

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

Hancheng Cao^{1,2}, Zhilong Chen², Fengli Xu², Yong Li², Vassilis Kostakos³

Stanford Univ.¹, Tsinghua Univ.², Univ. of Melbourne³

Sep. 11th, 2019

<http://hanchengcao.me/>

<http://fi.ee.tsinghua.edu.cn/~liyong/>

<https://people.eng.unimelb.edu.au/vkostakos/>



In this paper, we present the first large-scale analysis of Point of Interest (POI) revisitation patterns in cities...

Revisitation: The user behavior of returning to (i.e. re-visiting) the same service/location over time.



Every Day



Every Week

Why should we care about urban revisitation?

Motivation 1

Better understand human mobility patterns, which provide insights for designing future location-based services

- Location perspective: **'urban rhythm'**
- User perspective: Understand **user multilevel periodic visitation patterns** for mobility model.
- Previous works: lots of work on user/location visitation pattern, but seldom investigate how people revisit a previously visited place.
- Our work: **first work** to identify urban revisitation patterns and analyze their characteristics!



Every Day



Every Week



?

Motivation 2

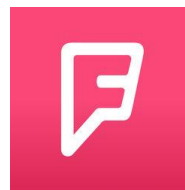
Better understand fundamental behavioral patterns in how people **explore, visit and revisit** services.

- Previous works have investigated online revisitation patterns.
- Do revisitation patterns in **urban space** and **cyber space** share similar properties? What's common and what's different?
- Our work: **first work** to systematically compare revisitation patterns across POIs, websites and smartphone apps.

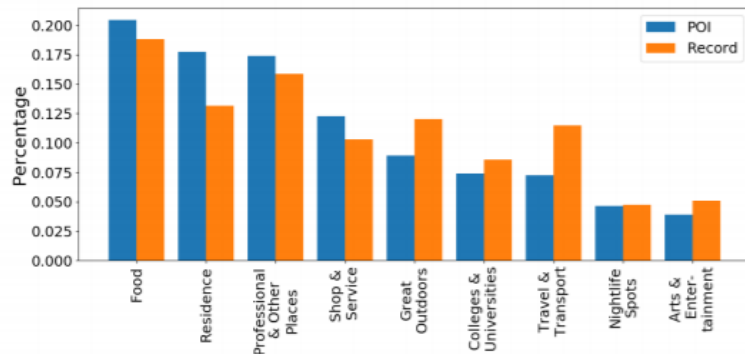


Datasets

Datasets

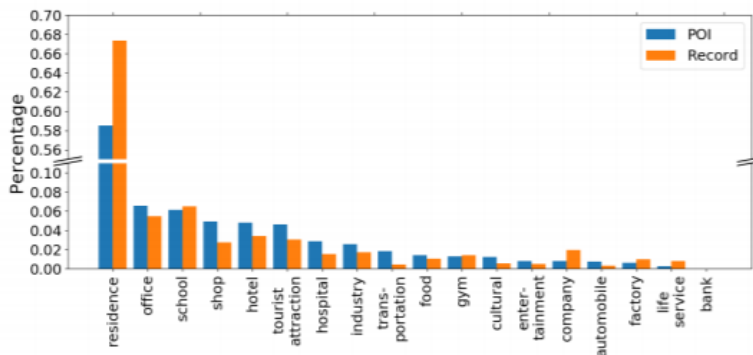


- **Global-scale check-in data from Foursquare:** publicly available, wide popularity, global coverage but 'biased' due to user active check-in
(User ID || POI || Time)



(a) Foursquare

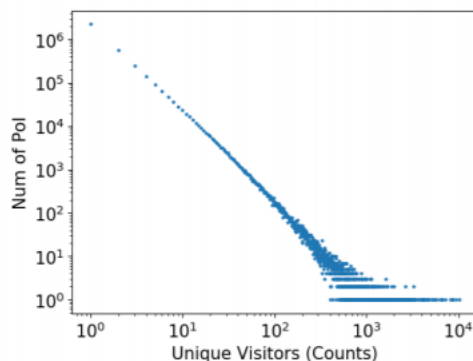
- **Localization data from Tencent in Beijing:** fine-grained, accurate (GPS-enabled), passively record user behavior thus 'objective' but not easily accessible
(User ID || POI || Time)



