasing the $sim(u^i, nb_l)$ values of more candidates nb_l . In other words, such a method may appear to achieve good performance in finding identical individuals, but it improves its probability of including u^i 's identical individual v^i in the resulting neighbor list by including as many nb as possible. In this way, as \bar{R} increases, more identical individuals v^i are identified.

Therefore, we further compared the four individual similarity measurements based on their ability to identify v^i in the first position (by absolute position rank), which reflects the probability with which they identify v^i as the most similar to u^i ; we also assessed the average rank score, which represents the ability to discern other different individuals. As shown in Figure 6, both the ISM-PSP and the MTP approaches have a high probability of successfully identifying v^i in the first position. However, the average rank score of the MTP is much lower than that of the ISM-PSP, which means that for those v^i in the first position obtained by the MTP, $\bar{R}(v^i)$ is generally large. This finding indicates that the MTP achieves its high probability of finding identical individuals at the cost of generating many false nearest neighbors. We also compared the ISM-PSP results with and without ps , as shown in Figures 5 and 6. These figures show that the ISM-PSP approach is more capable of discerning individuals when place semantics are considered with personal meaning than when only the semantics of public places without personal meaning are interpreted. This result suggests that it is insufficient to assess the distinctive characteristics of individuals based solely on their visits to public places, as most previous

studies have done. Thus, personally significant places are helpful in characterizing individuals and should not be neglected.

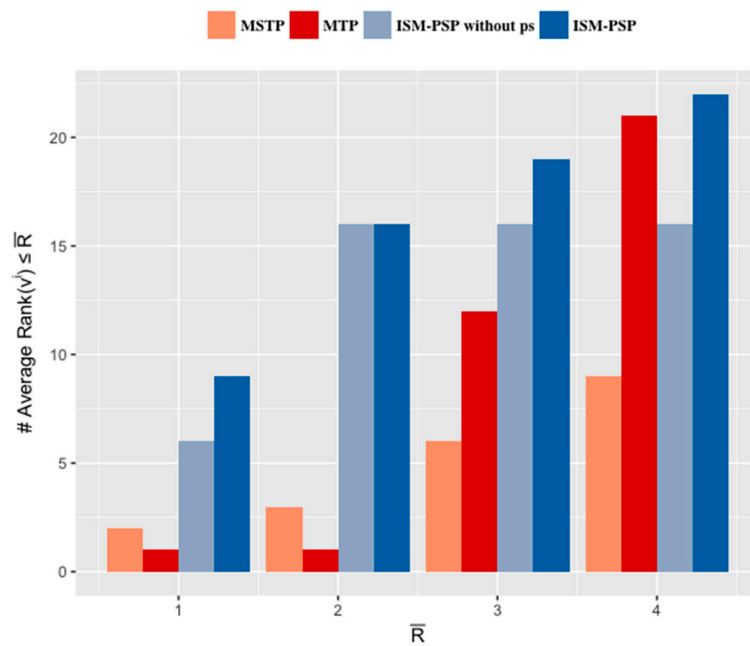


Figure 5. Numbers of v^j identified for which the average rank is less than or equal to \bar{R} .

