Revisitation in Urban Space vs. Online: A Comparison across POIs, Websites, and Smartphone Apps.*

HANCHENG CAO†, ZHILONG CHEN, FENGLI XU, and YONG LI‡, Beijing National Research Center for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, China
VASSILIS KOSTAKOS, University of Melbourne, Australia

We present the first large-scale analysis of POI revisitation patterns, which aims to model the periodic behavior in human mobility. We apply the revisitation analysis technique, which has previously been used to understand website revisitation, and smartphone app revisitations. We analyze a 1.5-year-long Foursquare check-in dataset with 266,909 users in 415 cities around the globe, as well as a Chinese social networking dataset on continuous localization of 15,000 users in Beijing. Our analysis identifies four major POI revisitation patterns and four user revisitation patterns of distinct characteristics, and demonstrates the role of POI functions and geographic constraints in shaping these patterns. We compare our results to previous analysis on website and app revisitation, and highlight the similarities and differences between physical and cyber revisitation activities. These point to fundamental characteristics of human behavior.

CCS Concepts: • **Social and professional topics → Geographic characteristics**; • **Information systems** → *Data mining*.

Additional Key Words and Phrases: Revisitation, urban space, POI, human mobility

ACM Reference Format:
Hancheng Cao, Zhilong Chen, Fengli Xu, Yong Li, and Vassilis Kostakos. 2018. Revisitation in Urban Space vs. Online: A Comparison across POIs, Websites, and Smartphone Apps.. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 4, Article 156 (December 2018), 24 pages. https://doi.org/10.1145/3287034

1 INTRODUCTION

We present the first large-scale analysis of Point of Interest (POI) revisitation patterns in cities. Revisitation refers to the user behavior of returning to (i.e. re-visiting) the same service/location over time. For example, users access their personal email multiple times every day, or may revisit the same restaurant every month. Previous work has conducted revisitation analyses to understand how users revisit websites on the Internet [1] or apps on their smartphones [2], where very similar revisitation patterns are observed. While there are extensive work on cyber

*This work was supported in part by the National Key Research and Development Program of China under grant 2017YFE0112300, the National Nature Science Foundation of China under 61861136003, 61621091 and 61673237, Beijing National Research Center for Information Science and Technology under 20031887521, and research fund of Tsinghua University - Tencent Joint Laboratory for Internet Innovation Technology.
†The author is now with Stanford University.
‡Corresponding author.

Authors' addresses: Hancheng Cao, hanchcao@stanford.edu; Zhilong Chen, chen-zl16@mails.tsinghua.edu.cn; Fengli Xu; Yong Li, liyong07@tsinghua.edu.cn, Beijing National Research Center for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing, 100084, China; Vassilis Kostakos, University of Melbourne, Australia, vassilis.kostakos@unimelb.edu.au.

2474-9567/2018/12-ART156 $15.00
https://doi.org/10.1145/3287034

revisitation activities, revisitation in the physical world has never been explicitly studied. As human mobility shows strong periodicity[3, 4], we are curious to investigate what are the revisitation patterns in the physical world. In this study, we conduct the revisitation analysis on POIs in cities around the globe by focusing on the following three aspects: (1) Understand what is the typical revisitation pattern for different types of POIs; (2) Investigate whether different users have different revisitation patterns; and (3) Compare revisitation patterns in urban space vs. online.

Understanding the revisitation patterns in human mobility is important, especially on a population scale. On the one hand, such analysis reveals latent urban rhythms as we get to know the frequency of people revisiting a particular location. On the other hand, we are able to characterize locations and users into different behavioral clusters, which provides key insights for urban planning, decision making and marketing (e.g., building recommendation system). Finally, by comparing human mobility to website usage and app usage, we can begin to identify fundamental behavioral patterns in how people explore, visit and revisit services.

In our analysis, rather than identifying a definite period for revisitation, we adopt a non-deterministic view to model revisitation as a probability distribution over time, which represents the likelihood of a POI being revisited – or a user revisiting a POI – after a certain time interval. The reason is that multiple revisitation time intervals may exist even for the same POI or user, which cannot be modeled as a definite period. For instance, doctors may revisit a hospital every day, but patients revisit the hospital ranging from weekly to monthly basis, making it hard to define a 'universal' revisitation period for the hospital. However, it is meaningful to model the hospital's probability distribution of revisitation time intervals, which in this case we characterized as 'hybrid'. We apply the revisitation curve method proposed in [1], which has been adopted more widely [2, 5], thus making our findings directly comparable to previous work. Leveraging a 1.5-year-long foursquare check-in data with 266,909 users in 415 cities around the globe, and a 1.5-month-long Chinese social networking dataset on continuous localization of 15,000 users, we derive revisitation curves for both POIs and users, and cluster these into distinct patterns. We analyze the relationship between revisitation clusters and POI functions, as well as the relationship between POIs and user revisitation patterns. Furthermore, we analyze the relationship between revisitation and distance to home/workplace.

Our key findings are summarized as follows:

- We recognize four distinct revisitation patterns for both POIs and users, and demonstrate that they correspond to POI functions. We discover that POI and user revisitation curves demonstrate strong correlation of statistic significance, and thus highlight the interplay between urban rhythms and human mobility.
- We study the revisitation patterns in cities across the world, and show the effects of cultural and economic factors on revisitation behaviors.
- We identify the similarities and differences between physical (POIs) and online revisitation patterns (websites and apps): physical revisitation tends to be longer in periods and subject to geographic constraints.
- We compare the revisitation results between Foursquare check-in data and Chinese social localization data, and highlight the fundamental differences between nature of active check-ins and passive localizations.

2 RELATED WORK

2.1 Periodic Pattern and Routine Mining in Spatial Temporal Data

Periodic pattern mining is one important topic in time series analysis [6]. As a special case of time series, mining periodic patterns in spatial temporal data has attracted much attention. Previous work mostly focused on identifying the popular location sequences shared by different trajectories and grouping trajectories based on their physical closeness [7, 8]. Jeung et al. [9] proposed the *T-patterns* framework to address the problem of detecting frequent sub-trajectories in spatiotemporal data, and Zheng et al. [10] investigated the problem of detecting frequent traveling paths between fixed locations. To detect periods in movement data through reference spots, Li et al. [11] proposed a Fourier transform and autocorrelation based algorithm. A probability measure for detecting periods was later introduced [12–14] to address the noise and incomplete observations in trajectory data. Yuan et al. [15] proposed a Bayesian non-parametric model to discover periodic mobility patterns by jointly modeling geographic and temporal information. Periodic mining has also been applied to summarize traffic conditions on road networks [16]. However, this line of works mostly focused on techniques rather than analysis, and therefore does not contribute much to our understanding of patterns in urban rhythm and human mobility. Moreover, it is very difficult to define a fixed periodicity for POI/user revisitation in our case since multiple periodic patterns exist. Instead of proposing methods to identify specific periods or periodic patterns, in this paper, we focus on characterizing different patterns of revisitation and understanding their underlying characteristics.

Meanwhile, spatial temporal data has been used for studying human routine behaviors. Gonzalez et al. [3] have shown that human periodically return to a few previously visited locations, and that human mobility can be modeled by a stochastic process around a fixed point. Jiang et al. [4] proposed the timeGeo model to capture the revisit and exploration pattern of individual mobility. Eagle et al. [17] utilized PCA algorithm to extract the features from semantic annotated spatial temporal data, and then identified clusters of activity patterns. Xu et al. [18] developed a framework to cluster popular temporal modes in human trajectory, which helped understand how people allocate time in their daily lives. Different from them, we model human routine from a novel perspective of revisitation analysis. Our work jointly analyzes urban dynamics and human mobility through POI revisitation, and provides insights for the interplay between human periodic behaviors and urban rhythms of POI revisitation. We show how the function of place, and geographic constraint influence revisitation patterns in human mobility.

2.2 Location and POI Visitation Modeling

Previous work has considered location and POI visitation modeling, providing insights for urban planning, decision making and advertising (e.g., recommendation system). A number of studies focused on user visitation modeling. For instance, Yang et al. [19] proposed the SEAL system to determine users' sentiments regarding locations from check-in data, and to characterize fine-grained user location visit preferences. Cho et al. [20] investigated the relationship between location visitation and friendship. Moreover, user location and POI visitation have been leveraged for POI recommendation [21–23]. Meanwhile, there are also studies that characterize location visitation patterns: Yuan et al. [24] developed a framework to discover regions of different functions in a city using both human mobility across regions and points of interests (POIs) located in a region. Fan et al. [25] proposed a novel tensor factorization approach to modeling city dynamics in a basic life pattern space (CitySpectral Space). Miranda et al. [26] designed a visual exploration framework that allows users to explore spatial temporal activities across multiple temporal resolutions in cities. Zhang et al. [27] modelled the semantic meaning of spatial-temporal points based on their co-occurrence with the texts in social media's check-ins. Works have also been done to model POI visitation demand [28] and POI evolution[29, 30]. However, to the best of our knowledge, none of the existing work has considered human revisitation patterns at specific locations, and thus fails to capture the close relationship between human mobility, urban periodic patterns, and the specific functions of a location. In our analysis, we show that revisitation patterns correspond to the nominal functions of any given location. In other words, urban function influences how people revisit the region, and this in turn shapes the unique urban rhythm in cities.

2.3 Revisitation Analysis on the Internet

Previous studies have investigated people's web surfing and browsing behaviors, especially webpage and app revisitation habits. It has been demonstrated that revisitation is a frequent pattern in online use, making up a large proportion (50% to 80%) of all online activities [31–33]. Obendorf et al. [34] found that users' online revisitation behaviors vary dramatically. The behaviors are strongly affected by personal habits and types of sites visited, which can be characterized into short-term, medium-term and long-term revisits. While short-term revisits are often a result of backtrack or undo, medium-term revisits come from re-utilizing and rediscovering. To study the effects of web content on revisitation, Adar et al. [1] proposed a clustering method based on revisitation curves to group webpages of similar revisitation activities, and discovered four primary revisitation patterns with unique behavioral and content characteristics. More recently, Jones et al. [2] applied the same technique on smartphone app usage data, and demonstrated much of smartphone app revisitation on a macro-level resembles web browsing on desktops, thus concluding that smartphone usage is driven by innate patterns in cyber space rather than technology characteristics. Revisitation curves have also been used to characterize users' online revisitation patterns [5]. Our work is well aligned with the aforementioned work in analyzing human revisitation behavior, which is perhaps the most important and interesting pattern in human behavior. However, unlike previous work, we focus on revisitation pattern in the physical space rather than cyber space. Visits and transitions in physical space are likely to bear a high cost in terms of time and resources, at least in comparison to visiting websites or smartphone apps. Nevertheless, we have analyzed our data using the revisitation curve method used in [1, 2, 5], so that our findings on physical revisitation can be directly compared to previous work.

