

# How Where Is When? On the Regional Variability and Resolution of Geosocial Temporal Signatures for Points Of Interest

Grant McKenzie<sup>1</sup>, Krzysztof Janowicz<sup>1</sup>, Song Gao<sup>1</sup>, Li Gong<sup>1,2</sup>

<sup>1</sup>*STKO Lab, Department of Geography, University of California, Santa Barbara, CA, USA*

<sup>2</sup>*Geosoft Lab, Institute of RS and GIS, Peking University, China*

---

## Abstract

The temporal characteristics of human behavior with respect to Points of Interest (POI) differ significantly with place *type*. Intuitively, we are more likely to visit a restaurant during typical lunch and dinner times than at midnight. Aggregating geosocial check-ins of millions of users to the place type level leads to powerful *temporal bands and signatures*. In previous work these signatures have been used to estimate the place being visited based purely on the check-in time, to label uncategorized places based on their individual signature’s similarity to a type signature, and to mine POI categories and their hierarchical structure from the bottom-up. However, not all hours of the day and days of the week are equally *indicative* of the place type, i.e., the information gain between temporal bands that jointly form a place type signature differs. To give a concrete example, places can be more easily categorized into weekend and weekday places than into Monday and Tuesday places. Nonetheless, research on the *regional variability* of temporal signatures is lacking. Intuitively, one would assume that certain types of places are more prone to regional differences with respect to the temporal check-in behavior than others. This variability will impact the predictive power of the signatures and reduce the number of POI types that can be distinguished. In this work, we address the regional variability hypothesis by trying to prove that all place types are created equal with respect to their temporal signatures, i.e., temporal check-in behavior does not change across space. We reject this hypothesis by comparing the inter-signature similarity of 321 place types in three major cities in the USA (Los Angeles, New York, and Chicago). Next, we identify a common core of least varying place types and compare it against signatures extracted from the city of Shanghai, China for cross-culture comparison. Finally, we discuss the impact of our findings on POI categorization and the reliability of temporal signatures for check-in behavior in general.

**Keywords:** Semantic Signature, Point Of Interest, Check-in, Geosocial, Time, Resolution

---

## 1. Introduction

Points of Interest (POI)<sup>1</sup> are inextricably linked to modern (mobile) search, recommender systems, location-based social networks, transportation studies, navigation and tourism systems, urban planning, predictive geo-analytics such as crime forecasting, and so forth. In terms of their computational representation, POI can be described and categorized in many different ways. Typical approaches are either based on features or functionality. The former describe POI based on attributes/properties such as price range, Wi-Fi availability, wheelchair access, ambience, noise level, room size, customer satisfaction, and so forth. Leaving pre-defined types such as *Restaurant*, *Hotel*, or *National park*, aside, POI can be grouped into ad-hoc categories [2] based on their common features such as “expensive places” or “attractions that offer wheelchair access.” A functionality-centric view describes and categorizes POI based on what they *afford*, e.g., dining, travel, trade, or shelter [14, 34]. While both approaches can be combined to account for their

---

<sup>1</sup>We use the term *Point of Interest* here to keep in line with related work in research and industry and because these places are typically represented by point geometries. On the long term and due to the increase in richer geometric representations, *Place of Interest* seems to be the more appropriate name.

distinct strengths and weaknesses, they are typically realized in a schema-first manner in which features or functionalities are defined top-down and then populated with data [11]. An example of such a schema is shown in Figure 1 which depicts properties defined for *museum* as well as the higher-level types from which these properties were inherited.

**Thing > Place > CivicStructure > Museum**  
 A museum.

Property	Expected Type	Description
<b>Properties from CivicStructure</b>		
<u>openingHours</u>	<u>Duration</u>	The opening hours for a business. Opening hours can be specified as a weekly time range, starting with days, then times per day. Multiple days can be listed with commas ',' separating each day. Day or time ranges are specified using a hyphen '-'. - Days are specified using the following two-letter combinations: Mo, Tu, We, Th, Fr, Sa, Su. - Times are specified using 24:00 time. For example, 3pm is specified as 15:00. - Here is an example: <code>&lt;time itemprop="openingHours" datetime="Tu,Th 16:00-20:00"&gt;Tuesdays and Thursdays 4-8pm&lt;/time&gt;</code> . - If a business is open 7 days a week, then it can be specified as <code>&lt;time itemprop="openingHours" datetime="Mo-Su"&gt;Monday through Sunday, all day&lt;/time&gt;</code> .
<b>Properties from Place</b>		
<u>address</u>	<u>PostalAddress</u>	Physical address of the item.
<u>aggregateRating</u>	<u>AggregateRating</u>	The overall rating, based on a collection of reviews or ratings, of the item.
<u>containedIn</u>	<u>Place</u>	The basic containment relation between places.
<u>event</u>	<u>Event</u>	Upcoming or past event associated with this place, organization, or action. Supersedes <u>events</u> .
<u>faxNumber</u>	<u>Text</u>	The fax number.
<u>geo</u>	<u>GeoCoordinates</u> or <u>GeoShape</u>	The geo coordinates of the place.

Figure 1: A fragment of the *Museum* type from Schema.org.

Alternatively, and assuming that meaning emerges from social structure [9], POI types can be described and categorized by aggregating how people behave towards places, e.g., when they visit them, what they say/write about them, and so forth. In addition to top-down schemata, such an approach reveals meaningful patterns suitable for a bottom-up, observations-first characterization of POI (types). To give a few concrete examples, certain types of places are visited mostly during the weekends, while others are visited primarily during the workweek. Similarly, some types have their visitation peaks during the evenings while others peak during typical business hours from 9am-5pm. Even the lack of such distinct peaks is indicative (e.g. of major airports). Textual descriptions and other sources of observations can be used accordingly. For instance, mining latent topics from social media such as textual user reviews of places from Los Angeles reveals very characteristic Spanish-language topics [23].

As an analogy to spectral signatures and bands in remote sensing, we have proposed *semantic signatures* that support the categorization of POI based on a multitude of spatial, temporal, and thematic bands [12]. Simply put, in the domain of remote sensing, geographic entities on the surface of the Earth are classified via their unique reflection and absorption patterns in different wavelengths of electromagnetic energy called *spectral bands* [28]. In some cases a particular band is sufficiently indicative to distinguish entity types (e.g., paved concrete from bare red brick), while in other cases a combination of multiple bands is required to form a unique spectral signature (e.g., deciduous and conifer trees cannot be distinguished via the visible light band alone).

Temporal signatures and bands are of particular interest as they are relatively easy to mine and at the same time are strongly indicative for a variety of POI types [29, 36]. Consequently, they have been successfully used for the labeling of uncategorized places, for data cleansing and deduplication, for the

construction of bottom-up POI hierarchies, for geolocation tasks such as estimating which place a user visited based on Global Positioning System (GPS) fixes, and further tasks that benefit from this kind of social sensing. Recognizing the role of time has also led to new fields of study such as time-aware POI recommendation [37]. Some POI types require additional (non-temporal) bands for their more fine-grained classifications [23]. However, we will exclusively focus on temporal signatures in this work.

Interestingly, not all hours of the day and days of the week are equally indicative for the classification of POI types, i.e., the information gain of temporal bands differs. Intuitively, places can be more easily categorized into evening and morning place types (e.g., *bars* versus *bakeries*) than into early morning and late afternoon places. To further exploit the analogy to spectral signatures, it is interesting to note that the *resolution* of temporal bands is characterized and bound by human behavior. While hourly, daily, and seasonal bands have predictive power, second or minute-based bands do not (at least not for POI). This leads to the question of whether temporal signatures also have a *platial*, i.e., place-based, resolution.<sup>2</sup> Note that we use the term *platial* (or *regional*) instead of *spatial* here as the variation is expected to be non-linear. For example, San Diego, CA and Tijuana, Mexico are neighboring cities, yet we expect them to vary more with regards to the temporal signatures (due to cultural differences between Mexico and the United States) than San Diego, CA and San Francisco, CA which are over 700km apart but within the same country. Conversely, non-spatial typically implies *platial* (regional) invariance.

