Understanding User Activity Patterns of the Swarm App: A Data-Driven Study

Shihan Lin[1,2,3], **Rong Xie**[1,2,3], **Qinge Xie**[1,2,3], **Hao Zhao**[1,2,3], **Yang Chen**[1,2,3]
[1] School of Computer Science, Fudan University, China
[2] Engineering Research Center of Cyber Security Auditing and Monitoring, Ministry of Education, China
[3] The State Key Laboratory of Integrated Services Networks, Xidian University, China

{shlin15,xieronglucy,17210240050,haozhao14,chenyang}@fudan.edu.cn

Abstract
Location-based social apps have been widely used by people to share their location information with friends. These apps provide rich spatial-temporal information for researchers to investigate user activity patterns. In this work, we collect check-in data from Swarm, and analyze the user behavior in a way of combining spatial and temporal features of check-ins. The results reveal users' different preferences for venue categories in different time of the day. Our work presents activity patterns of human behavior and the distinctions of life habits among three cities, Hong Kong, New York City, and San Francisco. Our findings can be further applied to Swarm's incentive mechanism and recommendation systems.

Author Keywords
Location-Based Services; Check-In; Swarm App; Spatial-Temporal Analysis; City Computing.

ACM Classification Keywords
[Human-centered computing]: Empirical studies in collaborative and social computing.

UbiComp/ISWC'17 Adjunct, September 11-15, 2017, Maui, HI, USA

ACM ISBN 978-1-4503-5190-4/17/09...$15.00
https://doi.org/10.1145/3123024.3123086

Introduction
The past several years have witnessed the rapid development of location based social networks (LBSNs), such as Foursquare [2], Momo and Skout. These LBSNs al-

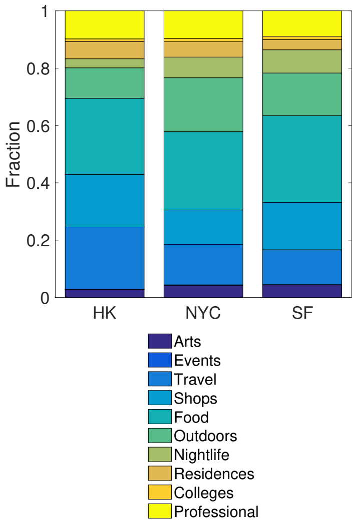

Figure 1: The fraction of check-ins in different venue categories in the three cities.

Time interval	From-To
morning	6:00-10:00
noon	10:00-14:00
afternoon	14:00-18:00
evening	18:00-22:00
night	22:00-2:00
late night	2:00-6:00

Table 1: Time of the day division criteria.

low users to share their location with friends by making check-ins. This kind of aggregated location information has inspired a large number of researchers to investigate human behavior patterns. Cheng et al. explore 22 million check-ins across 220,000 users. They report a quantitative assessment of human mobility patterns [3]. Preoţiuc-Pietro et al. [4] analyze the spatial and temporal patterns of Foursquare user activities by referring to their check-ins.

In this work, we collect the check-in data of Foursquare's Swarm app, a representative LBSN app on mobile platforms. We explore how check-ins of different venue categories are distributed in different time of the day. To the best of our knowledge, this is the first work to investigate the correlation between venue categories and different time of the day, and how the correlation changes according to different cities. The investigation of such correlation helps us know that what the typical activities are performed by people in different time slots within a day. It is an essential step for grasping the characteristics of human behavior, and also the extracted correlation can be applied by LBSN service providers and recommendation systems. Therefore, we consider it important to investigate the correlation.

Overall, our contributions can be summarized as below:

- We measure the correlation between venue categories and the time of the day, and extract the patterns to characterize human check-in behavior.
- We take a step further to study how the human check-in behavior changes over cities. We find out that people from different cities have different behavior patterns while they still have something in common.

Our findings can be referred by Swarm or other LBSN service providers to enhance the user experiences, which will be presented in detail in latter sections. In addition, our findings can be referred by recommendation systems to improve their recommendation accuracy.

