Understanding User Spatial Behaviors for Location-based Recommendations

Jun Zhang
Pitney Bowes Inc.
jun.zhang@pb.com

Chun-yuen Teng
University of Michigan
chun-yuen.teng@umich.edu

Yan Qu
PlaceNous.com
yan.qu@acm.org

ABSTRACT

In this paper, we introduce a network-based method to study user spatial behaviors based on check-in histories. The results of this study have direct implications for location-based recommendation systems.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine Systems

General Terms

Human Factors, Design

Keywords

Network analysis, location-based services, location-based recommendations, spatial behavior, human mobility

1. INTRODUCTION

T Rapid adoption of mobile phones has driven the growth of location-based recommendation systems. Finding a restaurant on a smart phone is a regular practice for travelers in a new city. In business, mobile marketing techniques like location-based coupons is one of the fastest areas of growth. Novel applications have been developed by researchers such as De Choudhury et al. [5] who proposed constructing travel itineraries with recommended paths from geo-tagged photos and Cranshaw et al. [3] who identified living neighborhoods for visitors using Foursquare check-ins.

To create stronger location-based recommendations, we need deepen our understanding of human mobility and spatial behavior. Excellent work has been done by González et al. [6] who built a model using call logs from millions of users, demonstrating that human mobility has "a high degree of temporal and spatial regularity" and "high predictability". They also demonstrated that the activity of an individual usually center around a small number of frequently visited locations. Cheng et al. [1] analyzed millions of Foursquare check-ins and found that location displacement follows a Levy-Flight like pattern: a mixture of short random movements with occasional long jumps.

These previous works established an excellent theoretical basis for understanding the basic laws governing mobility. However, they focused primarily on extracting patterns from large amounts of data, less attention was paid to user type, location context, and how individual behaviors and motivations shape the data generated. Specifically:

WWW 2013 Companion, May 13–17, 2013, Rio de Janeiro, Brazil.
ACM 978-1-4503-2038-2/13/05.

- **There are different types of users.** Lindqvist et al. [7] and Cramer et al. [4] found that the motivations and preferences of location based services users are very diverse, and their differences significantly impact the way data generated and its potential applications. For instance, a person who checks-in at home and at work will generate different location trails compared to a person who only checks-in when they travel. Thus a location-based recommendation system should consider what types of users they serve.
- **People have different needs based on their location context.** Places checked in by local residents are very different than places checked in by tourists. Also, users look for different kinds of recommendations based on their location context. Understanding these differences as well as finding ways to separating them will allow us to develop better location-based recommendations.

With this in mind, we introduce a network-based approach to analyze location histories. We focus on identifying activities in a location history as well exploring differences among users.

The paper is organized as following: first, we describe check-in data and ways to construct trajectory networks; second, we analyze a known user's trajectory network in detail to illustrate the process of how we analyze a check-in network; third, we analyze trajectory networks of a large group of foursquare users and compare their differences; fourth, we use two examples to illustrate the implications of our methods to the location-based recommendations; we conclude the paper with a short summary and discussion for future work.

2. LOCATION DATA AND TRAJECTORY NETWORK OF MOBILE USERS

2.1 Location Data of Mobile Users

Widespread usage of mobile phones enables us to collect rich location data from location-based services such as Foursquare check-ins, geo-tweets, and location-enabled mobile applications. Although collected and accessed in different ways, generally speaking, this data takes the following format:

[userID]	[check-in time]	[lati.]/[longi.]	[location]
ID1	2012-02-24T12:45:06	52.3648119/-2.37234658	placeA
ID1	2012-02-24T13:34:58	52.36051123/-2.36636901	placeB
ID1	2012-02-26T11:22:45	52.24949444/-2.30175644	placeC

Sometimes device ID is recorded instead of user ID. Here we treated both as user ID. Some dataset may not have detailed location names, but this information can easily be added using a reverse-geocoding process.

In this paper, we used Foursquare check-in data as a test case to illustrate our methods of analysis. However, we believe the same process can be applied to any location history dataset.