Clearly, as temporal signatures are mined from human behavior, certain POI types will be affected by cultural differences. For instance, the peak dinner time for restaurants in Italy is around 8pm while it is approximately 6pm in the United States. We may even expect differences between the West and East Coasts of the U.S. In contrast, meaningful differences between the neighboring cities of New York, NY and Newark, NJ are less likely. Understanding such regional variations, their resolution, and magnitude, is important as they will effect the indicativeness of the signatures and thus their contribution to the aforementioned tasks [8]. In other words and referring back to the wordplay in the title, we will ask how much the *where*, i.e., regional-effects, impacts the *when*, i.e., the time people tend to visit certain types of venues. We will put this *platial* resolution research question to the test by hypothesizing that all place types are born equal with respect to their temporal signatures, i.e., that the temporal check-in behavior does not vary across space.

The remainder of the paper is structured as follows. Section 2 outlines our research contributions. Section 3 introduces the data and the temporal signatures mined from these data. Next, in Section 4, methods, results and discussion on *platial* variation are presented. Section 5 discusses a small subset of the results in further detail, while Section 6 compares these results to another dataset from Shanghai, China. We discuss related work in Section 7 and finish with a discussion of the overall results and the conclusions in Section 8.

## 2. Research contribution

The *regional variability hypothesis* can be illustrated using the following intuitive example. Given a user location derived from a positioning fix of a mobile device and a set of POI in the vicinity of this fix; can we match the user’s *spatial* location (lat/long coordinate) to a *platial* location (venue)? In other words, can we estimate which place a user visited, e.g., the Hollywood Palladium, based on the spatial location, e.g., the latitude and longitude of the GPS fix 34.0981,-118.3249. Intuitively, the probability of *checking-in* at a particular place is inversely proportional to the distance between the spatial footprint of the POI and the user’s location fix. As argued previously, check-in times can be aggregated to *type-indicative* temporal signatures. Now, given the example above, if the GPS fix was recorded at 8am, the user is more likely to be at the nearby Waffle eatery than the spatially closer Hollywood Palladium since the check-in probability for a concert venue is negligibly low in the morning. In contrast, the same fix recorded on Friday at 7pm most likely indicates a visit to the Palladium.

In fact, performing such an experiment with real data from over 2,800 check-ins in Los Angeles, CA shows that incorporating temporal signatures (aggregated from multiple cities) improves the geolocation estimation as measured by the Mean Reciprocal Rank (MRR)<sup>3</sup> from 0.359 to 0.453, i.e., by about **26%** [22].

<sup>2</sup>We use the term *platial* here in reference to *place*, similar to how *spatial* refers to *space*.

<sup>3</sup>MRR is a measure for evaluating any process that produces a list of possible responses ordered by probability of correctness.

Simplifying, the check-in probability for a given place depends on the distance of its spatial footprint to the user’s location as well as the temporal check-in likelihood at this *type* of place. Now, if the temporal signatures for POI types would show no regional variability, then the geolocation estimation quality would not differ based on the origin of those signatures. Temporal signatures derived from New York or Chicago check-in data would lead to the same increase in MRR over the distance-only baseline as signatures derived from Los Angeles (or signatures aggregated over multiple cities). However, performing such an experiment shows that using New York signatures for the geolocation estimation in Los Angeles leads to an MRR of **0.425**. This is lower than the performance of the aggregated signatures (**0.453**) but higher than the distance-only method commonly used to date (**0.359**). Consequently, while the New York signatures still outperform the baseline, there must be a *regional* variation in the check-in behavior. Alternatively, we can hypothesize that the signature differences are explainable by *random* variations.

**This raises several interesting research questions (RQ); three will be explored in the following sections:**

- RQ1** Are POI types regionally invariant and the observed differences described above due to random fluctuations? We will try to reject this null hypothesis using the circular Watson’s Two-Sample test.
- RQ2** Are regional signature variations equally strong across all POI types, i.e., are there types that are affected more or less by such variations? Furthermore, given a POI type hierarchy, do certain supertypes form around more or less varying subtypes? We will address these questions by comparing the inter-signature similarity of POI types from three major cities in the USA (Los Angeles, New York, and Chicago). To ensure that these similarities are not merely artifacts of the used measure, we will use the Gini Coefficient, Jensen-Shannon Divergence, and Earth Mover’s Distance, and study the concordance of the resulting similarities by computing Kendall’s W.
- RQ3** Given a common core of *least* varying POI types determined by their signatures from major US cities, how do these signatures hold up when compared against data from a very different cultural region, e.g., against signatures extracted for Shanghai, China? To approach this research question, we will select POI types that can be *aligned* between the U.S. and Chinese POI schemata and then divide them into two groups, those that vary clearly within the U.S. and those that do not. Next, we will use Earth Mover’s Distance to test whether these groups remain *stable* when using the Chinese signatures, i.e., whether POI types in the regionally invariant group remain in this group and vice versa.

### 3. Raw data and Temporal Signatures

This section discusses the used data and the creation of temporal signatures.

#### 3.1. Data

Check-in information was accessed hourly via the public *Foursquare* API to collect a total of 3,640,893 check-ins to 938,031 venues from 421 POI types across three regions: Los Angeles (LA), New York City (NYC), Chicago (CHI), and New Orleans (NOLA).<sup>4</sup> The Foursquare POI type schema groups these 421 POI types in to 9 top-level classes. To gain a better understanding of the data, Figure 2 shows the percentage of user check-ins divided by those 9 classes and split by region. *Travel & Transport* is by far the most prominent POI class followed by *Arts & Entertainment* which is more pronounced in Los Angeles than in either New York City or Chicago. In contrast, both New York City and Chicago show a higher percentage of check-ins at *Outdoors & Recreation* POI types.

For the purposes of this research, these check-in data were accessed during the fall/winter of 2013. The goal was to access check-ins to 60 venues in each city from each of the 421 POI types.<sup>5</sup> Given the limited number of venues of some POI types in the selected cities (e.g., Belarusian Restaurant), this was not always

---

<sup>4</sup>Region boundaries are based on the *2010 Census Urban Areas* boundaries.

<sup>5</sup><https://developer.foursquare.com/categorytree>

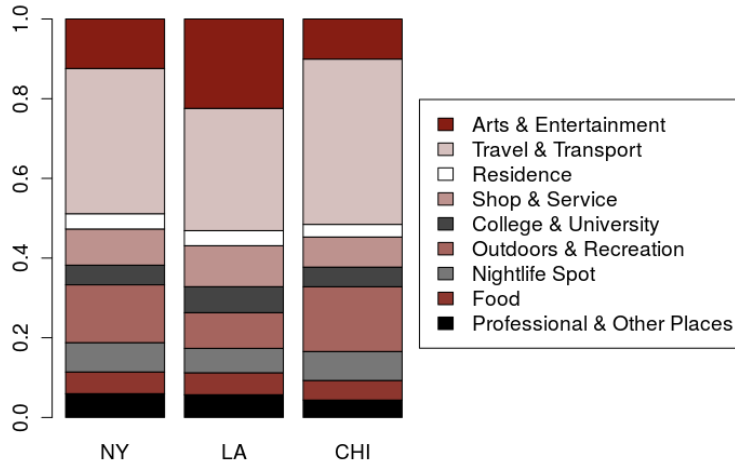


Figure 2: Stacked bar plots showing amount of check-ins to each parent class as a percentage of overall check-ins. Check-ins have been split in to regions.

possible. The hourly check-in data were aggregated by POI type, region, hour, and day of the week. Given 24 hours over 7 days, this resulted in 168 hourly *bands* used to construct a temporal signature normalized and aggregated to a single week. In order to ensure the robustness of the temporal signatures, any type whose venues appeared less than 30 times in a given region was removed from analysis. This reduced the number of POI types from 421 to 321. Additionally, the New Orleans dataset was dropped from analysis due to the limited availability of certain types which would have considerably restricted the categories available for comparison. The remaining 321 POI types in the three regions, Los Angeles, New York City and Chicago form the basis of the analysis to be discussed in the remainder of the paper. Lastly, the signatures were cleaned by removing data errors and outliers. Note that due to the usage restrictions of the Foursquare API, no individual check-ins or venues were stored for this research but merely type-level aggregates.