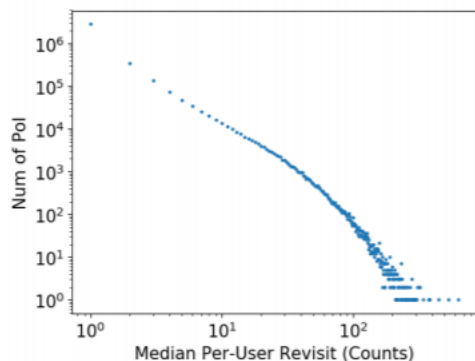
(b) Social localization

Global-scale check-in data from Foursquare

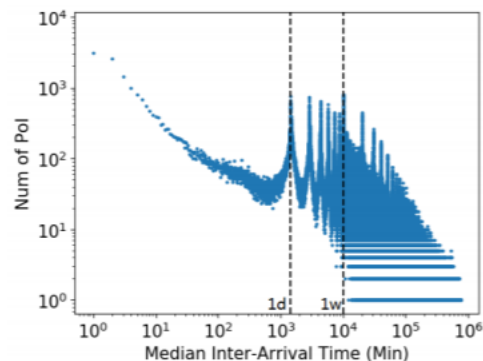
- The dataset includes **33,278,683 check-in records** of **266,909 users** at **3,680,126 unique POIs** (grouped into 9 major categories) between April 2012 and September 2013 (**1.5 years**) in the most checked **415 cities** worldwide.



(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.



Global-scale check-in data from Foursquare

- We used the complete Foursquare datasets to study revisitation patterns in urban space across the world
- We also selected five representative cities: New York, Tokyo, Sydney, Vienna and Rio de Janeiro to study **cultural economic influence** on revisitation based on GaWC.

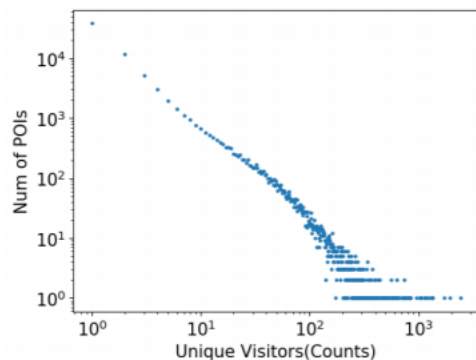
Table 1. Statistics of Foursquare records in 5 selected cities.

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

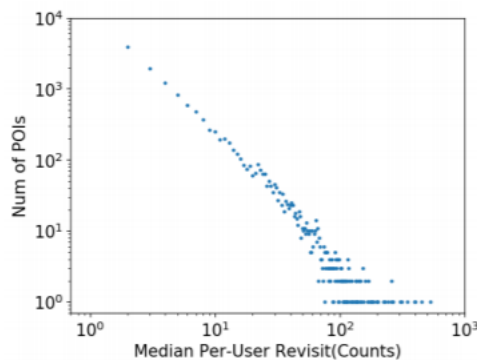


Localization data from Tencent in Beijing

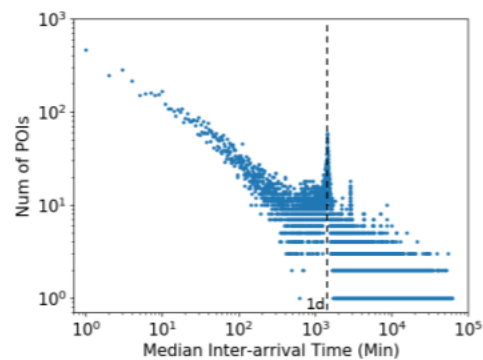
- The dataset includes **3,097,863 localization records** of **15,000 users** at **76,298 unique POIs** (grouped into 18 categories) between Sept. 17th - Oct. 31st, 2016 (**1.5 months**).