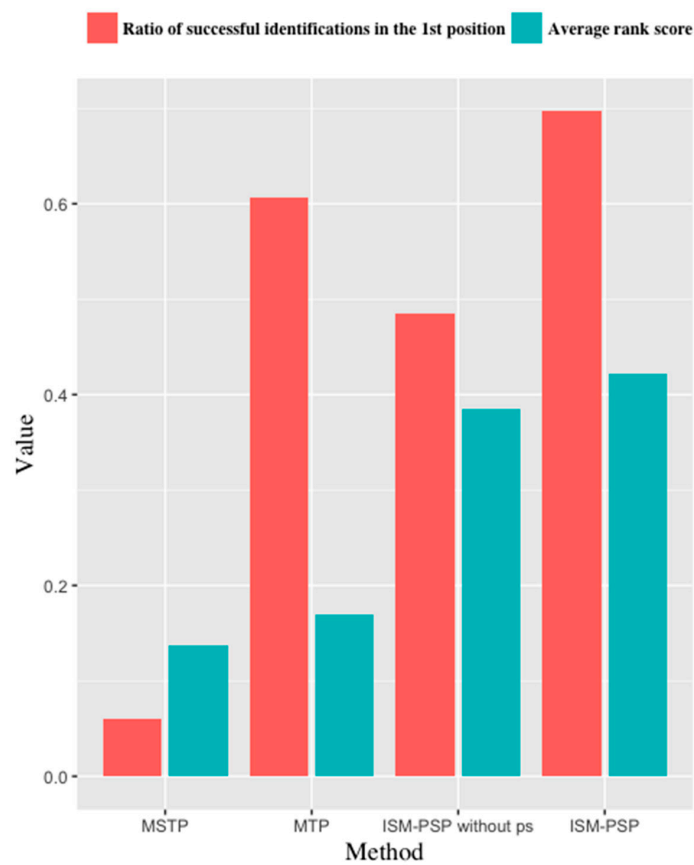


Figure 6. Ratio of identifying v^j in the first position and average rank scores using different methods.

4.4. Grouping Individuals

Our proposed framework allows for us to investigate individual similarity using any single or combined semantics. Specifically, by setting different personal semantics ps in the ISM-PSP, one can compute the individual similarities in terms of different semantics aspects. Here, only personal places of interest were used in the example application.

In this experiment, we first mined individual similarities based solely on their visits interpreted as “going to work” ($ps = \{Workplace\}$). We set $distThre_{spatial} = 300$ m and $distThre_{temporal} = 1$, and we assigned two individuals to the same group only when their similarity was equal to 1. Figure 7 shows that four groups were identified, all of which correspond to research institutes and universities in northwestern Beijing. The largest group (Group #1) is Microsoft Research Asia, located on Zhichun Road in Haidian District; Group #2 is near the School of Software of Tsinghua University; Group #3 is in the Chinese Academy of Sciences; and, Group #4 is the Beijing University of Aeronautics and Astronautics. The results reflect the fact that most participants in the Geolife project worked at the research institutes and universities listed above [40]. Our method successfully identified their workplaces and aggregated individuals into different groups based on the spatio-temporal similarity of their visits to their workplaces.

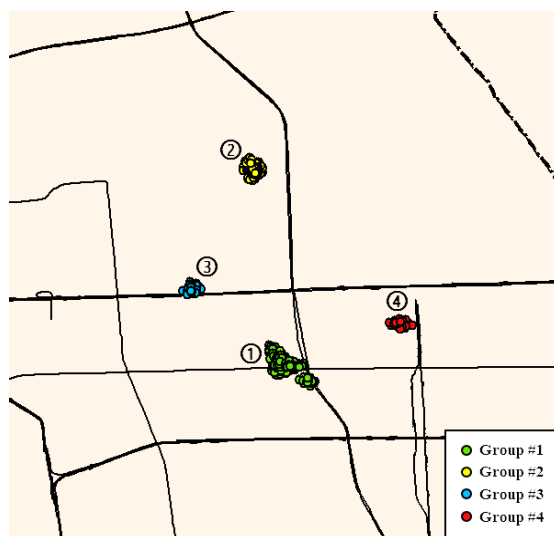


Figure 7. Aggregating individuals based on the similarity of their visits to their workplaces.

After identifying individuals who exhibited high similarity in their visits to their workplaces, we also took the individuals’ homes into consideration to discover groups whose visits to both their homes and workplaces were similar ($ps = \{Home, Workplace\}$). As shown in Figure 8, only two groups were identified under this constraint. Group #1 included individuals who worked at Microsoft Research Asia and lived at Tsinghua University, and Group #2 included individuals whose homes and workplaces were both at Tsinghua University. We inferred that Group #1 could include students from Tsinghua University who were interns at Microsoft Research Asia, whereas Group #2 consisted of students or staff who both lived and studied or worked at Tsinghua University. We separately calculated the temporal distributions of homes and workplaces visits for the individuals in Groups #1 and #2 (Figure 9). Substantial differences were found between the temporal signals of the home and workplace visits in both groups. When compared with the individuals in Group #2, the individuals in Group #1 were less likely to be found at home during the daytime, which is consistent with the daily routines of an internship.