### 3.2. Temporal Resolution

Before using these signatures, it is important to understand their *temporal resolution* [13], in this case, the smallest change in quantity (i.e., check-ins counts) that can be observed via a sensor (i.e., the check-in *Apps* and *Application Programming Interface (API)*). Reporting and using data below such a resolution may lead to erroneous results. For this reason, it is common practice to round data to their significant digits. While some location-based social network (LBSN) APIs return the check-in timestamp, others return check-in counts per venue and have to be scanned repeatedly at an interval that corresponds to the temporal resolution. More importantly, timestamps do not represent the time a user entered a place. For instance, a user would most likely enter a coffee shop, order an espresso, sit down, and *then* use his/her smartphone to check-in. In fact, this process and the behavior associated with *checking in* have been the subject of multiple studies [5, 7]. Other systems, however, may check-in users automatically; see [21] for the resulting differences between manual and automatic check-ins. Additionally, most LBSN platforms do not provide an option for *checking out* of a place and therefore, many services will typically check their users out automatically after a certain time, e.g., 2 hours. Consequently, reporting temporal signatures on the level of minutes (even for large aggregates of data) or trying to draw conclusions from check-out times invites misunderstanding (Figure 3c). This is of particular importance for the research at hand as we will compare signatures aggregated via Foursquare with those from *Jiepan*<sup>6</sup>, a leading Chinese LBSN services whose APIs return different temporal resolutions of data.

Consequently, we report the data at an hour-resolution as depicted in Figure 3a. If appropriate and necessary, the signatures can be smoothed via a kernel function; see Figure 3b.

<sup>6</sup>See Section 5 for further details on Jiepan

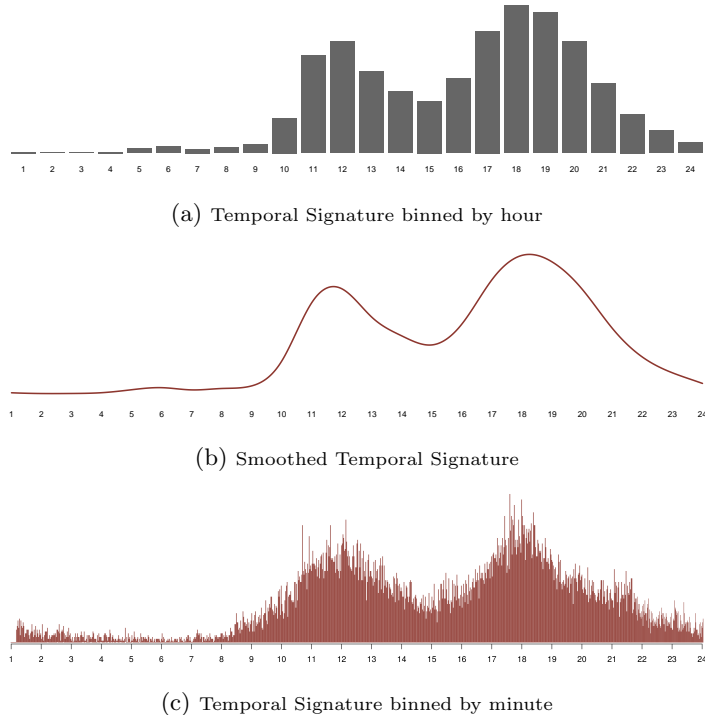


Figure 3: Check-in data represented as (a) a binned temporal signature by hour of the day and (b) a smoothed temporal signature. Both temporal signatures show averaged check-in behavior over 24 hours (a typical Tuesday) at a Mexican Restaurant.

## 4. Regional variation

In this section, a number of methods for analyzing regional variations between POI types are presented. First, the question of whether or not types are place specific is examined followed by an analysis of how much individual POI types vary regionally. Finally, POI hierarchies are examined in terms of their temporal signature homogeneity. We will define the terms and introduce the used measures in the corresponding subsection.

### 4.1. Significance of platial variations

Before we can explore the regional differences between particular POI type, we have to exclude the possibility that the temporal signature variations are merely a sampling artifact or produced through random fluctuations (**RQ1**). In order to do so, we start with the hypothesis that all types of POI are regionally equal, in other words they are platially invariant. Using *Watson's Non-parametric Two Sample  $U^2$  Test Of Homogeneity* [33, 39] we can test this hypothesis. The *Watson's  $U^2$  test* (Equation 1) starts with the assumption that all samples (temporal signatures) are drawn from the same population (region). The variable  $N$  is the sum of the number of values in each sample ( $n_1, n_2$ ) and  $d_k$  is the difference between the two cumulative signatures.

$$U^2 = \frac{n_1 n_2}{N^2} \left[ \sum d_k^2 - \frac{(\sum d_k)^2}{N} \right] \quad (1)$$

The test also assumes that the temporal signatures are circular in nature (e.g, Monday is equally as close to Sunday in temporal distance as Sunday is to Saturday). Figure 4 visually depicts circular representations of temporal signatures for the POI types of *Theme Park* and *Drugstore*. Clearly, temporal signatures for *Theme Park* tend to vary stronger with place than those for *Drugstore*.

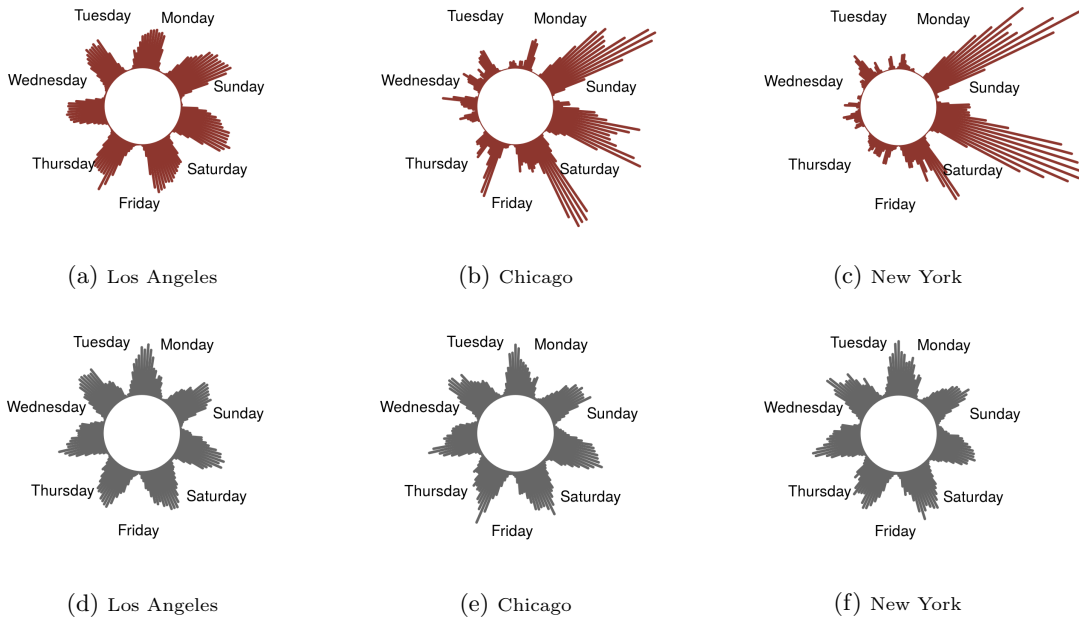


Figure 4: Circular histograms depicting temporal signatures for *Theme Park* (a,b,c) and *Drugstore* (d,e,f).

Altering the significance level<sup>7</sup>, the categorical circular distributions of 168 temporal bands are compared between each pair of regions (e.g., NYC & LA). The results shown in Table 1 present the percentage of POI types that are significantly different between regions based on the provided significance level. For example, a significance value of  $0.05$  shows that in **52%** of the cases, the hypothesis is rejected for the pair of LA & NYC meaning that 167 out of 321 POI types differ significantly between the two regions. Similarly, LA & CHI and NYC & CHI pairs reject the hypothesis for **50%** and **48%** of POI types respectively. Provided this information, arguments can be made that (1) measurable and meaningful regional variability does exist between POI types and (2) some types are regionally dependent while others are not.

	<b>0.01</b>	<b>0.05</b>	<b>0.1</b>
<b>NYC &amp; CHI</b>	33%	48%	57%
<b>LA &amp; CHI</b>	37%	50%	59%
<b>LA &amp; NYC</b>	36%	52%	63%

Table 1: Percentage of POI types that are statistically different between regions as determined by the *Watson's* non-parametric two sample  $U^2$  test of homogeneity. The results for three significance values (0.01, 0.05, 0.1) are reported.