(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.



Methodology

Data Preprocessing

- **Data filtering:** stay detection/ remove inactive users/ remove POIs with few records.

After the filtering, 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.

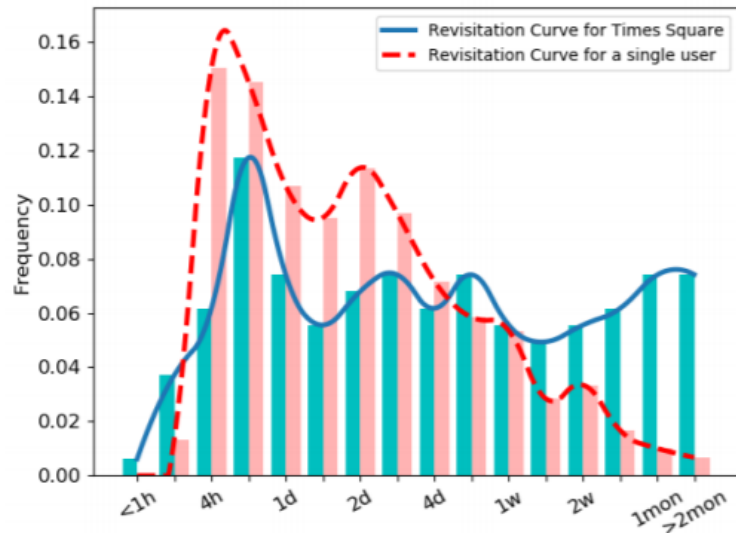
- **Home/workplace Detection:** For every user in Tencent localization data, we label the most recorded POI everyday from 7pm-8am as home, most recorded POI every weekday from 8am-7pm as workplace.

We identified home POI for 12,892 users, and the workplace POI for 13,584 users.



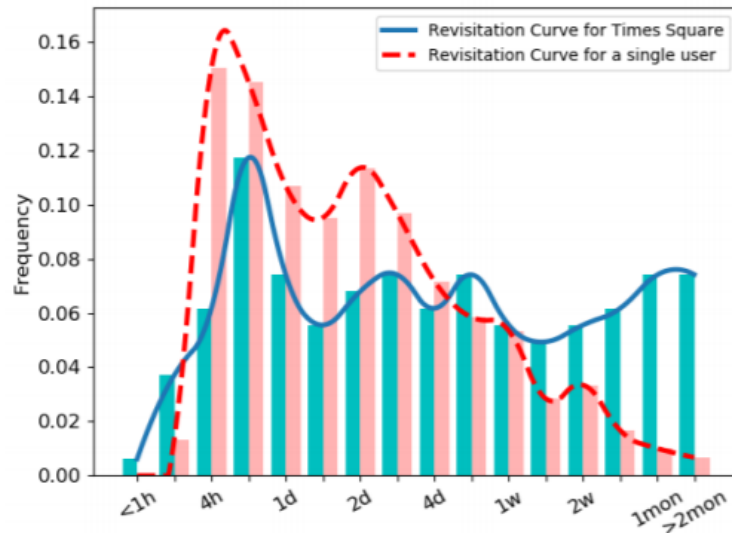
Revisitation Curve Representation

- The revisitation curve represents the number of revisitation within a predefined time interval.
- **POI revisitation curve:** represents how often any given user comes back to a certain POI.
- **User revisitation curve:** represents how often a certain user revisits any given POI.
- We use bins of exponential scale so that observations from the dataset lie approximately evenly in each bin.



Recognizing Revisitation Patterns

- We use K-Means to cluster POIs and users of similar revisitation behavior.
- We ran k-means algorithm by 'Euclidean distance' on multiple k values and constructed an elbow plot to decide the best k value.