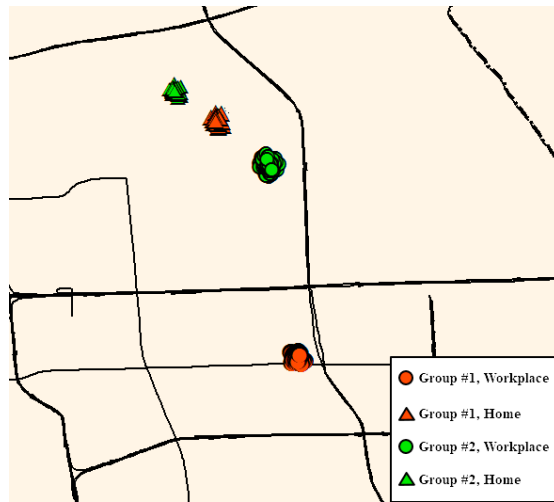


Figure 8. Grouping individuals based on the similarity of their visits to both their homes and workplaces.

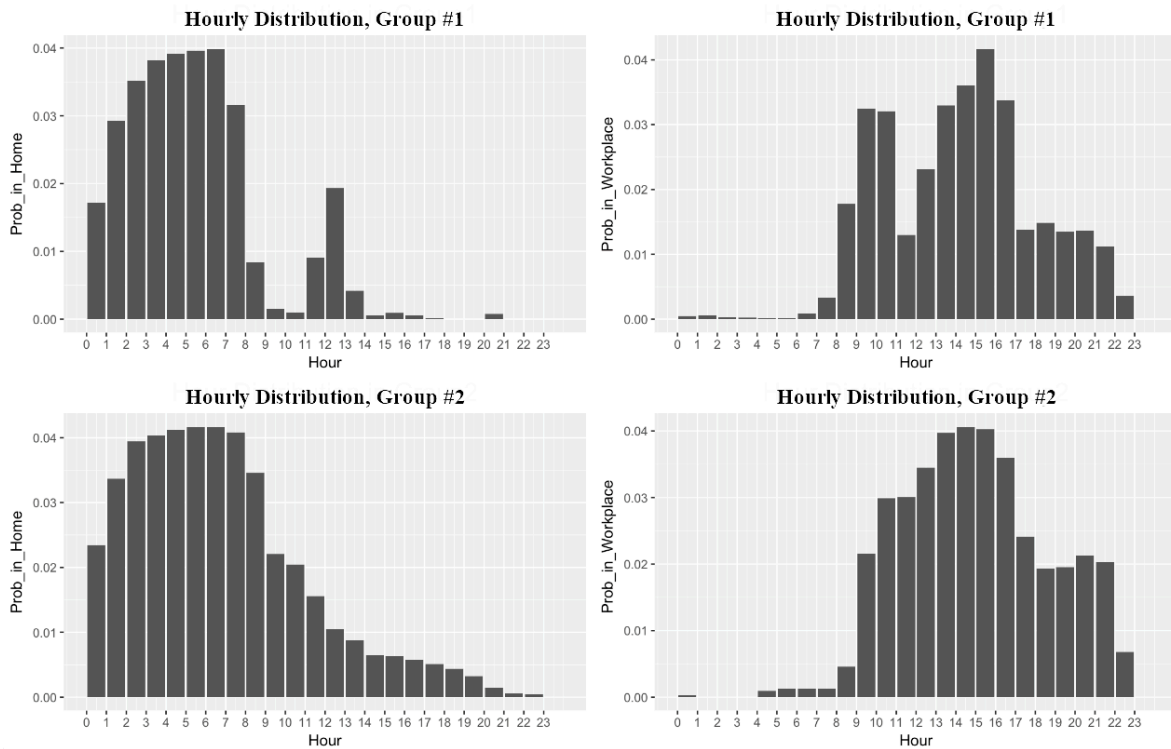


Figure 9. Temporal signatures of the visits of individuals in Groups #1 and #2 to their homes and workplaces.