These results confirm our intuition and reject the null hypothesis (**RQ1**). On the one hand, temporal signatures for POI types or check-in times in general have been successfully used in previous research because they are stable and generalizable over individual samples [8, 23, 29, 36, 37, 38]. On the other hand, even when applying the very conservative 0.01 alpha level, at least 106 POI types differ significantly between regions. Thus, understanding and quantifying these differences opens up new ways to substantially improve POI recommendation, classification, and so forth.

<sup>7</sup> *alpha parameter* in the `watson.two.test` function in R.

The question remains as to *which* POI types are spatially variant and by what amount? This will be answered in the following subsections.

#### 4.2. Variability between categories

In considering **RQ2** it is necessary to explore how temporal signatures of different POI types change based on region. In order to determine the amount by which some POI types are regionally dependent, we analyzed the variability using three dissimilarity measures.

##### 4.2.1. Difference in Gini coefficients

The *Gini coefficient* is a measure of the inequality of a given distribution. Originally intended to represent the income distribution of a country’s residents [10], a distribution of  $P$  is said to be equal (all values are the same) if  $G(P)$  results in 0 and completely unequal should  $G(P)$  equate to 1. As shown in Equation 2, this coefficient provides a rough value used to describe any given distribution where  $x$  is an observed value,  $n$  is the number of values and  $\mu$  is the mean value. In comparing two distributions, the *Gini coefficient* of one distribution can be subtracted from the other (which we refer to as the *difference in Gini Coefficients* or DGC) to give a broad indication of the (dis)similarity of two distributions.

$$G(P) = \frac{\sum_{i=1}^n \sum_{j=1}^n |x_i - x_j|}{2n^2\mu} \quad (2)$$

Table 2 lists the five most dissimilar POI types as well as the five most similar types based on the difference in Gini coefficients. The types are split based on region pairs. The value shown in parenthesis beside each type is the *difference in Gini coefficient* value normalized by the most dissimilar type (Theme Park) and the most similar type (American Restaurant). Normalization allows for comparison between POI type as well as between dissimilarity measures (cf. Table 3 and Table 4).

NYC & LA	NY & CHI	LA & CHI
Dissimilar POI Types		
Theme Park (0.844)	Recycling Facility (0.825)	Theme Park (1)
Real Estate Office (0.739)	Resort (0.797)	Resort (0.802)
E. European Restaurant (0.68)	Farm (0.703)	Baseball Stad. (0.74)
Recycling Facility (0.586)	Historic Site (0.702)	Donut Shop (0.711)
Farm (0.582)	Basketball Stad. (0.686)	Garden Cntr (0.704)
Similar POI Types		
Drugstore / Pharmacy (0.004)	Furniture / Home Store (0.003)	Monument / Landmark (0.005)
Gym (0.001)	Harbor / Marina (0.002)	Men’s Store (0.003)
Community College (0.001)	Yoga Studio (0.002)	Gym (0.002)
Pet Store (0.001)	Laboratory (0.001)	Community College (0.001)
Art Museum (0.001)	Wings Joint (0.001)	American Restaurant (0.000)

Table 2: Top five and bottom five dissimilar POI types based on normalized difference in Gini coefficient and split by region pairs.

##### 4.2.2. Jensen-Shannon Distance

While informative, the *difference in Gini coefficient* approach primarily focuses on the minima and maxima of a distribution. The *Jensen-Shannon Divergence (JSD)* is a method for measuring dissimilarity between two probability distributions ( $P, Q$ ) [19]. In this case, comparison between distributions is done through a one-to-one bin approach. The *distance metric* is calculated by taking the square root of the value resulting from the divergence and is bounded between 0 (identical distributions) and 1 (complete



dissimilarity). The  $JSD$  calculation is shown in Equation 3 where  $M = \frac{1}{2}(P + Q)$  and KLD represents the *Kullback-Leibler Divergence* specified in Equation 4. While useful as a dissimilarity metric,  $JSD$ 's one-to-one bin comparison does not take into account neighboring bins.

$$JSD(P \parallel Q) = \frac{1}{2}KLD(P \parallel M) + \frac{1}{2}KLD(Q \parallel M) \quad (3)$$

$$KLD(P \parallel Q) = \sum_i P(i) \ln \frac{P(i)}{Q(i)} \quad (4)$$

Table 3 shows the top five and bottom five dissimilar POI types split by region pair. As we saw with the *difference in Gini coefficient* approach (Table 2), the most dissimilar POI types are often *Theme Parks* or *Stadiums*. Interestingly, the most similar POI types are shown to be a variety of *Stores* (e.g., Grocery Store).

NYC & LA	NYC & CHI	CHI & LA
Dissimilar POI Types		
Football Stadium (1.000)	Theme Park (0.863)	Football Stadium (0.843)
Baseball Stadium (0.687)	Recycling Facility (0.677)	Theme Park (0.835)
Theme Park (0.603)	Food Truck (0.651)	Recycling Facility (0.733)
Basketball Stadium (0.594)	Funeral Home (0.627)	Skate Park (0.710)
Campground (0.584)	Basketball Stadium (0.586)	Food Truck (0.707)
Similar POI Types		
Electronics Store (0.021)	Grocery Store (0.000)	University (0.021)
Furniture / Home Store (0.039)	Residential Building (0.037)	Electronics Store (0.035)
Hospital (0.035)	Home (private) (0.023)	Hardware Store (0.030)
Grocery Store (0.032)	Department Store (0.021)	Drugstore / Pharmacy (0.024)
Department Store (0.031)	Mall (0.018)	Gym (0.022)

Table 3: Top five and bottom five dissimilar POI types based on normalized Jensen-Shannon Distance and split by region pairs.

#### 4.2.3. Earth Mover's Distance

Given  $JSD$ 's reliance on a one-to-one bin comparison, the *Earth Mover's Distance (EMD)* is utilized as well. Originally introduced by the computer vision community [26, 27],  $EMD$  compares each bin in a distribution ( $P$ ) to all bins in a second distribution ( $Q$ ) assigning a cost value based on bin distance. Simply put,  $EMD$  is the minimum amount of work it takes to convert one distribution into the other, where  $F_{i,j}$  is a flow matrix (amount of earth to move between bins) and  $C_{i,j}$  is the cost matrix representing the cost of moving the flow. The total cost is then shown in Equation 5.

$$EMD(P, Q) = \sum_{i=1}^n \sum_{j=1}^n F_{i,j} C_{i,j} \quad (5)$$

As with both  $DGC$  and  $JSD$ , calculating the  $EMD$  across all types for all pairs of regions allows us to rank POI types by their regional similarity with high values indicating high dissimilarity. Table 4 lists the five most and five least dissimilar types split by region. The normalized  $EMD$  values are shown in parenthesis next to the type name. Similarities between the regional pairs are apparent in both the highly dissimilar and similar (shaded gray) groups with *Theme Parks* and *Stadiums* again, showing to be the most dissimilar POI type and *Stores* and *Residences* proving to be the most similar.

NYC & LA	NYC & CHI	LA & CHI
Dissimilar POI Types		
Theme Park (0.789)	Theme Park (0.710)	Theme Park (1.000)
Football Stad. (0.686)	Resort (0.614)	Resort (0.600)
Real Estate Office (0.471)	Basketball Stad. (0.549)	Baseball Stad. (0.575)
East Euro Restaurant (0.416)	Winery (0.506)	Garden Center (0.442)
Farm (0.402)	Recycling Facility (0.453)	Donut Shop (0.423)
Similar POI Types		
College Residence Hall (0.013)	Home (0.005)	Monument / Landmark (0.015)
Shoe Store (0.011)	Hardware Store (0.005)	University (0.009)
Military Base (0.010)	Doctor’s Office 0.003	Drugstore/ Pharmacy (0.006)
Convenience Store (0.001)	Comm. College (0.002)	Home (0.004)
Drugstore/ Pharmacy (0.000)	Airport Gate (0.000)	Convenience Store (0.002)

Table 4: Top five and bottom five dissimilar POI types based on normalized Earth Mover’s Distance and split by region pairs.