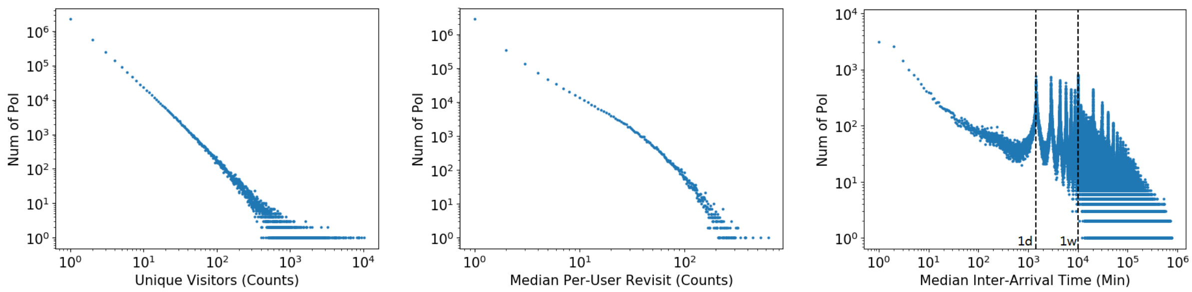
3 DATASET

In this paper, we reuse two large-scale datasets to model revisitation in cities across the world.

Global-scale Check-in data. This publicly available check-in dataset comes from Foursquare, one of the world's largest location-based social networking services. The dataset includes 33,278,683 check-in records of 266,909 users at 3,680,126 unique POIs between April 2012 and September 2013 in the most checked 415 cities worldwide[35, 36][1]. Each check-in record consists of user ID, venue ID, UTC time, and timezone offset in minutes. Venue information and city information are additionally offered as well. The POIs are grouped into 429 categories, and further clustered into 9 major categories: colleges & universities, great outdoors, shop & service, arts & entertainment, food, travel & transport, nightlife spots, residence and professional & other places[2]. We used the 9 major POI categories when studying the relationship between POI function and revisitation pattern, while using the fine-grained categories when comparing different cities. The statistical properties of this dataset are shown in Fig. 1. From the figure, we can observe that the dataset exhibits power law distributions, where lots of POIs have few unique visitors and revisits. Thus data filtering procedure has to be included as part of the analysis to remove noises and incomplete observations. Meanwhile, an interesting point to notice is that the dataset deviates at 1 day, 2 days, 1 weeks, etc. in median inter-arrival time from power law, indicating the significance of daily and weekly revisitation patterns in human check-in behaviors.

The global Foursquare dataset also makes it possible to compare POI revisitation patterns across cities, which we presented in section 5.1.3. Based on the city characterization by Globalization and World Cities Research Network (GaWC)[37], where cities are classified into levels of Alpha++, Alpha+, Alpha, Alpha-, Beta+, Beta, Beta-, Gamma+, Gamma, Gamma- according to their advanced producer services (e.g., accounting, finance, consultancy, etc.), we selected 5 cities worldwide for analysis, each of which is representative with abundant check-in records. Specifically, New York is chosen as a representative of western metropolis ranking Alpha++, Tokyo an eastern metropolis ranking Alpha+, Sydney an Alpha city in the southern hemisphere, Vienna an Alpha- European city rich in cultural heritage and Rio de Janeiro a Beta- city in a developing country. We present detailed statistics of Foursquare records in the five cities, as shown in Table 1.

[1]Data Source: https://sites.google.com/site/yangdingqi/home/foursquare-dataset
[2]https://developer.foursquare.com/docs/resources/categories



(a) Unique visitors per POI (b) Median per user revisit count per POI (c) Median inter-arrival time per POI

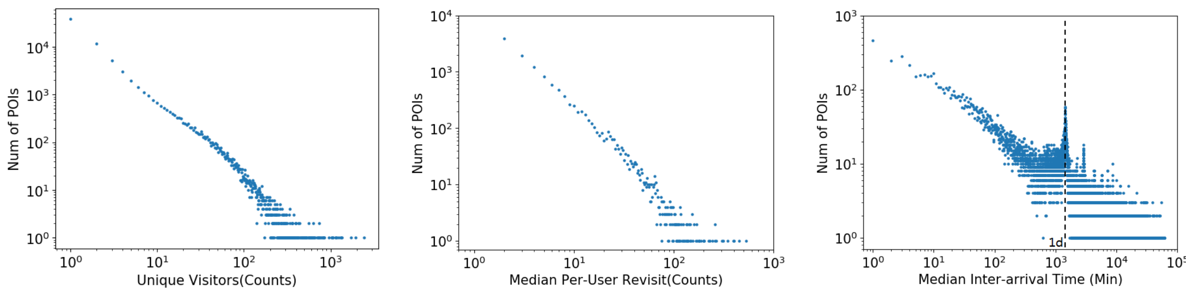
Fig. 1. Foursquare check-in dataset statistics.

Table 1. Statistics of Foursquare records in 5 selected cities.

City	Number of POIs	Number of Check-in Records	Number of Users
New York	8,276	212,919	33,663
Tokyo	17,514	732,032	44,116
Sydney	1,915	34,171	4,137
Vienna	878	16,016	2,679
Rio de Janeiro	1,825	37,326	3,267

Localization Data in Beijing. The dataset was collected by one of the largest Chinese social networking platforms, which is also the most popular localization platform in China. When third-party apps which leverages its localization API are recalled, the users' location information will be uploaded to the servers and recorded by this platform. Specifically, the localization module takes the signals of WiFi connection, GPS and base station connection as inputs and is able to determine the location with an error of less than 22.5 meters in more than 90% of all cases. Apart from users' active checking-in POIs, the platform also infers the POI of the user based on the platform's detailed POI boundary database with high accuracy. Through associating the actual locations in physical location trace with the spatial coverage of POIs, the platform is able to extract the actual POIs users locate in. POIs on the localization data are grouped into 18 categories, including food, company, institute, shop, life service, entertainment, gym, automobile, hospital, hotel, tourist attraction, cultural, school, bank, transportation, residence, factory and office.The utilized dataset in our analysis is collected from $15,000$ anonymous users, who were active between Sept. 17th - Oct. 31st, 2016 in Beijing. It contains 76,298 unique POIs and 3,097,863 localization records in total. Each record consists of the following fields: the anonymized user ID, time, location, the associated POI, and POI type information. Some statistical properties of this dataset are shown in Fig. 2, and similar distribution as Foursquare is observed on this data. The power law distribution of the two datasets motivate us to carry out analysis on an exponential scale in the analysis latter.

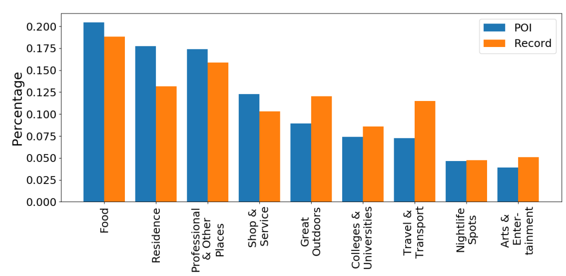
We noted that a major difference between these two datasets is that Foursquare records users' active check-in, which are more likely to happen at 'unusual' places, while the social localization platform passively records user locations. We plot the distribution of POI categories and check-in/localization record categories in Foursquare and social localization dataset in Fig.3. As shown in the figure, the social localization dataset has much more 'residence' POIs and records while Foursquare dataset record more recreation related POIs and check-ins. As a result, the social localization dataset is more likely to reflect users' real-life patterns such as home and workplace revisit compared to Foursquare dataset. Despite the fact that a localization platform is a more accurate data



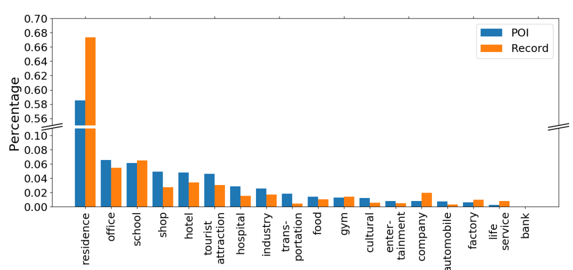
(a) Unique visitors per POI (b) Median per user revisit count per POI (c) Median inter-arrival time per POI

Fig. 2. The social localization dataset statistics.

source for studying user revisitation patterns, localization data is quite often not publicly available and limited in coverage. In comparison, check-in platforms such as Foursquare are open data sources and cover users around the globe. Therefore, in our work, we leverage both Foursquare check-in dataset and social localization dataset to ensure the generalizability and reliability of our findings. Moreover, as we will show later, in the very case of revisitation analysis, check-in dataset reveals similar patterns as that of localization data.



(a) Foursquare



(b) Social localization

Fig. 3. Distribution of POI categories and check-in/localization record categories in Foursquare and social localization dataset.

Ethical Considerations. We took careful steps to address privacy issues regarding the sharing and mining of user location and check-in data. In terms of localization data in Beijing, the Terms of Service for the social localization platform included consent for research studies. It shared user data with us after preprocessing the data to protect user privacy. All user identifiers have been replaced with secure hashcodes to improve anonymity. In terms of Foursquare check-in data, it is a public dataset that is available for everyone. Furthermore, our research protocol has been reviewed and approved by our local university institutional board. Finally, all data is stored in a secure off-line server, with access limited to only authorized members of the research team bound by strict non-disclosure agreements.