Findings

POI Revisitation Patterns

1. Foursquare check-in dataset for 415 cities

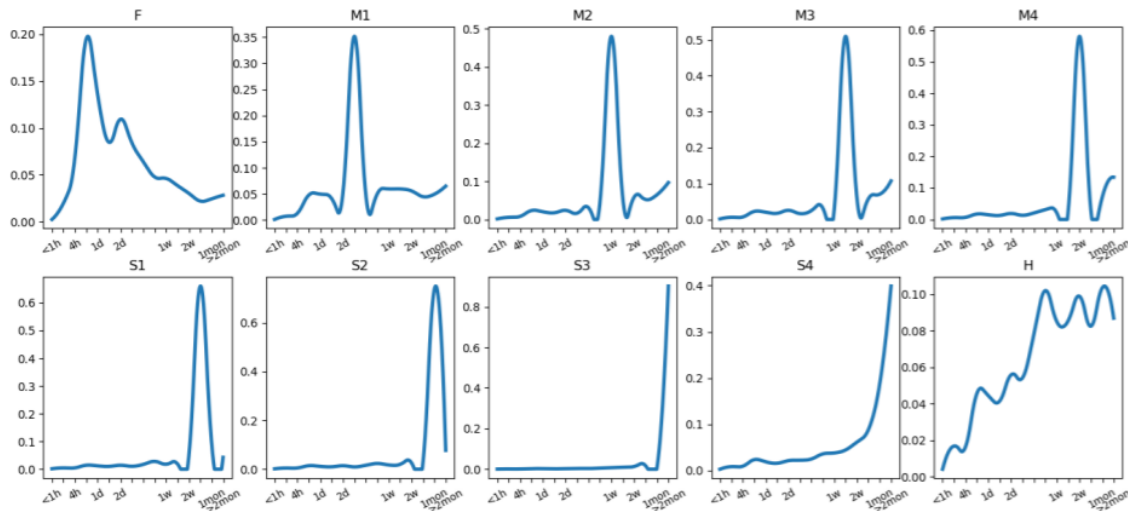


Fig. 5. Centroid POI revisitation curves for Foursquare data.











- Identified 10 clusters
- Can be manually classified into four groups: **slow, medium, fast and hybrid.**



POI Revisitation Patterns

1. Foursquare check-in dataset for 415 cities

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)

- Identified 10 clusters
- Can be manually classified into four groups: **slow, medium, fast and hybrid.**
- **Different pattern corresponds to distinct POI categories!**



POI Revisitation Patterns

2. Tencent localization dataset in Beijing

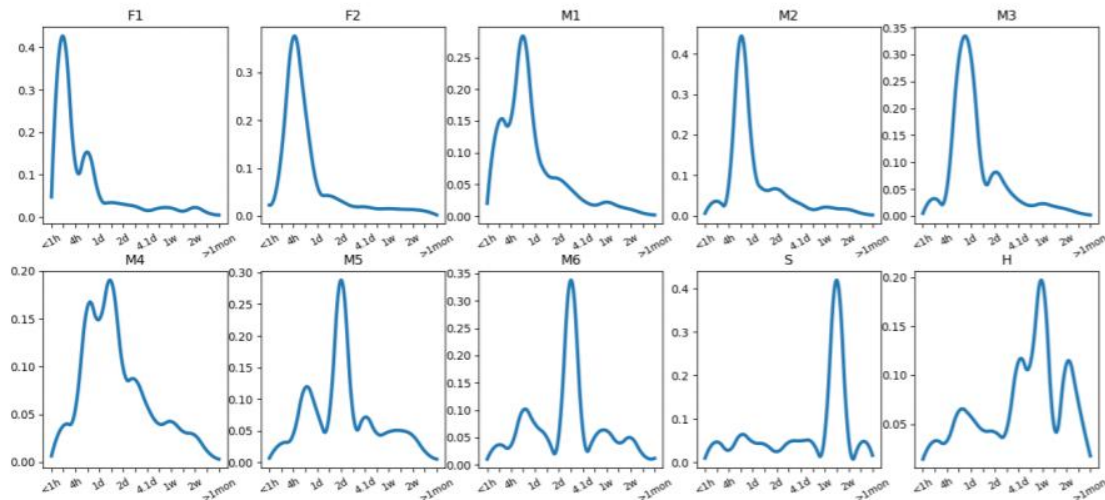












Fig. 6. Centroid POI revisitation curves for the social localization data.