Summing up, with respect to **RQ2**, these three dissimilarity measures show that there are clear differences between POI types. Some, e.g., *Theme Park*, show a strong regional variability, while others, e.g., *Convenience Store*, do not.

#### 4.3. Concordance between dissimilarity measures

While these three statistical dissimilarity measures yield individual results for inter-signature comparison, the real value of these measures is shown in their agreement. Here *Kendall’s coefficient of concordance* is employed [16]. Each of the three regions is compared to each other region using *Earth Mover’s Distance*, *Jensen-Shannon distance* and *difference in Gini coefficient*. These produce a single dissimilarity value from each region pair for each POI type. Kendall’s *W* is then used to calculate the measure of concordance between each dissimilarity measure across all POI types.

Measures	NYC & LA	LA & CHI	NYC & CHI
EMD & JSD	0.80	0.82	0.78
EMD & GINI	0.91	0.91	0.88
GINI & JSD	0.74	0.73	0.74

Table 5: Kendall’s coefficients of concordance *W* for pairs of regions and combinations of dissimilarity measures ( $p < 0.01$  in all cases).

A *Kendall’s W* value of 1 indicates complete concordance where a value of 0 represents no concordance at all. As shown in Table 5, all *W* values are greater than random with the values for *EMD & GINI* producing the highest coefficient of concordance followed closely by *EMD & JSD* and *GINI & JSD*. This indicates a high level of agreement between dissimilarity measures, thus excluding the possibility that the observed similarities are merely artifacts of choosing a specific measure. We will focus on EMD for the remaining analysis.

#### 4.4. Hierarchy homogeneity

Typically, POI types are not flat but form a hierarchy consisting of one or more root types followed by multiple type-levels. Figure 1 shows such a hierarchy from schema.org with *Thing* as the root type. The subsumption relation is transitive, i.e., as *Place* is a supertype of *CivicStructure* and *CivicStructure* is a

supertype of *Museum*, *Museum* is also a subtype of *Place*. Such hierarchies are not only important means for knowledge engineering but also key for various information retrieval techniques such as query expansion.

The second part of **RQ2** poses the interesting question of whether superotypes, e.g., *Retail*, in a POI hierarchy are homogeneous with respect to the temporal signature variability of their subtypes, e.g., *Hardware store*. To address this question, we grouped the top 100 most and top 100 least varying subtypes and then compared their distribution with respect to the superotypes. Intuitively, homogeneous superotypes should mainly contain subtypes from one group but not from both.

Figure 5a depicts the results of our analysis for the superotypes provided by Foursquare. By necessity, hierarchies introduce some arbitrariness by highlighting certain perspectives and hiding others. The Foursquare POI hierarchy is an interesting case as its superotypes seem like mixed bags, e.g., grouping *Cemeteries* under the *Outdoors & Recreation* root type and even introducing a *Professional & Other Places* “catch-all” type. While some POI types such as *Nightlife Spot* and *Travel & Transport* are homogeneous, the majority do not show a clear trend. This confirms our intuition. In fact, this very problem has been addressed before, combining spatial, thematic, and temporal signatures to construct a more appropriate POI type hierarchy for Foursquare from the bottom-up [23]. We can now use this hierarchy to compare it to the original Foursquare categorization. Intuitively, the bottom-up version should be more homogeneous, i.e., superotypes predominantly contain either similar or dissimilar subtypes (with regards to their temporal signatures between U.S. cities). Figure 5b confirms this assumption, the *Accommodation, Eating & Drinking* and *Attractions* types being particularly clear examples. It is interesting to note that in both hierarchies the transportation-centric types contain mostly similar POI types, while the service types consist of subtypes too diverse to show a clear picture.

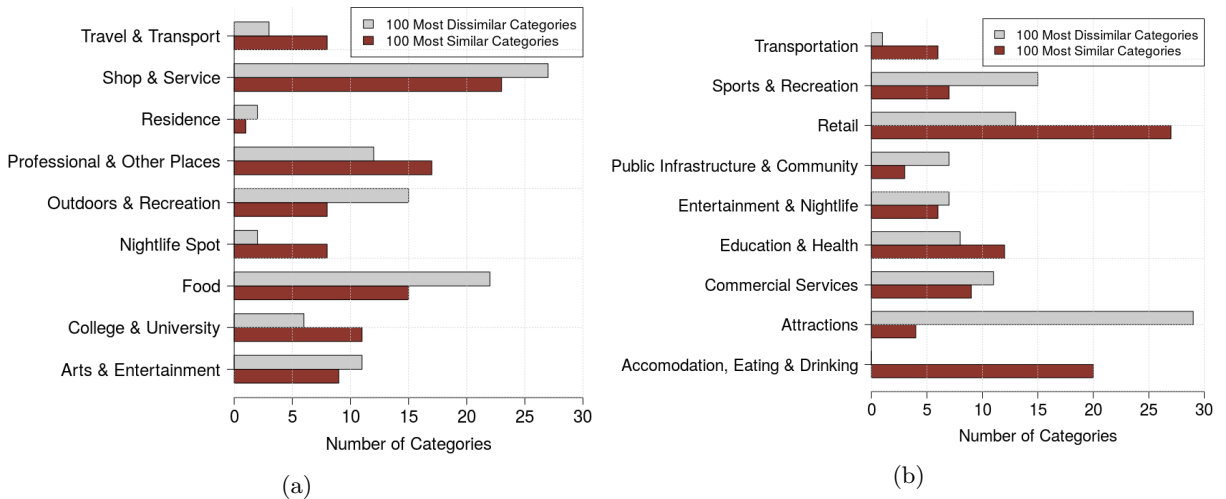


Figure 5: (a) Original Foursquare POI hierarchy superotypes by prevalence of the 100 most similar subtypes and the 100 most dissimilar subtypes. (b) Bottom-up signature-based POI hierarchy superotypes by prevalence of the 100 most similar subtypes and the 100 most dissimilar subtypes

Summing up, to answer the second part of **RQ2**, POI hierarchies are not generally homogeneous with respect to the regional variability of the temporal signatures of their types. Nonetheless, some superotypes show clear patterns even across different POI hierarchies.

## 5. Cross-cultural comparison

The next step in examining regional variation in POI types is a preliminary investigation into how cultural differences influence spatial variation in temporal signatures. With respect to research question **RQ3**, we examine the temporal signatures of POI in a city outside of the United States. In filling this role we chose to compare U.S. temporal signatures based on Foursquare data with temporal signatures constructed from

*Jiepan*<sup>8</sup> check-ins, which is one of the largest location-based social network platforms in China. Studying cross-cultural differences is important for a multitude of reasons including semantic translation and the alignment of place type hierarchies as well as understanding different motivations for contributing data and participating in online social networks[17].

### 5.1. Chinese check-in dataset

The *Jiepan* dataset on which this section is based contains more than 20 million location-based social check-in records from a one-year period starting in September of 2011. All check-in data is from the Shanghai region of China and has been used to explore inter-urban mobility in previous work [20, 18]. Approximately 75,000 venues from 156 POI types grouped within 8 root-level types were extracted from user check-ins. Note that the predefined classification system of the Chinese check-in dataset is different from the type schema used in Foursquare. For example, the (*American*) *football stadium* type is popular in the United States while it does not exist in the *Jiepan* POI hierarchy. Furthermore, *Doctor’s Office*, *Hospital* and *Medical Center* types from Foursquare are merged into a single Chinese LBSN POI type. Interestingly, and confirming our results from Section 4, the POI types that show clear regional differences within the U.S. are among those that are most difficult to align to the Chinese dataset, e.g., *Theme Park*, various types of sports facilities, and *Donut Shop*. In contrast, it was easier to find corresponding *Jiepan* POI types for the top *regionally invariant* types. As a sample comparison, 10 highly regionally invariant and 10 highly varying POI types were selected and manually matched between both datasets (see Table 6).

### 5.2. POI type similarity comparisons

The 20 sample POI types were selected based on dissimilarity analysis within the three U.S. cities presented in the previous sections. Similar to the methodology discussed in Section 4.2, we applied the Earth Mover’s Distance to calculate the dissimilarity of these POI types between the averaged temporal signatures for U.S. cities and the city of Shanghai.