5. Conclusions

Individuals have a remarkable propensity to return to their frequently visited places. Hence, the interactions between individuals and these places are likely to represent individuals’ characteristics. To facilitate the capture of these characteristics of individuals and the mining of their similarity, this study investigated how individuals spatially and temporally interact with their personally significant places. A framework was presented for mining individual similarity based on visits to personally significant places extracted from long-term trajectory data. Our framework includes two

major phases: extracting the semantics of personally significant places and mining individual similarity. In contrast to many previous studies, we extracted place semantics with personal meaning, and our semantic extraction process considered individuals' visits to both personal and public places of interest. We also proposed a new individual similarity measurement, the ISM-PSP, which incorporates both the spatio-temporal and semantic properties of individuals' visits to significant places. Experiments using a real-world GPS dataset suggest that (1) when compared with the existing approaches, the proposed ISM-PSP is more capable of finding identical individuals, while maintaining low numbers of false identifications; (2) more accurate identification of individuals can be achieved by considering the spatio-temporal properties of visits to significant places than by considering the sequential properties; and, (3) personal places of interest play a vital role in characterizing individuals, which indicates that the semantics of visits to significant places with personal meaning are important for assessing individual similarity. Therefore, we conclude that it is insufficient to measure individual similarity by only analyzing the sequential properties of visits to public places, as done in previous works.

Our study has several limitations. First, when extracting personal places of interest, we only identified the most common types (homes and workplaces) for illustration. However, personal places of interest actually include a much wider range of places. According to Definition 4, a personal place of interest to an individual could be any place that carries a special personal meaning that is distinct from its functionality for the general public. Different types of personal places of interest should be identified for specific applications. Second, during the semantic interpretation process, we inferred the semantics of personal places of interest by comparing the temporal distribution of a person's presence at a place against certain predefined typical temporal signatures. However, there are many people who deviate from a standard work schedule, and others may work at home. In these cases, our method presented in Section 3.2.2 could fail to extract the accurate semantics. An alternative method could be to identify homes as places where individuals spend most of their time at night and workplaces as places where individuals spend most of their time during the daytime on workdays. Third, our similarity measurement results could be significantly affected by the accuracy of semantic interpretation. This is because we designed the ISM-PSP based on the assumption that the spatio-temporal patterns of two visits are comparable only when they are driven by the same reason. In other words, we do not compare one individual's working behavior with another's dining behavior, although they might appear at the same restaurant. Therefore, errors in place semantics extraction could lead to poor results in measuring individual similarity. Fourth, although the proposed framework enables us to generate meaningful subgroups in any single or combined semantics by setting different ps in the ISM-PSP, in this paper, only personal places of interest were used to demonstrate the possible application of the proposed framework. The sparse records in our dataset restricted the types of significant places we were able to discover, thus restraining the types of groups that we could identify.

Future work will focus on improving the semantics enrichment process applied in the proposed framework. Other publicly available data in social networks (e.g., georeferenced posts on Twitter) can be used to explore place semantics [25,41–44]. Through the synthesis of individual similarities related to appropriate semantics, our similarity measurement could be applied in other fields using different datasets. For example, our approach could reveal meaningful relationships by identifying individuals who work together and who also share high a spatio-temporal similarity in their visits to certain restaurants or bars.

Acknowledgments: This study was financially supported by the National Key Technology Research and Development Program of China (Grant No. 2017YFB0503700) and the High-Resolution Earth Observation System National Key Foundation of China (Grant Nos. 11-Y20A02-9001-16/17 and 30-Y20A01-9003-16/17).