Data Collection

In the Swarm app, each user's check-in history is only visible to her friends. We use 10 Swarm accounts to perform the data collection from April 15th, 2017 to May 20th, 2017. We randomly select a number of Swarm users, and send friend requests with the following message to them, i.e., "We are from Fudan University, China. We add friends in order to do our research on user behavior modeling. Your data will be used for research purpose only, and will never be shared with a third party". Finally, 19,483 Swarm users have agreed to participate in our experiment, and we are able to crawl the complete check-in history of them. In this work, we analyze the check-ins generated in three cities, i.e., New York City (NYC), San Francisco (SF) and Hong Kong (HK). This dataset has 1,562,452 check-ins generated by 5,112 users.

The check-in venues are divided into 10 categories[1] as shown in Figure 1. We present the fraction of check-ins in different venue categories in the three cities in Figure 1.

Analysis of Check-ins

Our goal is to understand users' preferences for venue categories in different time of the day when they conduct check-ins. Therefore, as shown in Table 1, we divide 24 hours of a day into 6 time intervals [1]. Besides, we need to measure the relationship strength between venue categories and time intervals. Intuitively, if there is a large proportion of check-ins of a venue category in a time interval, we can infer that people prefer making check-ins in that venue category in that particular time interval. Hence, we

[1] https://developer.foursquare.com/categorytree

weekday / weekend	**morning**	**noon**	**afternoon**	**evening**	**night**	**late night**
Arts	0.002 / 0.001	0.005 / 0.009	0.008 / 0.017	0.020 / 0.016	0.004 / 0.005	0.001 / 0.001
Events	0.000 / 0.000	0.000 / 0.001	0.000 / 0.002	0.000 / 0.000	0.000 / 0.000	0.000 / 0.000
Travel	<u>0.062</u> / 0.031	0.021 / 0.034	0.035 / 0.032	0.030 / 0.027	0.015 / 0.014	0.013 / 0.010
Shops	0.013 / 0.009	0.040 / 0.047	<u>0.056</u> / <u>0.076</u>	0.037 / 0.034	0.003 / 0.003	0.001 / 0.001
Food	0.030 / 0.020	<u>0.093</u> / <u>0.082</u>	<u>0.051</u> / <u>0.076</u>	<u>0.088</u> / <u>0.096</u>	0.016 / 0.020	0.001 / 0.004
Outdoors	0.044 / 0.033	0.027 / <u>0.051</u>	0.035 / 0.046	0.033 / 0.024	0.013 / 0.012	0.014 / 0.007
Nightlife	0.000 / 0.000	0.001 / 0.002	0.005 / 0.012	0.046 / 0.026	0.049 / <u>0.090</u>	0.002 / 0.016
Residences	0.010 / 0.012	0.005 / 0.009	0.008 / 0.008	0.013 / 0.012	0.021 / 0.016	0.005 / 0.006
Colledges	0.004 / 0.000	0.005 / 0.002	0.003 / 0.002	0.001 / 0.001	0.000 / 0.000	0.000 / 0.000
Professional	0.044 / 0.010	0.047 / 0.019	0.025 / 0.013	0.009 / 0.007	0.002 / 0.003	0.002 / 0.003

Table 2: The values of $S_{v,t}$ in weekdays and weekends of the whole dataset.

Definition: We define V as the set of venue categories, and T as the set of time intervals. We define $C_{v,t}$ as the number of check-ins at the venue category v in the time interval t, $v \in V, t \in T$.
Let

$$X_v = \sum_{j \in T} C_{v,j}$$

and

$$Y_t = \sum_{i \in V} C_{i,t}$$

We define the *specificity* $S_{v,t}$ of the pair $(v,t), v \in V, t \in T$ as

$$S_{v,t} = \frac{C_{v,t}}{X_v} \times \frac{C_{v,t}}{Y_t}$$

As the formula above presents, the high value of the $S_{v,t}$ means that the venue category v has strong relationship with the time interval t, and vice versa.

can conclude that the relationship between them is strong. We propose the formula on the left to quantify this relationship. We calculate $S_{v,t}$ for each (venue category, time interval) pair (v,t) in weekdays and weekends separately and Table 2 shows the result. We underline the values larger than 0.05 for reading convenience.