Table 6 lists the 20 most and least dissimilar POI types along with the normalized EMD values for within the United States and between the United States and Shanghai. Please note that the *Mean EMD Within U.S.* is calculated by taking the mean of the EMDs reported from each regional pair while the *Shanghai vs. U.S. Mean EMD* is calculated as the EMD between the *regionally averaged* U.S. temporal signature and the *Jiepan* temporal signature. While the average EMD of temporal signatures between *the Shanghai vs. U.S. Mean* for all POI types is higher than that of *Within U.S.*, the magnitude difference between highly dissimilar POI types and highly similar POI types remains the same across cultures. In other words, POI types that are highly variable in the U.S. are also highly variable in Shanghai China (means of **0.62** & **0.68** respectively) while the most stable POI types remain stable across cultures (means of **0.07** and **0.23** respectively). The *Spearman* correlation coefficient of these sets of normalized EMD values is **0.64** (*Pearson* = 0.70) indicating above average similarities between the two.

In response to **RQ3**, this section shows that the most regionally invariant types in the U.S. show reasonable stability when compared to Shanghai, China, but that highly variable types within the U.S. are also high variable in the Chinese dataset. This is a very valuable insight as it indicates that some POI type may be represented by signatures with potentially global coverage.

## 6. Investigation of Temporal Signature differences through examples

The analysis presented in the previous sections shows that POI types do in fact vary regionally with some showing significant changes between the regions of Los Angeles, New York City, Chicago, and Shanghai, and others displaying no significant difference in their temporal signatures. In this section, we show a few select examples of these POI types with the purpose of illustrating why regional variability exists for some types but not others.

---

<sup>8</sup><http://jiepan.com>

POI Type	Shanghai vs. U.S. Mean $nEMD$	Mean $nEMD$ Within U.S.
Dissimilar POI Types		
Theme Park	0.89	1.00
Farm	0.89	0.82
Historic Site	0.43	0.69
Zoo	0.42	0.59
Cemetery	1.00	0.58
Gaming Cafe	0.63	0.58
Pool Hall	0.25	0.54
Burger Joint	0.89	0.53
Gas Station/Garage	0.42	0.46
Public Art	0.98	0.36
Similar POI Types		
Toy/Game Store	0.00	0.15
Furniture/Home Store	0.45	0.13
College Library	0.05	0.13
Shoe Store	0.28	0.11
Mall	0.19	0.10
Grocery Store	0.09	0.04
Hotel	0.37	0.01
University	0.35	0.01
Home (private)	0.10	0.00
Drugstore / Pharmacy	0.23	0.00

Table 6: Ten highly dissimilar POI types and ten highly similar POI types selected from the U.S. Foursquare dataset. The Earth Mover’s Distance was calculated between each Foursquare POI type its Chinese *Jiepan* counterpart. The values were normalized between the most dissimilar (1) and most similar (0) POI type.

The POI type that shows the highest level of dissimilarity across all pairs of regions and all dissimilarity measures is *Theme Park*. While this POI type may not immediately come to mind when thinking about regional differences, the reason is apparent when one examines the discretized temporal signatures (168 hourly bands of the week) shown in Figure 6. Check-in probabilities remain quite constant throughout the week for Los Angeles, while weekend peaks are much more pronounced for both New York and Chicago. Thinking of theme parks in the Los Angeles region, the *Disneyland Resort and Adventure Park* often come to mind. In 2013, the park hosted approximately 16.2 million guests making it the third most visited park in the world that year [31]. Provided this knowledge, it is reasonable to assume that the temporal signature for *Theme Park* in Los Angeles is dominated by check-ins to Disneyland. It is not surprising then, given this “holiday destination,” that the temporal signature displays very little difference between weekend and weekdays. To test the impact of Disneyland on the overall LA Theme Park signature, we used the Watson’s Non-parametric Two Sample  $U^2$  Test Of Homogeneity (described in Section 4.1). The hypothesis that there is no statistical difference between the temporal signatures for Los Angeles Theme Parks both with and without *Disneyland* was tested. The results indicate no significant difference between the distributions at a p-value of 0.01. This implies that while Disneyland is one of the most famous theme parks in the world, it exhibits similar temporal check-in patterns as other theme parks in Los Angeles, e.g., Universal Studios Hollywood.

There are many potential reasons why people behave differently towards theme parks in Los Angeles compared to the other cities. A strong argument can be made for the impact of weather on theme park

visits as well as seasonal effects in general. Weather variability in Southern California is minimal relative to the seasonal variability experienced in both Chicago and New York. In actuality, many theme parks in New York and Chicago close completely for the winter months (November - March) and a limited few remain open on the weekends for special events. Interestingly, check-in data from Shanghai shows similar weekend behavior but additionally we see a tendency toward a peak in the morning during the weekdays. In future research we plan to study seasonal temporal bands to capture these details. It is worth noting that we do not control for weather or any other environmental, cultural, or political factors here as we are interested in regional differences. Certainly, if one could remove all factors that differentiate Los Angeles from New York and Chicago, there would be no reason for people to behave differently towards POI types.

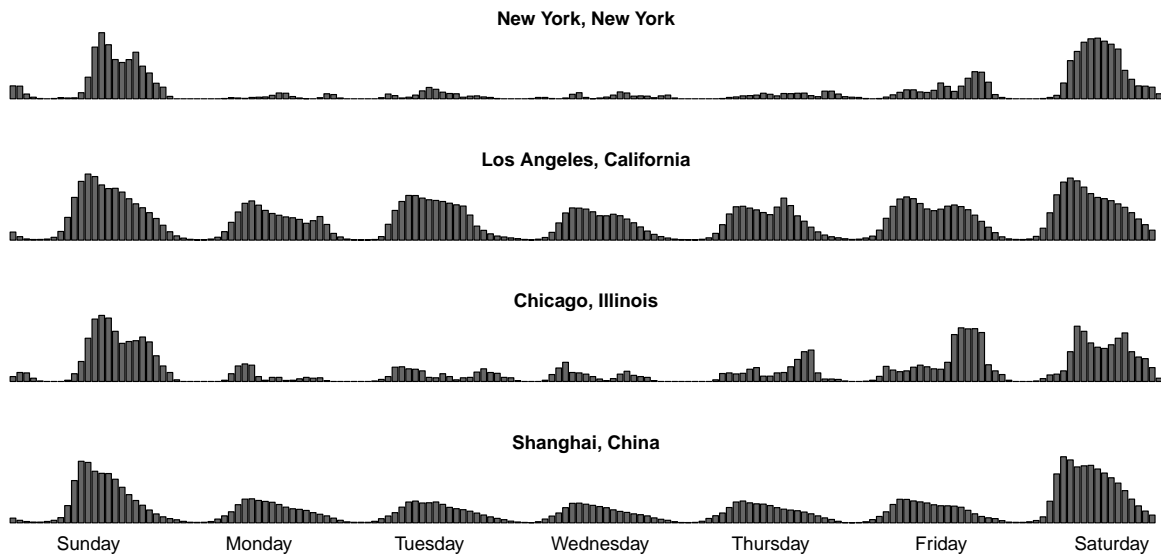


Figure 6: Temporal Signatures for the POI type *Theme Park* in New York City, Los Angeles and Chicago, United States and Shanghai, China.

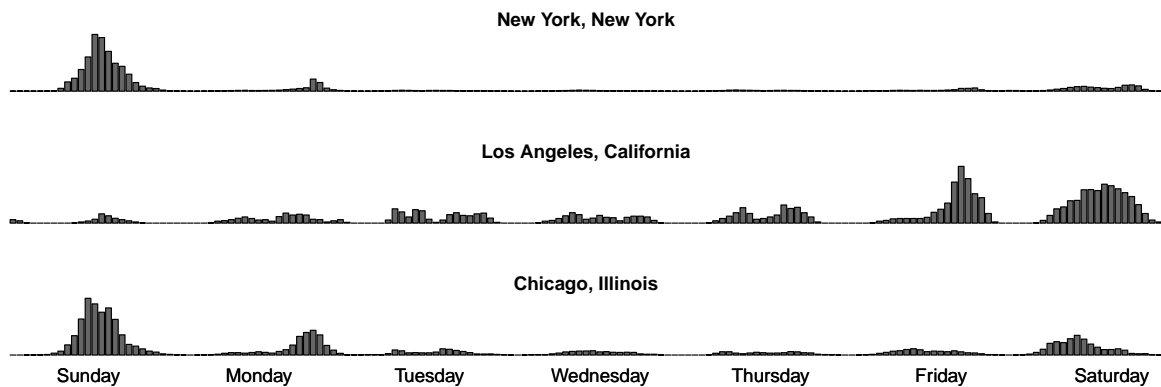


Figure 7: Temporal Signatures for the POI type *Football Stadium* in three cities in United States. Note that data from Shanghai China is not shown here as no matching POI type was found.