Author Contributions: Mengke Yang conceived, designed and performed the experiments, and wrote the manuscript; Chengqi Cheng supervised the study; and Bo Chen offered helpful suggestions and reviewed the manuscript. All authors have read and approved the submitted manuscript, have agreed to be listed, and have accepted this version for publication.

Conflicts of Interest: The authors declare that they have no conflicts of interest to disclose.

References

1. Yuan, Y.; Raubal, M. Measuring similarity of mobile phone user trajectories—A spatio-temporal edit distance method. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 496–520. [[CrossRef](#)]
2. Andrienko, G.L.; Andrienko, N.V. Extracting patterns of individual movement behaviour from a massive collection of tracked positions. In Proceedings of the Workshop on Behaviour Monitoring and Interpretation, Bremen, Germany, 10 September 2007; pp. 1–16.
3. Ye, Y.; Zheng, Y.; Chen, Y.; Feng, J.; Xie, X. Mining Individual Life Pattern Based on Location History. In Proceedings of the Tenth International Conference on Mobile Data Management: Systems, Services and Middleware, Taipei, Taiwan, 18–20 May 2009; pp. 1–10.
4. Kang, C.; Gao, S.; Xiao, Y.; Yuan, Y.; Liu, Y.; Ma, X. Analyzing and geo-visualizing individual human mobility patterns using mobile call records. In Proceedings of the 2010 18th International Conference on Geoinformatics, Beijing, China, 18–20 June 2010.
5. Zheng, Y.; Zhang, L.Z.; Ma, Z.X.; Xie, X.; Ma, W.Y. Recommending friends and locations based on individual location history. *ACM Trans. Web* **2011**, *5*. [[CrossRef](#)]
6. Zhu, L.; Xu, C.Q.; Guan, J.F.; Zhang, H.K. SEM-PPA: A semantical pattern and preference-aware service mining method for personalized point of interest recommendation. *J. Netw. Comput. Appl.* **2017**, *82*, 35–46. [[CrossRef](#)]
7. Trasarti, R.; Guidotti, R.; Monreale, A.; Giannotti, F. Myway: Location prediction via mobility profiling. *Inf. Syst.* **2017**, *64*, 350–367. [[CrossRef](#)]
8. Xiao, X.; Zheng, Y.; Luo, Q.; Xie, X. ST-DMQL: A Semantic Trajectory Data Mining Query Language. *J. Ambient Intell. Humaniz. Comput.* **2014**, *5*, 3–19. [[CrossRef](#)]
9. Ying, J.J.-C.; Lu, E.H.-C.; Lee, W.-C.; Weng, T.-C.; Tseng, V.S. Mining user similarity from semantic trajectories. In Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks, San Jose, CA, USA, 2 November 2010; ACM: New York, NY, USA, 2010; pp. 19–26.
10. Liu, Y.; Seah, H.S. Points of interest recommendation from gps trajectories. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 953–979. [[CrossRef](#)]
11. Zhu, L.; Xu, C.; Guan, J.; Yang, S. Finding top-k similar users based on trajectory-pattern model for personalized service recommendation. In Proceedings of the 2016 IEEE International Conference on Communications Workshops (ICC), Kuala Lumpur, Malaysia, 23–27 May 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 553–558.
12. Liu, J.; Wolfson, O.; Yin, H. Extracting semantic location from outdoor positioning systems. In Proceedings of the 7th International Conference on Mobile Data Management, (MDM 2006), Nara, Japan, 10–12 May 2006; IEEE: Los Alamitos, CA, USA, 2006; p. 73.
13. Comito, C.; Falcone, D.; Talia, D. Mining human mobility patterns from social geo-tagged data. *Pervasive Mob. Comput.* **2016**, *33*, 91–107. [[CrossRef](#)]
14. Spaccapietra, S.; Parent, C.; Damiani, M.L.; de Macedo, J.A.; Porto, F.; Vangenot, C. A conceptual view on trajectories. *Data Knowl. Eng.* **2008**, *65*, 126–146. [[CrossRef](#)]
15. Parent, C.; Spaccapietra, S.; Renso, C.; Andrienko, G.; Andrienko, N.; Bogorny, V.; Damiani, M.; Gkoulalas-Divanis, A.; Macedo, J.; Pelekis, N.; et al. Semantic trajectories modeling and analysis. *ACM Comput. Surv. (CSUR)* **2013**, *45*, 1–32. [[CrossRef](#)]
16. 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, Uppsala, Sweden, 21–24 March 2011; ACM: New York, NY, USA, 2011; pp. 259–270.
17. Cao, X.; Cong, G.; Jensen, C.S. Mining significant semantic locations from GPS data. *Proc. VLDB Endow.* **2010**, *3*, 1009–1020. [[CrossRef](#)]
18. Papandrea, M.; Jahromi, K.K.; Zignani, M.; Gaito, S.; Giordano, S.; Rossi, G.P. On the properties of human mobility. *Comput. Commun.* **2016**, *87*, 19–36. [[CrossRef](#)]
19. Alvares, L.; Bogorny, V.; Kuijpers, B.; de Macedo, J.; Moelans, B.; Vaisman, A. A model for enriching trajectories with semantic geographical information. In Proceedings of the 15th Annual ACM International Symposium on Advances in Geographic Information Systems, Seattle, WA, USA, 7–9 November 2007; ACM: New York, NY, USA, 2007; pp. 1–8.