4 METHDOLOGY

4.1 Data Pre-processing

To extract credible revisitation behaviors, we first removed noise and incomplete records from our dataset. Due to the location service and data collection mechanism, we observed that there can be multiple recorded visits at the same location within a short time. For example, in the Foursquare check-in dataset, we noticed that some users check-in many times at the same place within one visit (e.g., one user frequently checks-in at Central Park Zoo during one visit to share his thoughts with others). We did not consider such records as valid revisitations and removed them by asserting a revisitation is valid only if the inter check-in time at a location is longer than 30 minutes. We empirically derived this threshold by manually inspecting the data.

The localization nature of the social networking dataset makes it even more likely that a user's location is recorded multiple times during their stay at a place. For instance, a user may use one location-based service at place A, and 10 minutes later launch another location-based service at the same place, thus creating two records in the social localization dataset without moving at all. Furthermore, the localization data is very likely to capture locations where a user is actually passing by (e.g., looking up POI recommendation when walking in the street) and we also ignored such passing-by places since they are meaningless for revisitation analysis. Therefore, we removed such instances by applying a stay region detection algorithm on the localization dataset, where we only considered locations where the user stays longer than 30 minutes. Again, we empirically derived this threshold by manually inspecting the data.

Finally, to capture the main behavior patterns, we removed users with few remaining records to characterize their mobility: 5 and 10 for Foursquare check-in and localization datasets, respectively. We also filtered out POIs with few overall visits, with a threshold of 5 and 10 for Foursquare and localization data, respectively. Eventually, the Foursquare dataset retains 243,899 users, 951,427 POIs, and 17,136,200 check-ins. The localization dataset retains 11,448 users, 14,749 POIs, and 767,642 stays. The pre-processed two datasets remain large-scale both in user number and POI number, thus they are ideal for studying the characterization of urban revisitation activities in population.

Note that we have systematically tuned all threshold parameters so that we get the optimal revisitation patterns in analysis while filtering out the least data in the preprocessing step.

4.2 Revisitation Curve Representation

Next, we applied a revisitation curve method proposed in [1, 2] to extract features of POI revisitation patterns.

The revisitation curve represents the number of times a POI is revisited within a predefined time interval. We constructed two types of revisitation curves for later analysis: per POI, and per user. For each POI, we constructed a revisitation curve using all user's inter arrival time between revisits to this POI. Therefore, a POI revisitation curve indicates how often any given user comes back to the same POI. For each user, we constructed a revisitation curve by considering the user's inter arrival time between revisiting the same POIs. Therefore, a user revisitation curve indicates how often that user revisits any given POI.

We used an exponential scale to construct the revisitation curve bins, following the same method from Ref. [1, 2]. The following time bins were used: <1 hour, 4 hours, 12 hours, 1 day, 1.4 days, 2 days, 3 days, 4 days, 5.7 days, 1 week, 1.4 weeks, 2 weeks, 3 weeks, 1 month, 2 months, >2 months so as to ensure (1) observations from the entire dataset lie approximately evenly in each bin, (thus each bin is equally weighted, leading to better clustering result); and (2) the timescale of bins are highly interpretable. For the social localization dataset, since the total duration is 1.5 months, we merged the last 3 bins. For each POI or user curve, we obtained a count of the number of inter arrival times between revisits within these predefined bins. Therefore, the revisitation pattern for each POI or user can be represented by a feature vector consisting of those counts per bin.

We present two examples to concretize the revisitation curve representation. One exemplary revisitation curve is NYC Times Square's, which is delineated in Fig. 4 in blue. With revisitation information aggregated from all users visiting the POI, the average frequency/probability of a revisitation to the POI to happen within a specific time period (which x-axis represents) is manifested by the curve. For example, the curve indicates that once a person visits Times Square, he has a probability of 11.7% to revisit the place after half a day to one day's time. Similar curves are also portrayed for each user, as suggested in [5], where one's revisitation patterns for each POI he visits are aggregated so as to measure his probability of revisiting the POI after a specific period of time. For example, a revisitation curve for a single user is illustrated in Figure 4 in red, which shows that once we spot his visit to a POI, we are most likely to spot him in the same place again after a period of 4 to 12 hours.

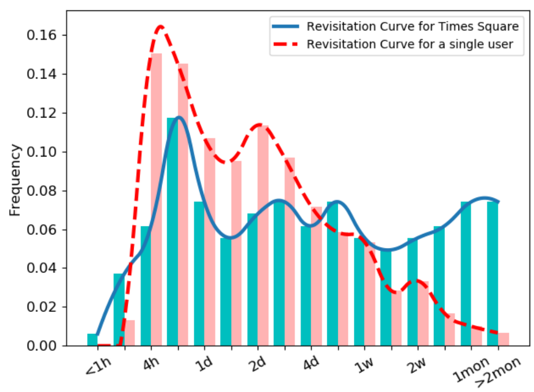


Fig. 4. Revisitation histogram and curve for a single POI (Times Square) and for a single user.

4.3 Pattern Clustering

After constructing the revisitation curves for each POI and user, we applied the K-means algorithm to cluster POIs and users of similar revisitation patterns. We ran k-means algorithm by 'Euclidean distance' on multiple k values and constructed an elbow plot to decide the best k value: by increasing k from 2 to 30, and calculating the sum of error from samples to the cluster center, we choose k at elbow point where the sum of error does not drop significantly compared to other points as the number of clusters.

4.4 Home/workplace Estimation

To investigate the effect of geographic influence on the POI revisitation pattern, we estimated the home and workplace region for each user in the social localization dataset. We first removed users with too few records (fewer than 20 records) in the original dataset and then leveraged a widely-used home and workplace inference techniques [38–40]: for every user, their most recorded POI everyday from 7pm-8am is labelled as their home, while the most recorded POI every weekday from 8am-7pm in the day is labelled as their workplace. We eventually identified the home POI for 12,892 users, and the workplace POI for 13,584 users. We ignored this analysis on the Foursquare dataset due to its sparsity and the fact that many people do not check-in at their homes/workplace.

5 ANALYSIS

5.1 POI Revisitation Patterns

5.1.1 POI Revisitation Patterns: Foursquare. We carried out analysis on global Foursquare dataset to investigate revisitation patterns in 415 cities across the world, where we first leveraged the nine major Foursquare POI categories to characterize POI function on a global scale. In the Foursquare study, our analysis identified 10 clusters of POIs, as shown in Fig. 5. Based on the curves' shapes, these clusters were then ordered, named and manually classified into four groups: slow, medium, fast and hybrid, as in previous work. Basic information about the clusters are shown in Table 2, including curve shape, cluster size, and characteristic POI categories. Characteristic POI categories of the patterns are selected by judging if the proportion of a certain POI category belonging to the pattern is significantly greater than the global proportion. We present the ratio of in-cluster proportion of characteristic POI categories to the global proportion in brackets in 'Characteristic POI Categories' of Table 2. For instance, Residence (1.88) for F pattern means that the proportion of the residence category within revisitation pattern F is 1.88 times the proportion of the residence category observed in the entire dataset.

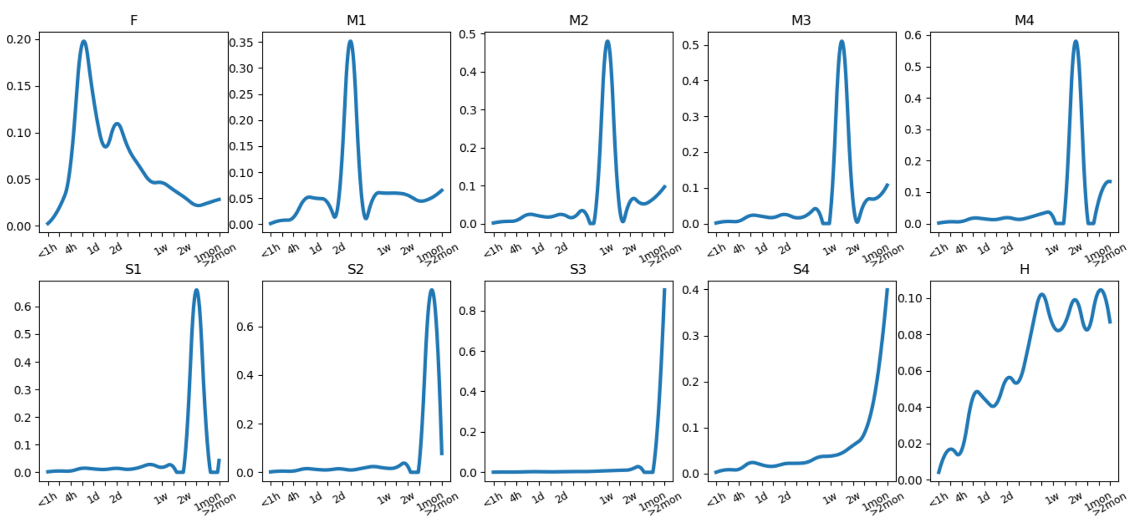


Fig. 5. Centroid POI revisitation curves for Foursquare data.

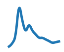
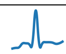
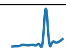
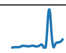
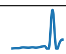
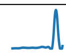
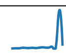
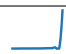
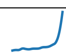
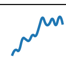
Fast Revisits In terms of fast revisitation, one cluster (F) accounting for 22.1% of all POIs falls into the group, with people revisiting the POIs mostly within a single day's time. POIs in this group are more likely to be within the categories of residence, travel & transport and professional & other places.

Residence: The most over-represented category in the group of fast revisitation pattern is residence. We deemed that this higher likelihood to observe residences in this group is associated with people's daily mobile movements: people visit their homes at least once a day.

Travel & Transport: POIs identified as travel & transport, such as hostels and hotels, have a higher tendency to be within the scope of fast revisitation pattern when compared to the average. These POIs serve as temporary residences for people and this similarity can explain their revisitation patterns' analogy with residence's.

Professional & Other Places: The portion that the category of professional & other places occupies in the group of fast revisitation pattern is significantly larger than the average. This higher likelihood is associated with the

Table 2. POI revisitation cluster groups for Foursquare data.