- Identified 10 clusters
- Can be manually classified into four groups: **slow, medium, fast and hybrid.**

POI Revisitation Patterns

2. Tencent localization dataset in Beijing

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)

- Identified 10 clusters
- Can be manually classified into four groups: **slow, medium, fast and hybrid.**
- **Different pattern corresponds to distinct POI categories!**



POI Revisitation Patterns

3. Cross-city Comparisons

Table 4. POI Revisitation Comparison 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, Coffee Shop
Tokyo	10.20% (0.47)	Train Station, Convenience Store, Subway	15.50% (0.86)	Japanese Restaurant, Ramen/Noodle House, Convenience Store	50.90% (1.60)	Ramen/Noodle House, Sake Bar, Mall	23.33% (0.83)	Convenience 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 Building, Bus Line	17.70% (0.98)	Neighborhood, Church, Gas Station	22.25% (0.70)	Bar, Neighborhood, Restaurant	31.99% (1.14)	Church, Gym, School

- Consistent revisitation patterns.
- **Routine -> Fast revisitation, Leisure Activities -> Slow revisitation**
- Developed cities: more POIs demonstrating fast revisitation



User Revisitation Patterns

1. Foursquare check-in dataset for 415 cities

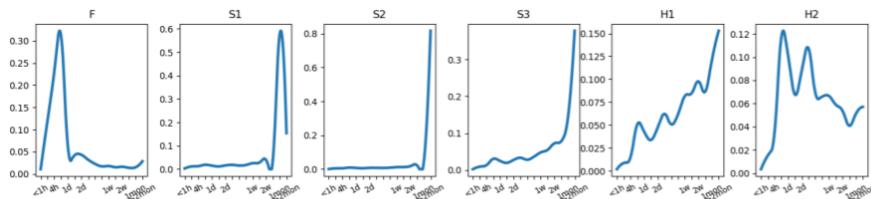


Fig. 7. Centroid user revisitation curves for the Foursquare data.

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%)	

- Identified 6 clusters
- Can be manually classified into four groups: **slow (explorer), medium, fast (routine) and hybrid.**
- **Different pattern corresponds to distinct POI categories!**



User Revisitation Patterns

2. Tencent localization dataset in Beijing

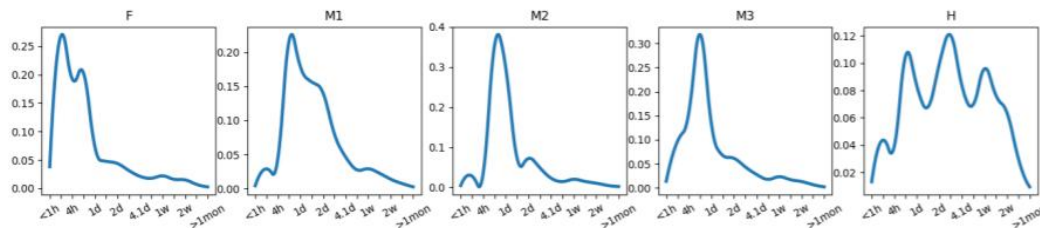


Fig. 8. Centroid user revisitation curves for the social localization data.

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)

- Identified 5 clusters
- Can be manually classified into four groups: **medium, fast and hybrid.**
- **Different pattern corresponds to distinct POI categories!**

POI Revisitation vs. User Revisitation Patterns

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)

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)

- We found **significant associations** between POI revisitation pattern and user revisitation pattern (Chi-Square test)
- User with fast revisitation pattern tends to revisit POIs with fast revisitation pattern.