First and foremost, the values of (Food, noon), (Food, afternoon), (Food, evening) are apparently much higher than other pairs, no matter in weekdays or weekends. This is consistent with our intuition that people tend to have meals at noon and in the evening, and delectable food motivates them to inform friends by making check-ins. In addition, nightlife activities are centralized in the evening and at night. Comparing with weekdays, (Nightlife, night) is much higher in weekends. It might be because that people tend to have more fun in weekends. On the contrary, check-ins at work spots, i.e., Professional and Colleges, have stronger relationship with daytime than nighttime in weekdays. Meanwhile, the values of pairs with these spots are dramatically lower in weekends than those in weekdays. Empirically, we know that most people rest in weekends, having fewer check-ins in work spots.

The conclusions drawn above can be applied to Swarm's check-in incentive mechanism, i.e., awarding users making check-ins in venues of a certain category during a certain time interval. For example, Swarm can award users making check-ins in "Food" category at noon and in the evening, and encourage users to make check-ins in colleges or offices in the daytime of weekdays. Since users tend to appear in different venues according to the time of the day, it is convenient for them to make check-ins in venues of a category during the time interval that has a strong relation with that category. Furthermore, they are more willing to do so with awarding. Therefore, our findings can help increase the number of active users and benefit Swarm.

Distinctions Among Cities

Our dataset spans three cities, NYC, SF, and HK, which raises our interest in finding whether people's life habits change over cities. We calculate the *specificity* of (v,t)

	weekday/weekend
HK-NYC	0.886 / 0.913
HK-SF	0.841 / 0.887
NYC-SF	0.934 / 0.923

Table 3: The cosine similarity of each city pair.

pairs in these three cities, respectively. For each city, we also consider both weekdays and weekends, just like in the previous section. We omit to show concrete values in this paper for space saving.

As expected, the rules we find before are also applicable to each of the three cities. However, there are also interesting differences among them. The values of (Shops, afternoon) pair in HK are overall higher than those in another two cities, no matter in weekdays or weekends. The reason may be that the title “Shopping Heaven” of HK attracts quantities of tourists to shop there. Moreover, (Travel, morning) in HK in weekdays is 0.12, which is much higher than that in another two cities. Though the nightlife check-in patterns in the three cities are similar, the values of pairs with nightlife is lower in HK than those in NYC and SF. There is no significant distinction between NYC and SF according to $S_{v,t}$. The reason may be that they are in the same country, i.e., USA, while HK is in China.

To measure the user behavior patterns' similarity in the three cities, we calculate the cosine similarities. All $S_{v,t}$ pairs of each city can be constructed as a 10×6 matrix S, whose elements $S_{i,j}$ represent the *specificity* value of the pair (i, j). For simplicity, we transform all the 10×6 matrices into 60×1 vectors, and regard similarity between cities as the cosine value between their vectors.

As shown in Table 3, the cosine values of city pairs are all larger than 0.8. This means that human beings share similar patterns of behavior in different cities and countries, and this supports the conclusion in the previous section. Furthermore, the cosine values of NYC-SF are higher than those of the other two city pairs. This confirm the point proposed before, i.e., there is no significant distinction between NYC and SF, since they are in the same country.

Conclusion

In this work, we collect the complete check-in data of more than 19 thousand Swarm users, and analyze check-ins created in NYC, SF and HK. Our findings reveal the users' preferences for different venue categories in different time of the day when they make check-ins. Besides, we find both common laws and distinctions of human behaviors among the three cities. For future work, we would like to explore user behavior and validate our findings at a larger geographic scale.

Acknowledgement

This work is supported by National Natural Science Foundation of China (No. 61602122), Natural Science Foundation of Shanghai (No. 16ZR1402200), Shanghai Pujiang Program (No. 16PJ1400700). Yang Chen is the corresponding author.

REFERENCES

1. Emre Celikten, Géraud Le Falher, and Michael Mathioudakis. 2017. Modeling urban behavior by mining geotagged social data. *IEEE Transactions on Big Data* (2017).

2. Yang Chen, Yuxi Yang, Jiyao Hu, and Chenfan Zhuang. Measurement and Analysis of Tips in Foursquare. In *Proc. of IEEE PerCom Workshops, 2016*.

3. Zhiyuan Cheng, James Caverlee, Kyumin Lee, and Daniel Z Sui. Exploring millions of footprints in location sharing services. In *Proc. of AAAI ICWSM, 2011*.

4. Daniel Preoţiuc-Pietro and Trevor Cohn. Mining User Behaviours: A Study of Check-in Patterns in Location Based Social Networks. In *Proc. of ACM WebSci, 2013*.