Description	Label	Curve Shape	Cluster Size	Characteristic POI Categories
Fast Revisits (within a day)	F		209984 (22.1%)	Residence (1.88), Travel & Transport (1.38), Professional & Other Places (1.27)
Medium Revisits (around 1 week)	M1		63605 (6.7%)	Colleges & Universities (1.13), Shop & Service (1.08),Food (1.08)
	M2		38296 (4.0%)	
	M3		36365 (3.8%)	
	M4		34086 (3.6%)	
Slow Revisits (around 1 month)	S1		25108 (2.6%)	Food (1.74), Nightlife Spots (1.56), Shop & Service (1.38)
	S2		41164 (4.3%)	
	S3		68888 (7.2%)	
	S4		167961 (17.7%)	
Hybrid	H		265970 (28.0%)	Colleges & Universities (1.46), Great Outdoors (1.25), Residence (1.33)

fact that these POIs are very likely to be workplaces. With the assertion that workplaces are most likely to be revisited at least daily, the result can be viewed as an interpretation of people's daily routine.

Medium Revisits Medium revisitation pattern make up 18.1% of all POIs. This group consists of four clusters (M1-M4), the revisitation peaks of which are on the weekly level. In specific, cluster M1 has peak revisitation after a period slightly less than 1 week. Peaks for POIs in cluster M2 can be observed at the time interval of approximately 1 week. POIs that belong to cluster M3 are most likely to be revisited after a period of 1 to 2 weeks. Cluster M4 reaches its highest probability of revisitation at approximately 2 weeks. Characteristic POIs of the group, when viewed collectively and compared to the average distribution, include the categories of colleges & universities, shop & service and food.

Colleges & Universities: POIs in the category of colleges & universities are more likely to appear in clusters belonging to the group of medium revisitation. This is probably because these places for studies are revisited neither too often nor too infrequently.

Shop & Service: Shop & service stands out within medium revisitation pattern. This category covers tailor shops, hunting supplies, pet services, etc. These POIs serve people's daily routines, but once people's needs are satisfied through one visit, there is no need to revisit the place in a while and medium revisitation patterns emerge.

Food: A major portion of POIs in the medium revisitation group falls into the category of food. One interpretation is that food places are visited once in a while, such as when treating others or meeting old friends.

Slow Revisits Overall, 31.9% of all POIs have slow revisitation patterns (clusters S1-S4). Specifically, the revisitation peak of cluster S1 occurs between 2 weeks and 1 month. Cluster S2 shares the same shape as cluster S1, but its revisitation peak moves to approximately 1 month. POIs in S3 are most likely to be revisited when more than 2 months have passed since the last visit. The probability of revisitation for POIs in cluster S4 increases as time passes. When these four clusters are considered as a whole, their POIs tend to belong to the categories of food, nightlife spots and shop & service.

Food: Food is once again observed when characterizing categories with slow revisitation pattern, indicating people do not have frequent needs for some restaurants.

Nightlife Spots: Many POIs with slow revisitation pattern are nightlife spots, including pubs, bars and nightclubs. We assumed that these POIs are recreational places and their nature of amusement is likely to result in long intervals between two consecutive pairs of visits and thus reveals slow revisitation patterns.

Shop & Service: Shop & service emerges within slow revisitation pattern again, demonstrating people's demands for certain services are on monthly basis.

Hybrid Revisits One cluster (H) with 28.0% of all POIs tends to be the combinations of the revisitation patterns aforementioned, and is thus labelled as hybrid. These POIs tend to fall into the categories of college & universities, great outdoors, arts & entertainment and travel & transport.

Multi-functional Places: POIs in this group have a higher probability to perform multiple functions. For example, a location may be a school for students but can be a 'park' for citizens, and a park could be visited with different purposes for different kinds of people (e.g., visitors, regular exercisers, workers). The diversity of people's motivations in visiting a single place contributes to the variety of people's revisitation patterns. Combining these weighted patterns together, we obtain hybrid revisitation pattern.

5.1.2 POI Revisitation Patterns: Social Localization Data. Our analysis identified 10 clusters of POIs in the social localization dataset, shown in Fig. 6. After the analysis of the curves' shapes, ordering, naming and manual classification were done. As Table 3 demonstrates, slow, medium, fast and hybrid revisitation patterns were also found.

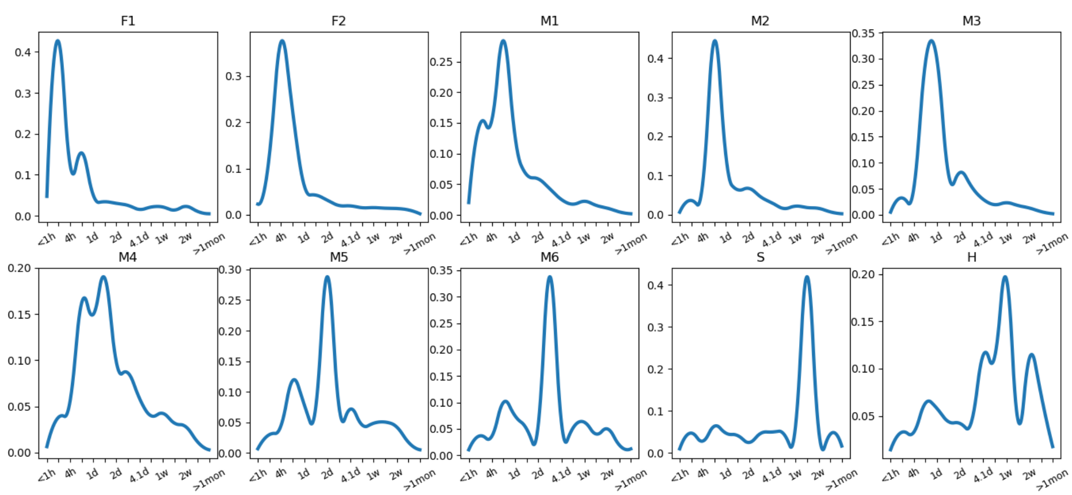
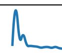
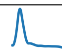
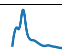
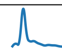
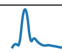
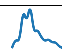
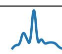
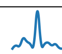
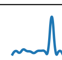
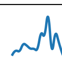


Fig. 6. Centroid POI revisitation curves for the social localization data.

Fast Revisits The group of fast revisitation takes up 11.9% of all POIs and includes 2 clusters (F1 & F2) where the POIs are revisited mostly within a time period of half a day. Specifically, cluster F1 peaks at approximately 1 hour, with another local maximum at around 12 hours. Cluster F2 has only one peak at around 4 hours. When

Table 3. POI revisitation cluster groups for social localization data.

Description	Label	Curve Shape	Cluster Size	Characteristic POI Categories
Fast Revisits (shorter than 1 half day)	F1		885 (6.0%)	Hotel (1.62), transport (1.81), cultural (1.69), tourist attraction (2.02)
	F2		867 (5.9%)	
Medium Revisits (around 1 day)	M1		3440 (23.3%)	Life service (1.63), company (1.51), institute (1.36), residence (2.48), industry (1.61), office (1.76), school (1.25), entertainment (1.31), restaurant (1.27)
	M2		1648 (11.1 %)	
	M3		2772 (18.8%)	
	M4		2069 (14.0%)	
	M5		1151 (7.8%)	
	M6		684 (4.6%)	
Slow Revisits (around 1 week)	S		297 (3.0%)	Entertainment (3.86), gym (2.13), tourist attraction (1.56), shop (1.44)
Hybrid	H		946 (6.4%)	Food (2.21), hospital (1.63), shop (1.52), cultural (1.64), transport (3.80)

considering these clusters as a group, POIs in this group are more likely to belong to the categories of hotel, transport, cultural and tourist attraction.

Hotel: POIs of hotel category are more likely to exhibit fast revisitation patterns. Possibly this is because hotels are temporary stops for tourists. Compared to home residence that show regular daily basis revisitation, the nature of hotels makes them more probable to be revisited within a shorter period of time (possibly the time to visit a famous spot) and thus exhibit fast revisitation patterns.

Transport: We observed that if a POI is pertinent to transport, its probability of being found in fast revisitation pattern is significantly higher than average. This probably owes to the fact that transport-related POIs, such as bus stops and subway stations, are visited both before and after people go to a certain place.

Cultural and Tourist Attraction: Cultural POIs and tourist attractions, e.g., museums and scenic spots, are over-represented in fast revisitation pattern. We believe that this phenomenon is due to the fact that during a single trip to a POI of these kinds, people may visit somewhere else for other purposes (for example, having lunch) and go back after a short period of time, leading to fast revisitation behaviors.

Medium Revisits The majority of POIs (79.6%) in this group are revisited on a daily basis, and six clusters (M1-M6) show up. Specifically, peaks of cluster M1, M2 and M3 are at approximately 12 hours, but cluster M1 has a local maximum at around 4 hours. The revisitation curves of cluster M2 and M3 share a similar shape, but cluster M3's is not as centralized. Cluster M4's peak revisitation interval is between half a day to 2 days. The likelihoods of POI revisitation in cluster M5 and cluster M6 reach their climaxes at approximately 2 days and more than 2 days respectively. When these clusters are considered as a group, we observe a higher tendency for the medium revisitation group to be within the categories of life service, company, institute, residence, industry park, office, school, entertainment and restaurant.

Daily routines: Similar to Foursquare data, characteristic POIs that exhibit medium revisitation patterns are for daily routines. For example, companies, institutes, industry and offices are places for everyday work; residences are where people go home for daily; some restaurants are routinely for meals and some places for relaxation are for daily relief.

Slow Revisits One small cluster (S) accounting for 3.0% of all POIs and peaking around 1 to 2 weeks is classified as slow revisitation pattern. POIs within this group are more likely to relate to the categories of entertainment, gym, tourist attraction and shop.

Places for non-daily needs: Similar to Foursquare data, POIs in this group have a higher likelihood to correlate with people's non-daily needs than the average case. Similar to the aforementioned Foursquare analysis, this higher probability could be interpreted by the assumption that people's demands for these spots are on a scale of longer time.

Hybrid Revisits One cluster (H) with 6.4% of all POIs is recognized as hybrid. POIs in this cluster are more likely to fall into the categories of food, hospital, shop, cultural and transport when compared to the average.

Multi-function Places: Similar to Foursquare data, a large portion of categories pertinent to hybrid revisitation pattern are places with multiple functions. For example, hospitals show diverse functions for doctors, care workers and patients in need of long-term medication. Here the combination of different people's diversified patterns of visitation leads to the hybrid revisitation pattern.