Based on the variability analysis done in Section 4, *Football Stadium* are shown to be another POI type high in dissimilarity between regions. Since professional American football is traditionally played on Sundays, one might expect temporal signatures to be quite similar between regions in the United States. Upon further examination we find a number of different factors contributing to this dissimilarity ranking.

First, while *professional* football is played on Sundays, College football is often played on Saturdays and High School football is typically played on Friday nights. It is important to know that Los Angeles does not have a professional football team which means that the peak one would expect on Sunday afternoon (which is seen in Chicago and New York City) is not found in the Los Angeles signature (Figure 7). Instead, we see the influence of both College and High School Football with peaks on Friday night and Saturday afternoon. Furthermore, football stadiums as with other types of stadiums, routinely host events other than just football matches. Major music concerts, trade fairs, and other sporting events often take place in large football stadiums which would also contribute to the regional difference in temporal signatures.

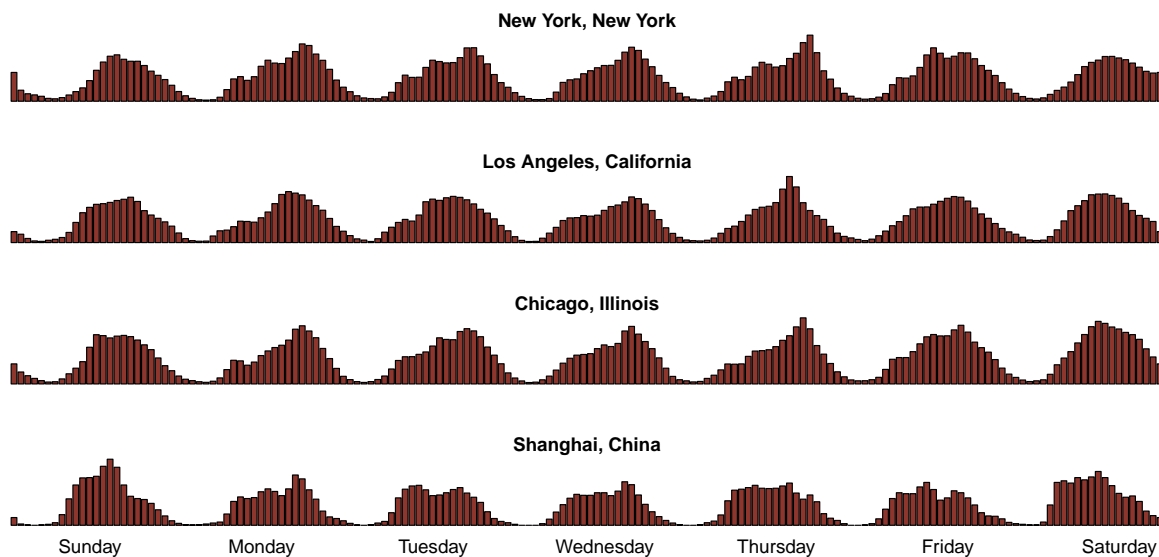


Figure 8: Temporal Signatures for the POI type *Drug Store / Pharmacy* in New York City, Los Angeles and Chicago, United States and Shanghai, China.

Lastly, the category of *Drug Store / Pharmacy* presents the highest regional similarity across all POI types. From a conceptual perspective, one often thinks of a drug store or pharmacy as being an *atemporal* type of category. One would be hard-pressed to list the *typical* times of day that an individual would choose to visit a drug store as drug stores offer a wide range of products. Found on many street corners in the United States and China, drug stores are often the “closest” place to pick up anything from sunscreen to birthday cards. Figure 8 shows the *atemporal* nature of drug stores with check-in values shown for most of the daylight hours and less check-ins late at night when many drug stores may be closed. Furthermore, Figure 8 shows the lack of regional variability in the temporal signatures between Los Angeles, New York City, Chicago and Shanghai.

In summary, while statistical methods applied to temporal signatures show that there are regional differences in POI types, a better understanding of the data behind these variations can be gained through a detailed examination of a subset of POI types.

## 7. Related work

Previous studies have explored the role of LBSN data in analyzing human behavior and urban dynamics. For instance, Cheng et al.[4] found that users follow the “Lévy Flight” mobility pattern and adopt periodic behaviors in check-ins, which were bounded with their social ties as well as geographic and economic constraints. Wu et al.[35] further analyzed the temporal transition probability of different activities (e.g., working, dining and entertainment) using social media check-in data. In work by Noulas et al. [24], the authors reveal different temporal rhythms in the top 10 most popular Foursquare categories (e.g., Home,

Café, Highway, and Bar) between weekdays and weekends. The distinct temporal bands of POI types can be useful for data cleaning, place recommendation and decision making in LBSN [36]. From the urban informatics perspective, the POI data mined from user-generated content provides a fresh and updated view on the *city-in-use* versus the *city-in-plan*. Thus, it can help study neighborhood variations and monitor land-use changes [25]. Cranshaw et al. [6] spatially clustered POI as urban neighborhoods and studied how multiple factors shape urban dynamics. Recently, McKenzie et al. [23] introduced a multi-granular, semantic signatures-based approach for the interactive visualization of the city pulse using millions of POI data in the Greater Los Angeles area. A data-driven and theory-informed POI classification approach has also been introduced in this work focusing on the multi-dimensional (spatial, temporal and thematic) characteristics of POI types. A large amount of research has relied on spatial (as opposed to platial) social media data such as *tweets* for answering domain-specific questions (e.g., [3, 15]). Although there is a large volume of literature studying POI location recommendation based on users' historical check-in records and spatio-temporal patterns [1, 32, 40, 8, 37], to the best of our knowledge, no existing research has addressed the *platial perspective* and the role of regional variability in categorically defined temporal signatures.

## 8. Conclusions and future work

In this work we have discussed the regional variability and resolution of temporal signatures for Points of Interest. To study the variability, we assumed that POI type signatures are regionally invariant and hypothesized that the observed differences are merely random fluctuation. We rejected this hypothesis using Watson's Two-Sample test. Consequently, there are measurable and meaningful regional differences between POI types. This is an important finding as temporal signatures are a valuable social sensing methodology for various tasks including data cleansing, geolocation, POI recommendation, and categorization.

Next, we discussed the magnitude and the distribution of these differences within the U.S. by comparing major cities. To ensure that the comparison is not driven by the choice of similarity measure, we tested three measures and determined the concordance between them. The results confirm that the regional temporal signature variation is not homogeneous across POI types. A POI type that does not show regional variations when comparing New York to Los Angeles data, is also likely to show no substantial variation when comparing any of these cities to Chicago. Interestingly, the picture is more difficult for types that display strong regional variability. These types differ in unique ways, so to speak, independent of the compared cities. Finally, we compared U.S.-based signatures to those from Shanghai, China to test whether types that show less variance would also remain stable when compared to data from a very different culture. While additional research is necessary, our first results indicate that this is the case.

Summing up, temporal signatures built from social media data (here user check-ins) show clear regional differences. They vary to a degree where methods and applications would benefit from region-specific signatures. However, this does not mean that one would have to generate and store a multitude of local signatures. First, as the geolocation example in Section 2 demonstrates, *aggregated* signatures are very powerful and second, not all types vary to a degree that would justify the additional overhead. The suitable platial resolution for regionally varying signatures depends on the concrete application needs and the expected benefits. Defining country-wide signatures may be an appropriate resolution for some tasks but not for others. However, using the same signatures world-wide will only prove useful for a certain subset of relatively invariant types. Once again, this highlights the local nature of information and the role of space and place in studying Physical-Cyber-Social relations in general [30]. Our findings are important as today's research applies temporal POI and check-in data uniformly across space. In addition, temporal signatures from different regions could potentially aid urban planners and policy makers in improving place-type based services according to local characteristics.