Comparisons across POIs, Websites and Apps

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

- Revisitation in urban share **strong similarities** with revisitation in cyber space.
- Can all be categorized into short-term, medium-term, long-term and hybrid.
- Can all be explained by characteristics of POIs/sites. (e.g. slow revisitation corresponds to leisure activities)



Comparisons across POIs, Websites and Apps

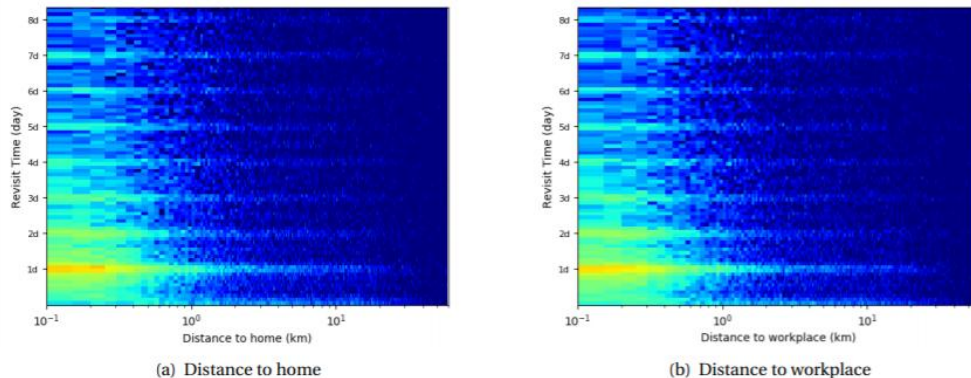


Fig. 9. Revisit time with respect to distance.

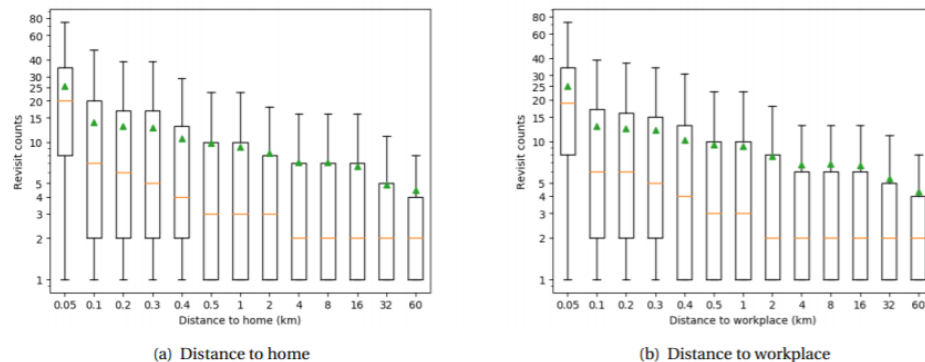


Fig. 10. Revisit count with respect to distance.

- Revisitation in urban space have some **unique properties** compared with revisitation in cyber space.
- Physical revisitation tends to be longer in periods.
- Physical revisitation subjects to **geographic constraints** (distance): the closer the location to home/workplace, more likely it's to be revisited; Farther revisitation typically happens at day level.



Conclusions & Discussions

- Four distinct revisitation patterns for both POIs and users: **fast, medium, slow and hybrid**, which correspond to POI functions. **Consistent** across different cultures, countries and population.
- Similarities between physical (POI) and online (websites and apps) revisitation patterns: slow revisitation corresponds to **leisure activities**, while fast/medium revisitation come from **daily routines or shallow explorations**. Yet physical revisitation differs in that it subject to **geographic constraints**.
- **Design implication:** automatic personal event reminder; better location-based recommender systems that consider revisitation patterns.



Thanks! Questions?

Hancheng Cao^{1,2}, Zhilong Chen², Fengli Xu², Yong Li², Vassilis Kostakos³

Stanford Univ.¹, Tsinghua Univ.², Univ. of Melbourne³

Sep. 11th, 2019

<http://hanchengcao.me/>

<http://fi.ee.tsinghua.edu.cn/~liyong/>

<https://people.eng.unimelb.edu.au/vkostakos/>