Our result shows that Foursquare and social localization dataset reveal consistent POI revisitation patterns: slow, medium, fast and hybrid, and demonstrate close correlation with different POI categories. As our findings are well aligned between passive localization data which reflects user 'real' living pattern, and active check-in datasets covering cities across world, we argue both reliability and generalizability of our findings. However, we also note the proportion difference of revisitation patterns identified by the two datasets, where percentage of slow revisitation pattern POIs in social localization dataset is far less than Foursquare dataset, and percentage of medium revisitation pattern POIs in social localization dataset is much more prevailing, highlighting the phenomenon that users often check-in at 'unusual' places which demonstrates slow revisitation rather than routine POIs of medium revisitation pattern.

5.1.3 Cross-city Comparison. We further examined revisitation patterns across the five selected cities: New York, Tokyo, Sydney, Vienna and Rio de Janeiro.

We present cluster size, and characteristic POI categories for different revisitation patterns of the five cities in Table 4. We also show the ratio of each cluster size over global percentage for better comparison. For instance, 20.0% of all POIs in New York demonstrates fast revisitation pattern, while 22.1% of all POIs in the Foursquare dataset shows fast revisitation, thus the 'fast revisitation pattern' cluster size in New York is 0.91 times of the global 'fast revisitation pattern' cluster size. We present the ratio, 0.91 in the corresponding bracket. For better comparison across different cities, we present the fine-grained venue instead of major POI categories in this analysis.

As shown in Table 4, similarities across different cities are observed. We notice that the four revisitation patterns: fast, medium, slow and hybrid are reflected in each city. Moreover, the function of POIs for each revisitation pattern show strong resemblance across the five cities: POIs related to daily routines (e.g. residence, office, daily transport) stand out in fast revisitation pattern, while POIs of non-necessities fall into groups with longer revisitation intervals.

Meanwhile, we can observe obvious differences across cities. First, the distribution of the four revisitation patterns vary from city to city, with a trend that more developed cities have fewer 'fast revisitation' POIs and more 'slow revisitation' POIs. We ascribe this phenomenon to the fact that more developed cities generally have

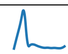
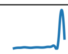
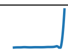
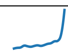
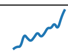
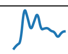
Table 4. POI Revisitation Comparision across Cities.

City	Fast		Medium		Slow		Hybrid	
	Size	Typical Venues	Size	Typical Venues	Size	Typical Venues	Size	Typical Venues
New York	20.00% (0.91)	Home, Hotel, Office	17.77% (0.98)	Coffee Shop, Subway, Park	36.00% (1.13)	Bar, American Restaurant, Grocery Store	26.23% (0.94)	Home, Office, Coffe Shop
Tokyo	10.20% (0.47)	Train Station, Conve-nience Store, Subway	15.50% (0.86)	Japanese Restaurant, Ramen/ Noodle House, Conve-nience Store	50.90% (1.60)	Ramen/ Noodle House, Sake Bar, Mall	23.33% (0.83)	Conve-nience Store, Train, Ramen/ Noodle House
Sydney	14.15% (0.64)	Hotel, Home, Train Station	18.66% (1.03)	Café, Train Station, Pub	42.74% (1.34)	Café, Pub, Bar	24.45% (0.87)	Café, Train Station, Gym
Vienna	25.99% (1.18)	Hotel, Home, Office	18.62% (1.03)	Train Station, Subway, Café	28.81% (0.90)	Café, Grocery Store, Restaurant	26.58% (0.95)	Office, Light Rail, Restaurant
Rio de Janeiro	28.06% (1.27)	Home, Residential Buiding, Bus Line	17.70% (0.98)	Neighbor-hood, Church, Gas Station	22.25% (0.70)	Bar, Neighbor-hood, Restaurant	31.99% (1.14)	Church, Gym, School

greater percentage of recreational & cultural POIs in POI constitution. Since POIs of fast revisitation patterns are more likely to be POIs related to daily routine, and POIs of slow revisitation patterns are associated with amusement and entertainment, a greater percentage of recreational & cultural POIs in developed cities will result in a larger portion of fast revisitation pattern. Second, characteristic POI categories of the four revisitation pattern groups differ from city to city. For instance, although food related POIs are found in slow revisitation patterns in five cities, in Tokyo food related POI is reflected as ramen/noodle house while in New York it is American Restaurant.

The result across cities indicates that although revisitation pattern is influenced by economic and cultural factors, the characterization of the four POI revisitation patterns: fast, medium, slow and hybrid, and their correlation with different urban functions (POI types), are universal, which further verifies the generalizability of our findings.

Table 5. User revisitation cluster groups for the Foursquare data.

Description	Label	Curve Shape	Cluster Size	Characteristic POI Categories
Fast Revisits (within a day)	F		6276 (2.6%)	Travel & Transport (1.90), Residence (1.60)
Slow Revisits (around 1 month)	S1		4008 (1.6%)	Arts & Entertainment (1.63), Nightlife Spots (1.59), Food (1.56)
	S2		7515 (3.1%)	
	S3		46501 (19.1%)	
Hybrid	H1		89704 (36.8%)	Residence (1.11), Colleges & Universities (1.08), Professional & Other Places (1.06)
	H2		89893 (36.9%)	

5.2 User Revisitation Patterns

So far we have investigated the revisition patterns of specific POIs. Now, we turn to the patterns for individual users. Fig. 7 and Fig.8 show the centroids of the 6 identified user clusters in Foursquare dataset and the 5 identified user clusters in social localization dataset, respectively. Similar to the analysis of POI clusters, we ordered, named and manually classified the clusters into the groups of slow, medium, fast and hybrid ones referring to the shapes of the curves. Table 5 and Table 6 present description, curve shape, cluster size, and characteristic POI categories (ratio of in-cluster proportion of characteristic POI categories to global proportion in brackets) for each revisitation pattern.

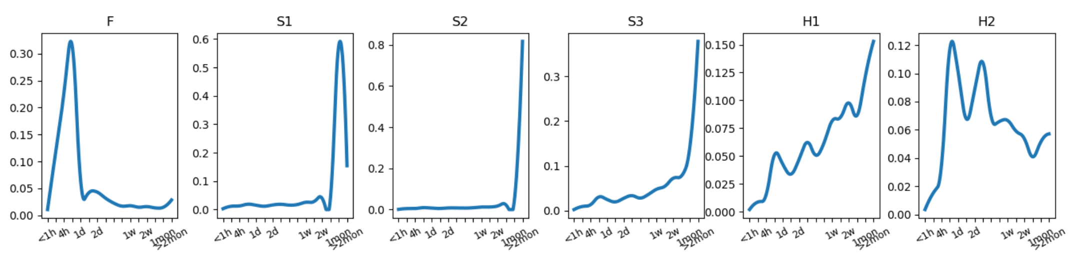


Fig. 7. Centroid user revisitation curves for the Foursquare data.

Fast Revisits. In the Foursquare dataset, one user cluster (F) that contains 2.6% of all users falls into the group of fast revisitation, while in the social localization dataset, 15.1% of the users constitute cluster F with fast revisitation behaviors. The revisitation behaviors of people in these groups are most likely to take place within a day's time. Compared to the average case, POIs concerning travel &transport, residence, food, entertainment, gym, cultural and school are more likely to be visited by these people.

Medium Revisits. In the social localization dataset, three user clusters (M1-M3) that collectively account for 72.5% of all users form the medium revisitation group, whose revisitations are most likely to be spotted daily. A preference of POIs within the categories that relate to daily routines, such as company, institute, industry, bank

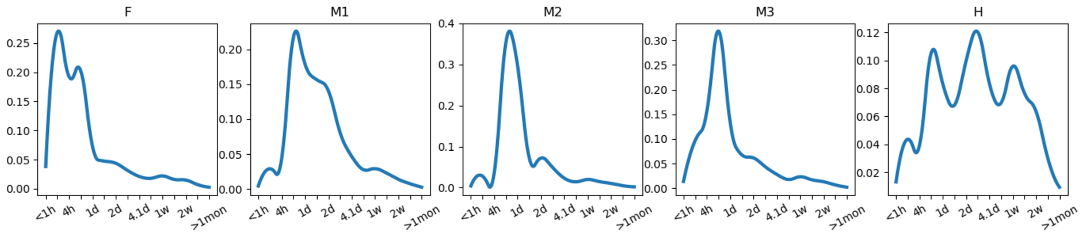
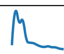
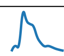
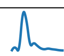
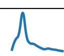
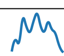


Fig. 8. Centroid user revisitation curves for the social localization data.

Table 6. User Revisitation Cluster Groups for the social localization data.

Description	Label	Curve Shape	Cluster Size	Characteristic POI Categories
Fast Revisits (within half a day)	F		1726 (15.1%)	Food (1.40), entertainment (1.34), gym (1.29), cultural (1.29), school (1.26)
Medium Revisits (around a day)	M1		2512 (21.9%)	Company (1.33), Institute (1.40), industry (1.21), bank (1.26), automobile (1.26)
	M2		2625 (22.9%)	
	M3		3166 (27.7%)	
Hybrid	H		1419 (12.4%)	Company (1.13), life service (1.21)

and automobile was characteristic among the clusters in this group. However, medium revisitation pattern has not been spotted in the Foursquare dataset.

Slow Revisits. In the Foursquare dataset, the feature of slow revisitation pattern is shared by three user clusters (S1, S2 & S3) that hold 23.8% of all users. These users are not likely to revisit a specific POI until one month or more. Users in these clusters are more likely to visit POIs pertaining to categories of arts & entertainment, nightlife spots and food. However, slow revisitation pattern has not been spotted in the social localization dataset.

Hybrid Revisits. In the Foursquare dataset, a vast majority of users (73.7%) constitute the group of hybird, where two clusters of H1 (36.8%) and H2 (36.9%) have been recognized. In the social localization dataset, one cluster (H) with 12.4% of all users form the group of hybrid revisitation. These POIs display combinations of the revisitation patterns aforementioned. With the likelihood to visit POIs pertaining to categories of residence, colleges & universities, shop & service (e.g. life service) and professional & other places (e.g. companies) more, these clusters tend to reflect the routine behaviors of most users.