20. Bogorny, V.; Kuijpers, B.; Alvares, L.O. ST-DMQL: A semantic trajectory data mining query language. *Int. J. Geogr. Inf. Sci.* **2009**, *23*, 1245–1276. [[CrossRef](#)]
21. Xiao, X.; Zheng, Y.; Luo, Q.; Xie, X. Finding similar users using category-based location history. In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, San Jose, CA, USA, 2–5 November 2010; ACM: New York, NY, USA, 2010; pp. 442–445.
22. Spinsanti, L.; Celli, F.; Renso, C. Where you stop is who you are: Understanding people’s activities by places visited. In Proceedings of the Behaviour Monitoring and Interpretation (BMI) Workshop, Karlsruhe, Germany, 21 September 2010; pp. 38–52.
23. Ye, M.; Janowicz, K.; Mülligann, C.; Lee, W.-C. What you are is when you are: The temporal dimension of feature types in location-based social networks. In Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Chicago, IL, USA, 1–4 November 2011; ACM: New York, NY, USA, 2011; pp. 102–111.
24. Shen, J.N.; Cheng, T. A framework for identifying activity groups from individual space-time profiles. *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 1785–1805. [[CrossRef](#)]
25. Andrienko, G.; Andrienko, N.; Fuchs, G.; Raimond, A.-M.O.; Symanzik, J.; Ziemlicki, C. Extracting semantics of individual places from movement data by analyzing temporal patterns of visits. Presented at the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL GIS 2013), Orlando, FL, USA, 5–8 November 2013; ACM: New York, NY, USA, 2013; pp. 9–16.
26. Li, Q.; Zheng, Y.; Xie, X.; Chen, Y.; Liu, W.; Ma, W.-Y. Mining user similarity based on location history. In Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Irvine, CA, USA, 5–7 November 2008; ACM: New York, NY, USA, 2008; pp. 1–10.
27. Chen, X.; Pang, J.; Xue, R. Constructing and comparing user mobility profiles. *ACM Trans. Web (TWEB)* **2014**, *8*, 1–25. [[CrossRef](#)]
28. Mazumdar, P.; Patra, B.K.; Lock, R.; Korra, S.B. An approach to compute user similarity for GPS applications. *Knowl.-Based Syst.* **2016**, *113*, 125–142. [[CrossRef](#)]
29. Lv, M.; Chen, L.; Chen, G. Mining user similarity based on routine activities. *Inf. Sci.* **2013**, *236*, 17–32. [[CrossRef](#)]
30. González, M.C.; Hidalgo, C.A.; Barabási, A.-L. Understanding individual human mobility patterns. *Nature* **2008**, *453*, 779–782. [[CrossRef](#)] [[PubMed](#)]
31. Song, C.; Koren, T.; Wang, P.; Barabási, A. Modelling the scaling properties of human mobility. *Nat. Phys.* **2010**, *6*, 818–823. [[CrossRef](#)]
32. Furtado, A.S.; Kopanaki, D.; Alvares, L.O.; Bogorny, V. Multidimensional similarity measuring for semantic trajectories. *Trans. GIS* **2016**, *20*, 280–298. [[CrossRef](#)]