Future work in this area will involve expanding the dataset to include additional regions from major cities around the world. We will also explore the difference between rural and urban settings as well as the influence of weather and seasonal effects on certain types of POI. Along the same lines, we focused on regional differences here while demographic variation, urban planning and environmental conditions may also be key drivers.



Finally, the work at hand is part of a long-term project [12] to publish an openly available library of semantic signatures with the hope that it will be equally as transformative as *spectral signature libraries* have been to the field of remote sensing. Signatures are difficult and time consuming to mine; the research community will benefit from having common access to well described and documented spatial, temporal, and thematic signatures for Points Of Interest and other features.

## References

- [1] Bao, J., Zheng, Y., Mokbel, M. F., 2012. Location-based and preference-aware recommendation using sparse geo-social networking data. In: Proceedings of the 20th International Conference on Advances in Geographic Information Systems. ACM, pp. 199–208.
- [2] Barsalou, L. W., 1983. Ad hoc categories. *Memory & cognition* 11 (3), 211–227.
- [3] Chen, X., Yang, X., 2014. Does food environment influence food choices? A geographical analysis through “tweets”. *Applied Geography* 51, 82–89.
- [4] Cheng, Z., Caverlee, J., Lee, K., Sui, D. Z., 2011. Exploring millions of footprints in location sharing services. In: Proceedings of the International Conference on Weblogs and Social Media (ICWSM). pp. 81–88.
- [5] Cramer, H., Rost, M., Holmquist, L. E., 2011. Performing a check-in: emerging practices, norms and ‘conflicts’ in location-sharing using foursquare. In: Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services. ACM, pp. 57–66.
- [6] Cranshaw, J., Schwartz, R., Hong, J. I., Sadeh, N. M., 2012. The livelihoods project: Utilizing social media to understand the dynamics of a city. In: Proceedings of the International Conference on Weblogs and Social Media (ICWSM).
- [7] Frith, J., 2014. Communicating through location: The understood meaning of the foursquare check-in. *Journal of Computer-Mediated Communication* 19 (4), 890–905.
- [8] Gao, H., Tang, J., Hu, X., Liu, H., 2013. Exploring temporal effects for location recommendation on location-based social networks. In: Proceedings of the 7th ACM conference on Recommender systems. ACM, pp. 93–100.
- [9] Gärdenfors, P., 1993. The emergence of meaning. *Linguistics and Philosophy* 16 (3), 285–309.
- [10] Gini, C., 1912. Variabilità e mutabilità. Reprinted in *Memorie di metodologica statistica* (Ed. Pizetti E, Salvemini, T). Rome: Libreria Eredi Virgilio Veschi 1.
- [11] Glushko, R. J., 2014. *The Discipline of Organizing*. O’Reilly Media, Inc.
- [12] Janowicz, K., 2012. Observation-driven geo-ontology engineering. *Transactions in GIS* 16 (3), 351–374.
- [13] Jensen, J. R., Cowen, D. C., 1999. Remote sensing of urban/suburban infrastructure and socio-economic attributes. *Photogrammetric engineering and remote sensing* 65, 611–622.
- [14] Jordan, T., Raubal, M., Gartrell, B., Egenhofer, M., 1998. An affordance-based model of place in gis. In: 8th Int. Symposium on Spatial Data Handling, SDH. Vol. 98. pp. 98–109.
- [15] Karagiorgou, S., Pfoser, D., Skoutas, D., 2014. Geosemantic network-of-interest construction using social media data. In: *Geographic Information Science*. Springer, pp. 109–125.
- [16] Kendall, M. G., Smith, B. B., 1939. The problem of m rankings. *The annals of mathematical statistics* 10 (3), 275–287.
- [17] Kim, Y., Sohn, D., Choi, S. M., 2011. Cultural difference in motivations for using social network sites: A comparative study of american and korean college students. *Computers in Human Behavior* 27 (1), 365–372.
- [18] Lian, D., Zhu, Y., Xie, X., Chen, E., 2014. Analyzing location predictability on location-based social networks. In: *Advances in Knowledge Discovery and Data Mining*. Springer, pp. 102–113.
- [19] Lin, J., 1991. Divergence measures based on the shannon entropy. *Information Theory, IEEE Transactions on* 37 (1), 145–151.
- [20] Liu, Y., Sui, Z., Kang, C., Gao, Y., 2014. Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. *PloS one* 9 (1), e86026.
- [21] Malmi, E., Do, T. M. T., Gatica-Perez, D., 2012. Checking in or checked in: comparing large-scale manual and automatic location disclosure patterns. In: Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia. ACM, p. 26.
- [22] McKenzie, G., Janowicz, K., 2015. Where is also about time: A location-distortion model to improve reverse geocoding using behavior-driven temporal signatures. *Computers, Environment and Urban Systems* 54, 1–13.
- [23] McKenzie, G., Janowicz, K., Gao, S., Yang, J.-A., Hu, Y., 2015. POI pulse: A multi-granular, semantic signatures-based approach for the interactive visualization of big geosocial data. *Cartographica: The International Journal for Geographic Information and Geovisualization* 50, 71–85.
- [24] Noulas, A., Scellato, S., Mascolo, C., Pontil, M., 2011. An empirical study of geographic user activity patterns in foursquare. *ICWSM* 11, 70–573.
- [25] Quercia, D., Saez, D., 2014. Mining urban deprivation from foursquare: Implicit crowdsourcing of city land use. *Pervasive Computing, IEEE* 13 (2), 30–36.
- [26] Rubner, Y., Tomasi, C., Guibas, L. J., 1998. A metric for distributions with applications to image databases. In: *Computer Vision, 1998. Sixth International Conference on*. IEEE, pp. 59–66.
- [27] Rubner, Y., Tomasi, C., Guibas, L. J., 2000. The earth mover’s distance as a metric for image retrieval. *International Journal of Computer Vision* 40 (2), 99–121.
- [28] Schowengerdt, R. A., 2006. *Remote sensing: models and methods for image processing*. Academic press; 3 edition.

- [29] Shaw, B., Shea, J., Sinha, S., Hogue, A., 2013. Learning to rank for spatiotemporal search. In: Proceedings of the sixth ACM international conference on Web search and data mining. ACM, pp. 717–726.
- [30] Sheth, A., Anantharam, P., Henson, C., 2013. Physical-cyber-social computing: An early 21st century approach. *Intelligent Systems, IEEE* 28 (1), 78–82.
- [31] TEA, 2014. Tea/aecom 2013 global attractions report. [http://www.aecom.com/deployedfiles/Internet/Capabilities/Economics/\\_documents/ThemeMuseumIndex\\_2013.pdf](http://www.aecom.com/deployedfiles/Internet/Capabilities/Economics/_documents/ThemeMuseumIndex_2013.pdf), retrieved June 6, 2014.
- [32] Wang, H., Terrovitis, M., Mamoulis, N., 2013. Location recommendation in location-based social networks using user check-in data. In: Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, pp. 364–373.
- [33] Watson, G. S., 1961. Goodness-of-fit tests on a circle. *Biometrika*, 109–114.
- [34] Winter, S., Freksa, C., 2014. Approaching the notion of place by contrast. *Journal of Spatial Information Science* (5), 31–50.
- [35] Wu, L., Zhi, Y., Sui, Z., Liu, Y., 2014. Intra-urban human mobility and activity transition: Evidence from social media check-in data. *PloS one* 9 (5), e97010.
- [36] Ye, M., Janowicz, K., Mülligann, C., Lee, W.-C., 2011. 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. ACM, pp. 102–111.
- [37] Yuan, Q., Cong, G., Ma, Z., Sun, A., Thalmann, N. M., 2013. Time-aware point-of-interest recommendation. In: Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval. ACM, pp. 363–372.
- [38] Yuan, Q., Cong, G., Sun, A., 2014. Graph-based point-of-interest recommendation with geographical and temporal influences. In: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management. ACM, pp. 659–668.
- [39] Zar, J. H., 1976. Watsons nonparametric two-sample test. *Behavior Research Methods* 8 (6), 513–513.
- [40] Zhang, J.-D., Chow, C.-Y., 2013. igslr: personalized geo-social location recommendation: a kernel density estimation approach. In: Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, pp. 324–333.