Joint analysis of POI clusters and user clusters: To examine the relationship between the POI clusters and user clusters, the popularity of different POI clusters for each user cluster was determined and per user cluster visitation distribution was measured both in the Foursquare dataset and the social localization dataset (see Table 7, Table 8). In Table 7 and Table 8, we present the number of check-ins/records within POI cluster i left by users from user cluster j, as well as the ratio of POI check-ins/records proportion within POI cluster i left by users from user cluster j to the global proportion of POI check-ins/records proportion within POI cluster i (the ratio presented in bracket). For instance, in Table 7, the upper-left cell value 259945(2.08) represents that users belonging to fast revisitation pattern checked-in POIs of fast revisitation pattern 259945 times in total, and that the proportion of check-ins at POIs within fast revisitation pattern is 2.08 times of the global average for users belonging to fast revisitation pattern. Significant associations between the POI clusters and user clusters are found by Chi-Square tests both on the Foursquare dataset (χ^2(45)=4034578, p<0.01) and the social localization dataset (χ^2(36)=2373, p<0.01).

We highlighted cells showing strong correlations between POI clusters and user clusters in red, which we determined by judging if the proportion of a POI cluster in one user cluster is significantly greater than the global proportion, i.e., ratio in bracket significantly larger than 1. As is shown in Table 7 and Table 8, we observe that the frequency a POI is revisited positively correlates with the frequency users revisit places. For example, in Foursquare dataset, POIs in POI cluster F are 2.08 times as likely as the average to be visited by users in user cluster F; in social localization dataset, at POIs in POI cluster M3, we have a probability 1.45 times the average to observe visitors from user cluster M2.

Therefore, we conclude that in a similar way as POI revisitation, user revisitation also demonstrates distinct patterns, e.g., slow, medium, fast and hybrid, and is correlated with different POI functions. Moreover, POI revisitation pattern shows strong relevance with user revisitation pattern, which highlights the interplay between urban POI revisitation rhythm and individual periodic visitation pattern.

Table 7. Popularity of POI clusters for each user cluster in the Foursquare data, with significant association highlighted in red.

POI clusters	User clusters					
	F	S1	S2	S3	H1	H2
F	259945(2.08)	10879(0.33)	20016(0.32)	317332(0.32)	1511130(0.54)	5600967(1.52)
M1	8814(0.62)	2216(0.59)	3587(0.50)	65525(0.57)	287148(0.89)	513470(1.22)
M2	4368(0.74)	1764(1.14)	3266(1.11)	51420(1.09)	150999(1.13)	152273(0.88)
M3	3805(0.68)	1837(1.25)	3312(1.18)	54110(1.20)	146131(1.15)	137306(0.83)
M4	3552(0.67)	2002(1.45)	3868(1.47)	56522(1.34)	140859(1.18)	118938(0.77)
S1	2497(0.67)	1628(1.68)	2798(1.51)	41594(1.40)	98830(1.18)	81599(0.75)
S2	4061(0.63)	5493(3.27)	5339(1.67)	79343(1.54)	170263(1.17)	131727(0.70)
S3	8094(0.64)	6516(1.98)	22034(3.51)	213678(2.13)	308587(1.09)	216269(0.59)
S4	63479(0.57)	50688(1.74)	106177(1.91)	1574422(1.77)	2973701(1.18)	2105506(0.64)
H	94817(0.58)	35775(0.84)	55992(0.69)	1175122(0.90)	4451253(1.21)	4270467(0.89)

5.3 Comparisons of Revisitations in Urban Space vs. Online

5.3.1 Similarity. Through our analysis, we find that human physical revisitation shows strong similarity with webpage and app revisitation: both POIs and users fall into distinct pattern clusters, can be roughly categorized into short-term, medium-term, long-term and hybrid, and can be explained by the contents/properties of POIs/sites.

For example, previous work reports that medium revisitations are observed in email and phone communication apps [2], and communication websites, .edu domains, and browser homepages [1]. In our work we find that company and institute locations belong to the medium revisitation group, along with residences, suggesting a semantic similarity across all these types of apps, websites and locations. A comparison of POI types, apps and websites, is shown in Table 9, and semantic similarities are revealed between POIs, apps and websites. From the table, we observe that slow revisitation pattern is closely related to leisure activities, which users do from time to time as a break from regular work. They represent users' long-term interests. Medium revisitation pattern is generally associated with daily routines, such as working, which constitutes the essential part of people's lives. People tend to follow strict schedule for these activities on a regular basis. In the meantime, fast

Table 8. Popularity of POI clusters for each user cluster in the social localization data, with significant association highlighted in red.

POI clusters	User clusters				
	F	M1	M2	M3	H
F1	2412(1.36)	835(0.79)	872(0.78)	2393(0.96)	744(0.87)
F2	1430(1.28)	558(0.84)	494(0.70)	1626(1.03)	472(0.87)
M1	12379(1.04)	6557(0.92)	6741(0.89)	17428(1.04)	5553(0.96)
M2	1502(0.76)	1139(0.97)	1521(1.22)	2998(1.08)	916(0.96)
M3	3898(0.77)	3442(1.10)	4694(1.45)	6457(0.90)	2439(0.99)
M4	3368(0.93)	2700(1.25)	2036(0.89)	4761(0.93)	1907(1.09)
M5	1197(0.95)	871(1.16)	677(0.85)	1786(1.00)	619(1.02)
M6	814(1.10)	416(0.94)	390(0.83)	982(0.94)	440(1.22)
S	381(1.02)	213(0.96)	184(0.78)	548(1.04)	200(1.11)
H	1621(1.13)	701(0.81)	800(0.88)	2015(0.99)	759(1.09)

revisitation pattern is related to shallow exploration which users frequently conduct within a short period of time but normally does not last long, while hybrid revisitation pattern generally happens with activities targeting a variety of audience of different purposes.

5.3.2 Difference. In the meantime, we also notice significant differences between cyber and physical revisitation patterns. First, cyber and physical revisitation have distinct time scales. For cyber revisitation behaviors, the periods generally last a few minutes to a few hours for fast and medium revisitation patterns, while for POI revisitation, the periods for fast and medium are at the day level. Second, physical revisitation is strongly affected by geographic constraints, which is not the case with websites or apps. Previous work considering revisitation patterns on websites and smartphone apps has assumed that there is minimal cost to visiting any particular website or smartphone app. This is not the case when it comes to POI revisitation, since there can be substantial cost to visiting a particular location, both in terms of time and resources. In our analysis we choose to encapsulate these inherent costs by using the metric of physical distance as a proxy for all associated costs: smaller distance to a POI suggests smaller costs associated with visiting that POI, and vice versa.

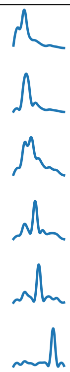
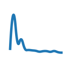
To uncover the relationship between physical distance and revisitation behavior, two heatmaps were generated. In Fig. 9, we visualize how users' revisitation patterns of POIs vary in relation to the physical distance of that POI from the user's (a) home, and (b) work. In these heatmaps, brighter colors indicate a larger number of revisitation records. The distance on the x-axis is measured in kilometers and presented on a logarithmic scale, while the revisitation time intervals were measured in minutes and presented in days for visibility.

The result shows that most revisitation behaviors occur near people's homes and workplaces, with a periodicity of approximately 1 day. We also observe a correlation between distance and periodicity. For places close to home or workplace (e.g., within 1 km), the revisitation intervals are relatively random. As the distance increases, the intervals become more and more unitary and POIs are more and more likely to be revisited after many days.

To investigate the relationship between physical distance and number of revisitations, we constructed the box plots in Fig. 10. Here, the green triangles represent the average of each bar, while the red lines represent the median. The results show that the revisitation number negatively correlates with distance to home and workplace: the larger the distance, the fewer revisitation behaviors are expected to be found.

Thus we conclude that revisitation pattern in urban space and online bear great similarity in terms of revisitation pattern characteristic and drivers. Meanwhile, revisitation in urban space is strongly influenced by geographic constraints, which is not a case in online revisitation behavior.

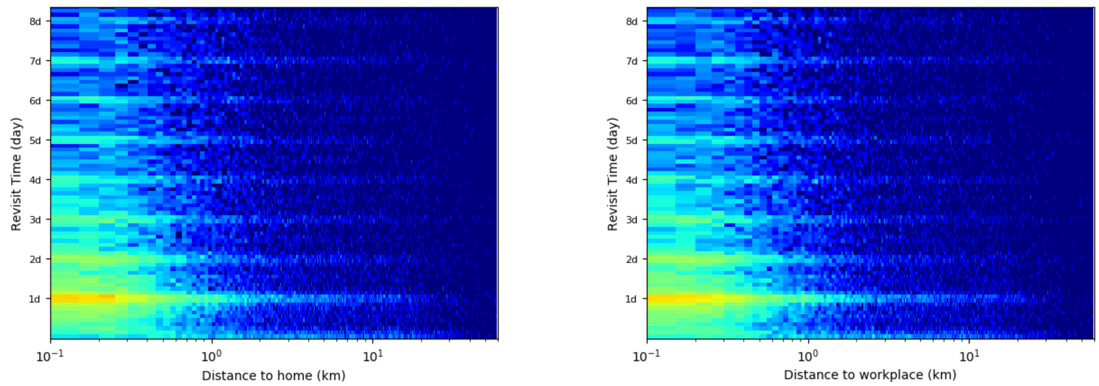
Table 9. Revisitation in urban space vs. online.