33. Palma, A.; Bogorny, V.; Kuijpers, B.; Alvares, L. A clustering-based approach for discovering interesting places in trajectories. In Proceedings of the 2008 ACM Symposium on Applied Computing, Fortaleza, Brazil, 16–20 March 2008; ACM: New York, NY, USA, 2008; pp. 863–868.
34. Rocha, J.A.M.R.; Times, V.C.; Oliveira, G.; Alvares, L.O.; Bogorny, V. DB-SMoT: A direction-based spatio-temporal clustering method. In Proceedings of the 2010 5th IEEE International Conference Intelligent Systems (IS), London, UK, 7–9 July 2010; pp. 114–119.
35. Zheng, Y.; Zhang, L.; Xie, X.; Ma, W.-Y. Mining interesting locations and travel sequences from GPS trajectories. In Proceedings of the 18th International Conference on World Wide Web, Madrid, Spain, 20–24 April 2009; ACM: New York, NY, USA, 2009; pp. 791–800.
36. Huang, G.; He, J.; Zhou, W.; Huang, G.-L.; Guo, L.; Zhou, X.; Tang, F. Discovery of stop regions for understanding repeat travel behaviors of moving objects. *J. Comput. Syst. Sci.* **2016**, *82*, 582–593. [[CrossRef](#)]
37. Ankerst, M.; Breunig, M.M.; Kriegel, H.-P.; Sander, J. Optics: Ordering points to identify the clustering structure. In Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data, Philadelphia, PA, USA, 31 May–3 June 1999; Volume 28, pp. 49–60.
38. Liu, Y.; Liu, X.; Gao, S.; Gong, L.; Kang, C.G.; Zhi, Y.; Chi, G.H.; Shi, L. Social sensing: A new approach to understanding our socioeconomic environments. *Ann. Assoc. Am. Geogr.* **2015**, *105*, 512–530. [[CrossRef](#)]
39. Lin, D. An information-theoretic definition of similarity. In Proceedings of the Fifteenth International Conference on Machine Learning, Madison, WI, USA, 24–27 July 1998; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 1998; pp. 296–304.

40. Huang, W.; Li, S.; Liu, X.; Ban, Y. Predicting human mobility with activity changes. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 1569–1587. [[CrossRef](#)]
41. Gabrielli, L.; Rinzivillo, S.; Ronzano, F.; Villatoro, D. From tweets to semantic trajectories: Mining anomalous urban mobility patterns. In *Citizen in Sensor Networks*; Springer: Cham, Switzerland, 2014; pp. 26–35.
42. Steiger, E.; Westerholt, R.; Resch, B.; Zipf, A. Twitter as an indicator for whereabouts of people? Correlating twitter with uk census data. *Comput. Environ. Urban Syst.* **2015**, *54*, 255–265. [[CrossRef](#)]
43. Steiger, E.; Resch, B.; Zipf, A. Exploration of spatiotemporal and semantic clusters of twitter data using unsupervised neural networks. *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 1694–1716. [[CrossRef](#)]
44. Gao, S.; Janowicz, K.; Couclelis, H. Extracting urban functional regions from points of interest and human activities on location-based social networks. *Trans. GIS* **2017**, *21*, 446–467. [[CrossRef](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).