Cluster Group	Curve Shape	Description	Corresponding cluster group descriptions from Adar et al. [1]	Corresponding cluster group descriptions from Jones et al. [2]
Fast (F1, F2)		Hotel, transport, cultural, tourist attraction	Hub & Spoke, Shopping & Reference, Auto refresh, Fast monitoring, Pornography & Spam	Instant Messaging, Browser, Social Media
Medium (M1, M2, M3, M4, M5, M6)		Life service, company, institute, residence, industry, office, school, entertainment, restaurant	Popular homepages, communication, .edu domain, browser homepages	Email, Phone Communication
Slow (S)		Entertainment, gym, tourist attraction, shop	Entry pages, Weekend activity, Search engines used for Revisitation, Child-oriented content, Software updates	Utilities, Multimedia, Health and Fitness, Games, Dating, Phone Settings
Hybrid (H)		Food, hospital, shop, cultural, transport	Popular but infrequently used, Entertainment & Hobbies, Combined Fast & Slow	Documents, Notes, Video, Satnav

6 DISCUSSION

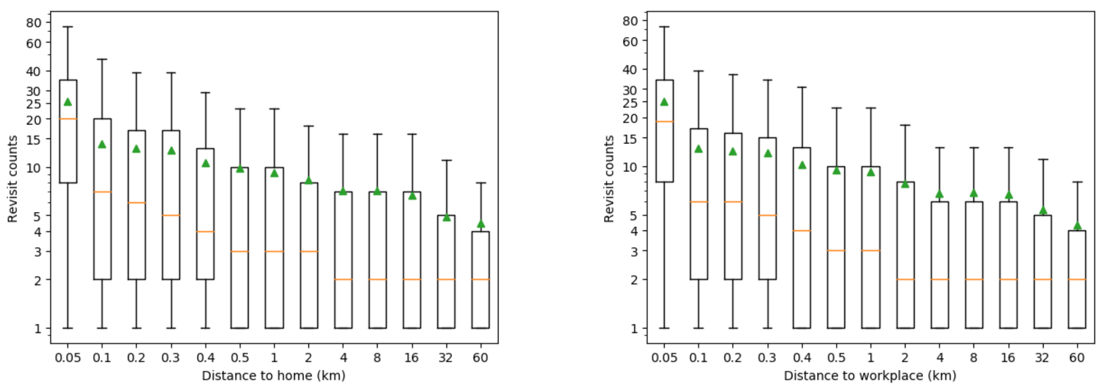
6.1 Contribution to Ubicomp Literature

With the increasing popularity of personal mobile devices and location-based applications, large-scale Point of Interest (POI) data, as well as POI-annotated semantic-rich trajectories of individuals are being recorded and accumulated at a rate faster than ever[41]. Many platforms, some of which are free, provide access to these data, including Google Maps, Foursquare, etc. The wide availability of POI data and semantic-rich spatial temporal data thus offers tremendous opportunity for mobile sensing and ubiquitous computing, as it not only reveals human movement pattern, but also makes possible tracking of user motivation behind mobility and city's social-economic status. In the Ubicomp community, recent work has considered user-POI characterization using POI and semantic-rich spatial temporal data, with promising results. For instance, previous studies have focused on mobility pattern recognition for prediction[19, 42], user profiling[19], POI demand modeling[29, 30], urban problem sensing[43], location recommendation[22, 23], as well as trust and privacy issues concerning POI and location-based services[41, 44–46]. Conversely, in our study, we aim at understanding fundamental pattern behind human mobility, and propose the novel idea of studying revisitation in urban space. We characterize POI



(a) Distance to home (b) Distance to workplace

Fig. 9. Revisit time with respect to distance.



(a) Distance to home (b) Distance to workplace

Fig. 10. Revisit count with respect to distance.

and user revisitation pattern, and identify close link between revisitation pattern and POI function. Our study highlights the interplay between human periodic behaviors and urban rhythms of POI revisitation.

Meanwhile, our study echoes previous discussion and self-criticism in the HCI and Ubicomp community on re-use of research methodologies[47, 48]. In this paper, we leverage revisitation curves, a classical temporal analysis method previously adopted in web and smartphone app revisitation analysis[1, 2] to study physical revisitation pattern in urban space. This enables direct comparison between our findings and previous works. Our analysis demonstrates insightful similarities and differences between online and offline revisitation pattern, which contributes to an understanding of fundamental human behavioral pattern, as well as intrinsic characteristic difference between physical and cyber space. While further attempts should be made to develop new methodologies and tools to support novel research ideas, we call on more emphasis on bringing novelty to established methodologies and findings in our community so as to enable deeper and systematic cognition of important problems.

6.2 Research Implications

Our study contributes to the understanding of periodic visitation pattern in human mobility. While it is widely acknowledged that periodic visit to the same location is one fundamental characteristic of human mobility, little is known as to why and how such periodic revisitation patterns emerge. We present the first study on the role of urban functions and geographic constraints on revisitation pattern.

Secondly, our analysis provides insightful comparisons of revisitation patterns in urban space and online. We find that human physical revisitation shows strong similarities with webpage and app revisitation: both POIs and users fall into distinct pattern clusters, and they can be roughly categorized into short-term, medium-term, long-term and hybrid, which at the same time can be explained by the contents/properties of POIs/sites. While a slow revisitation pattern is closely related to leisure activities, middle and fast revisitation patterns come from daily routines or shallow explorations, as shown in Table 9. Given that the results are obtained from vastly different datasets, we argue that the similarities we identify in this table point to fundamental human behaviors and needs that manifest across different technologies, populations, and timescales. On the other hand, cyber and physical revisitation patterns differ in time scales, and physical revisitation is strongly affected by geographic constraints. For instance, a particular Starbucks store can be a daily destination for a nearby inhabitant, yet rarely visited by someone who lives in a different neighborhood. Through our analysis, we show that the farther the POI is from a person's home/workplace, the less regularly the user revisits that POI. Furthermore, the inter-arrival time for nearby POIs can be quite random while distant POIs are far more likely to follow 'regular' periods with daily granularity. This can be explained by the fact that it takes people increasing time and energy to travel farther, thus they have to specially set their schedules instead of choosing a random time to revisit. This influence of geographic constraints is unique to physical revisitation pattern as opposed to cyber revisitations. Perhaps the most valuable implication of our study for future research is the insight that user behaviors bear great similarities in physical space and cyber space, since they are both driven by user innate needs. While there are lots of works in analyzing user physical behavior and cyber behavior, few try to jointly analyze the two to enable a more global and deeper understanding of user behavior pattern. These findings are beneficial for researchers to think of 'transferable' patterns or methodologies between the two domains.

Finally, our study sheds light on the usability of check-in datasets in studying human mobility. As accurate localization data are hard to obtain, many works on human mobility have turned to publicly available check-in datasets from social media (e.g., Foursquare, Twitter) for analysis [15, 20, 27], while a number of works (e.g., [45]) question their findings since check-in datasets are biased and sometimes even inaccurate. In the case of POI revisitation analysis, we argue that check-in dataset is usable and can obtain good understanding of user real revisitation pattern. In our analysis, Foursquare and social localization datasets both reveal the same categories of revisitation patterns: fast, medium, slow and hybrid. In both datasets, slow and medium revisitations correspond to POIs for leisure activity, such as food and entertainment, while fast revisits, which are on a daily basis, correspond to routine locations such as residence and workplace. The two datasets, however, do produce different results in regard to the relative proportion of each pattern. While almost 75% of POIs in the social localization data show revisitation pattern around 1 day, only 13.3% of POIs in Foursquare dataset belong to this group. Conversely, while almost 25% of Foursquare POIs show monthly revisit periods, only 3% of POIs in the social localization dataset do. We attribute this to the fact that on Foursquare, users tend to check-in at 'unusual' places, such as food and entertainment, rather than routine places such as homes and workplaces, as pointed out by earlier works [49, 50]. Therefore, we are inclined to treat the social localization data as a more complete source for understanding revisitation patterns. Nevertheless, check-in data provides us with a unique lens to study people's exploratory behavior.

6.3 Design Implications

Our work provides valuable insights for the application design. One direct application of revisitation patterns in urban space is automatic personal event reminders. By considering the location visit history of an individual and drawing their revisitation pattern, an app could potentially learn the user's routine visitation behavior and set up automatic event reminder for him/her. For instance, the app could remind a user to go to the gym every other day, go to the market every week and visit certain friends every year etc. without manual input.

Furthermore, our analysis can help design better POI recommendation systems. While there are many existing works on POI recommendation [21], few have taken into consideration user revisitation pattern. One important finding of our analysis is that human revisitation pattern is closely related to the function of the place the user visits: slow revisitation pattern emerges when the user visits 'leisure activity' POIs such as nightlife spots and restaurants, while 'rountine' POIs such as working places and residences generally indicate fast revisitation. For instance, if a user has visited a bar for leisure activities, it is quite unlikely for the user to visit it again soon afterwards since bars show slow revisitation pattern. Therefore, the system should avoid recommending bars and other nightlife spots within a short period of time. On the other hand, if the system learns that the user has not visited bars for nearly a month, the system should recommend the bar he/she prefers to him/her since the user is quite likely to revisit the bar soon. By leveraging user revisitation pattern, the system could recommend the right things at the right time to improve user experience.

6.4 Limitations and Future Work

Our work has a number of limitations. Firstly, the social localization dataset covers a period of 1.5 months, and therefore it is not possible to study revisitation patterns across longer periods. Secondly, although we have tried our best to seek dataset of greatest coverage and duration for analysis, bias still exists in the dataset: while Foursquare dataset records over 400 cities around the globe, most are western cities and tourist cities. Cities in developing regions are not well captured, especially in Asia and Africa. Finally, our study mainly aims at taking the first glance into revisitation pattern in physical space and comparing it with existing works on online revisitation pattern, thus we adopt the classical revisitation curve approach. Due to limit of space, we focus on empirical data analysis in this paper. In the future, we plan to develop better tools and methodologies, add more factors into analysis(e.g., user social demographic factors), and carry out detailed user studies to further strengthen our findings.

7 CONCLUSION

Our paper presents the first analysis to contrast POI revisitation patterns to website and smartphone app revisitation patterns. We identify a number of similarities in how individuals revisit POIs, websites, and apps, but we also highlight important difference. We show that the clusters that emerge for POIs and users match the clusters identified in previous work, both in terms of intermittency and semantics. In terms of the differences, we model how geographic constraints are likely to impact revisitation in urban space, and how revisitation is less likely for POIs that are increasingly far from users' homes or workplaces. Finally, our work points to interesting differences and similarities when considering proactive check-in datasets versus passive localization datasets.

REFERENCES

[1] Eytan Adar, Jaime Teevan, and Susan T. Dumais. Large scale analysis of web revisitation patterns. In *Sigchi Conference on Human Factors in Computing Systems*, pages 1197–1206, 2008.
[2] Simon L. Jones, Denzil Ferreira, Simo Hosio, Jorge Goncalves, and Vassilis Kostakos. Revisitation analysis of smartphone app use. In *ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 1197–1208, 2015.
[3] Marta C Gonzalez, Cesar A Hidalgo, and Albert-Laszlo Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779, 2008.

[4] Shan Jiang, Yingxiang Yang, Siddharth Gupta, Daniele Veneziano, Shounak Athavale, and Marta C González. The timegeo modeling framework for urban mobility without travel surveys. *Proceedings of the National Academy of Sciences*, 113(37):E5370–E5378, 2016.

[5] Philipp Pushnyakov and Gleb Gusev. User profiles based on revisitation times. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 359–360. ACM, 2014.

[6] Jiawei Han, Hong Cheng, Dong Xin, and Xifeng Yan. Frequent pattern mining: current status and future directions. *Data mining and knowledge discovery*, 15(1):55–86, 2007.

[7] Jae Gil Lee, Jiawei Han, and Kyu Young Whang. Trajectory clustering:a partition-and-group framework. In *ACM SIGMOD International Conference on Management of Data*, pages 593–604, 2007.

[8] Nikos Mamoulis, Huiping Cao, George Kollios, Marios Hadjieleftheriou, Yufei Tao, and David W. Cheung. Mining, indexing, and querying historical spatiotemporal data. In *Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, Washington, Usa, August*, pages 236–245, 2004.

[9] Hoyoung Jeung, Lung Yiu Man, and Christian S. Jensen. Trajectory pattern mining. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 330–339, 2007.

[10] Yu Zheng, Lizhu Zhang, Xing Xie, and Wei Ying Ma. Mining interesting locations and travel sequences from gps trajectories. In *International Conference on World Wide Web, WWW 2009, Madrid, Spain, April*, pages 791–800, 2009.

[11] Zhenhui Li, Bolin Ding, Jiawei Han, Roland Kays, and Peter Nye. Mining periodic behaviors for moving objects. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1099–1108. ACM, 2010.

[12] Zhenhui Li, Jingjing Wang, and Jiawei Han. Mining event periodicity from incomplete observations. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 444–452. ACM, 2012.

[13] Zhenhui Li, Jingjing Wang, and Jiawei Han. eperiodicity: Mining event periodicity from incomplete observations. *IEEE Transactions on Knowledge and Data Engineering*, 27(5):1219–1232, 2015.

[14] Zhenhui Li and Jiawei Han. Mining periodicity from dynamic and incomplete spatiotemporal data. In *Data Mining and Knowledge Discovery for Big Data*, pages 41–81. Springer, 2014.

[15] Quan Yuan, Wei Zhang, Chao Zhang, Xinhe Geng, Gao Cong, and Jiawei Han. Pred: Periodic region detection for mobility modeling of social media users. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pages 263–272. ACM, 2017.

[16] Tanvi Jindal, Prasanna Giridhar, Lu-An Tang, Jun Li, and Jiawei Han. Spatiotemporal periodical pattern mining in traffic data. In *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*, page 11. ACM, 2013.

[17] Nathan Eagle and Alex Sandy Pentland. Eigenbehaviors: identifying structure in routine. *Behavioral Ecology and Sociobiology*, 63(7):1057–1066, 2009.

[18] Fengli Xu, Tong Xia, Hancheng Cao, Yong Li, Funing Sun, and Fanchao Meng. Detecting popular temporal modes in population-scale unlabelled trajectory data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(1):46, 2018.

[19] Dingqi Yang, Daqing Zhang, Zhiyong Yu, and Zhiwen Yu. Fine-grained preference-aware location search leveraging crowdsourced digital footprints from lbsns. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pages 479–488. ACM, 2013.

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

[21] Jie Bao, Yu Zheng, David Wilkie, and Mohamed Mokbel. Recommendations in location-based social networks: a survey. *GeoInformatica*, 19(3):525–565, 2015.

[22] Jinfeng Zhuang, Tao Mei, Steven CH Hoi, Ying-Qing Xu, and Shipeng Li. When recommendation meets mobile: contextual and personalized recommendation on the go. In *Proceedings of the 13th international conference on Ubiquitous computing*, pages 153–162. ACM, 2011.

[23] Miao He, Weixi Gu, and Ying Kong. Group recommendation: by mining users' check-in behaviors. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, pages 65–68. ACM, 2017.

[24] Jing Yuan, Yu Zheng, and Xing Xie. Discovering regions of different functions in a city using human mobility and pois. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 186–194. ACM, 2012.

[25] Zipei Fan, Xuan Song, and Ryosuke Shibasaki. Cityspectrum: a non-negative tensor factorization approach. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 213–223. ACM, 2014.

[26] Fabio Miranda, Harish Doraiswamy, Marcos Lage, Kai Zhao, Bruno Gonçalves, Luc Wilson, Mondrian Hsieh, and Cláudio T Silva. Urban pulse: Capturing the rhythm of cities. *IEEE transactions on visualization and computer graphics*, 23(1):791–800, 2017.

[27] Chao Zhang, Keyang Zhang, Quan Yuan, Haoruo Peng, Yu Zheng, Tim Hanratty, Shaowen Wang, and Jiawei Han. Regions, periods, activities: Uncovering urban dynamics via cross-modal representation learning. In *Proceedings of the 26th International Conference on World Wide Web, WWW 2017, Perth, Australia, April 3-7, 2017*, pages 361–370, 2017.

[28] Yanchi Liu, Chuanren Liu, Xinjiang Lu, Mingfei Teng, Hengshu Zhu, and Hui Xiong. Point-of-interest demand modeling with human mobility patterns. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages

947–955. ACM, 2017.

[29] Xinjiang Lu, Zhiwen Yu, Leilei Sun, Chuanren Liu, Hui Xiong, and Chu Guan. Characterizing the life cycle of point of interests using human mobility patterns. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 1052–1063. ACM, 2016.

[30] Xinjiang Lu, Zhiwen Yu, He Du, Fei Yi, and Bin Guo. Discovery of booming and decaying point-of-interest with human mobility data. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, pages 137–140. ACM, 2017.

[31] Linda Tauscher and Saul Greenberg. Revisitation patterns in world wide web navigation. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, pages 399–406. ACM, 1997.

[32] Linda Tauscher and Saul Greenberg. How people revisit web pages: Empirical findings and implications for the design of history systems. *International Journal of Human-Computer Studies*, 47(1):97–137, 1997.

[33] Andy Cockburn and Bruce McKenzie. What do web users do? an empirical analysis of web use. *International Journal of human-computer studies*, 54(6):903–922, 2001.

[34] Hartmut Obendorf, Harald Weinreich, Eelco Herder, and Matthias Mayer. Web page revisitation revisited: implications of a long-term click-stream study of browser usage. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 597–606. ACM, 2007.

[35] Dingqi Yang, Daqing Zhang, and Bingqing Qu. Participatory cultural mapping based on collective behavior in location based social networks. *ACM Transactions on Intelligent Systems and Technology*, 2015. in press.

[36] Dingqi Yang, Daqing Zhang, Longbiao Chen, and Bingqing Qu. Nationtelescope: Monitoring and visualizing large-scale collective behavior in lbsns. *Journal of Network and Computer Applications*, 55:170–180, 2015.

[37] GaWC. The world according to gawc 2016. http://www.lboro.ac.uk/gawc/world2016.html, 2017.

[38] Serdar Çolak, Lauren P Alexander, Bernardo G Alvim, Shomik R Mehndiratta, and Marta C González. Analyzing cell phone location data for urban travel: current methods, limitations, and opportunities. *Transportation Research Record: Journal of the Transportation Research Board*, (2526):126–135, 2015.

[39] Francesco Calabrese, Francisco C Pereira, Giusy Di Lorenzo, Liang Liu, and Carlo Ratti. The geography of taste: analyzing cell-phone mobility and social events. In *International conference on pervasive computing*, pages 22–37. Springer, 2010.

[40] Santi Phithakkitnukoon, Zbigniew Smoreda, and Patrick Olivier. Socio-geography of human mobility: A study using longitudinal mobile phone data. *PloS one*, 7(6):e39253, 2012.

[41] Hancheng Cao, Jie Feng, Yong Li, and Vassilis Kostakos. Uniqueness in the city: Urban morphology and location privacy. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(2):62, 2018.

[42] Rong Du, Zhiwen Yu, Tao Mei, Zhitao Wang, Zhu Wang, and Bin Guo. Predicting activity attendance in event-based social networks: content, context and social influence. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*, pages 425–434. ACM, 2014.

[43] Yu Zheng, Tong Liu, Yilun Wang, Yanmin Zhu, Yanchi Liu, and Eric Chang. Diagnosing new york city's noises with ubiquitous data. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 715–725. ACM, 2014.

[44] Julie Boesen, Jennifer A Rode, and Clara Mancini. The domestic panopticon: location tracking in families. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, pages 65–74. ACM, 2010.

[45] Shion Guha and Stephen B Wicker. Spatial subterfuge: an experience sampling study to predict deceptive location disclosures. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 1131–1135. ACM, 2015.

[46] Dingqi Yang, Daqing Zhang, Bingqing Qu, and Philippe Cudré-Mauroux. Privcheck: privacy-preserving check-in data publishing for personalized location based services. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 545–556. ACM, 2016.

[47] Yong Liu, Jorge Goncalves, Denzil Ferreira, Simo Hosio, and Vassilis Kostakos. Identity crisis of ubicomp?: mapping 15 years of the field's development and paradigm change. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 75–86. ACM, 2014.

[48] Yong Liu, Jorge Goncalves, Denzil Ferreira, Bei Xiao, Simo Hosio, and Vassilis Kostakos. Chi 1994-2013: mapping two decades of intellectual progress through co-word analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3553–3562. ACM, 2014.

[49] Janne Lindqvist, Justin Cranshaw, Jason Wiese, Jason Hong, and John Zimmerman. I'm the mayor of my house: examining why people use foursquare-a social-driven location sharing application. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2409–2418. ACM, 2011.

[50] Sameer Patil, Gregory Norcie, Apu Kapadia, and Adam Lee. Check out where i am!: location-sharing motivations, preferences, and practices. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, pages 1997–2002. ACM, 2012.

Received May 2018; revised August 2018; accepted October 2018