2.2 Construct and Analyze a Trajectory Network

We define a trajectory network as a network that reflects the sequential structure of a user's location histories. A node is a location that was visited. An edge is drawn between two locations that were visited consecutively within a time threshold. If consecutive locations were logged within the threshold, we viewed them as a part of the same trip and drew an edge between them. Otherwise these locations were treated as separate trips and no edge was created. Figure 1 show a simple network created using the above sample data. Based on earlier work on check-in displacement time distribution [2] and our own analysis of twenty people's check-in trips, we set 3 hours as the time threshold. Note there is no edge between place B and C because the time gap was more than 3 hours. If a user went from place A to place B multiple times, the weight of the edge reflects this. Thus, the network generated is weighted and directed.

Figure 1: The construction of a trajectory network

2.3 Analyze a Sample Trajectory Network

We used a sample user to illustrate how a trajectory network can be used to analyze user spatial behaviors. Sam, an active Foursquare user, gave us his six months of check-in data, comprising of 1,288 check-ins at 370 unique locations. Figure 2 shows Sam's trajectory network, which has 370 nodes and 425 edges. Note: only 329 nodes are visible because 41 nodes without edges are hidden. To explore spatial aspect of the network, we colored nodes more than 20km away from Sam's home in red (his office is close to his home).

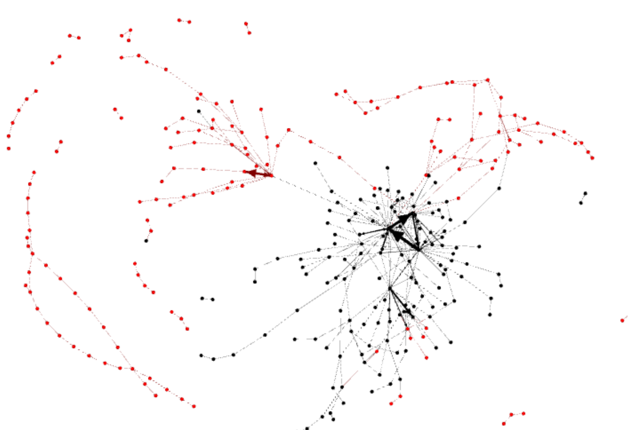


Figure 2: Sam's check-in network

2.3.1 Node Characteristics

We next examined the connectivity of nodes by checking their degrees (i.e. in-degree pluses out-degrees). Figure 3 shows the skewed distribution of the node connectivity. A clear pattern in this network is that there are several nodes with high degrees while most nodes have a degree of 0 or 1 or 2. 11% of nodes have a degree of 0, indicating they are isolated check-ins. 44.3% of nodes have a degree of 1 or 2, indicating they are the only parts of a single trip. 1.6% of nodes have a degree more than 20. We call these high degree nodes location hubs because many trips start or end at them.

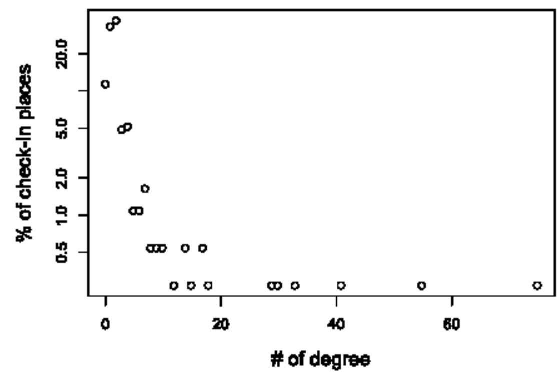


Figure 3: The degree distribution of Sam's check-in network

A detailed examination of Sam's data reveals that the top five location hubs are his home, his office, his favorite gym, a restaurant nearby his office, and a supermarket near his home. All of these location hubs are within 20km of his home. There are two red nodes with degree higher than 8. One is his parents' home and the other one is his hotel during a two week vacation. Identifying location hubs in a trajectory network has commercial implications. For instance, application designers can send different recommendations to hub locations (where routine activities are performed) than to leaf nodes (where casual activities are performed).

2.3.2 Edge Characteristics

Next, we looked at the edge patterns and found that the network can be separated into two parts: nodes in black (within 20km of home) are highly connected, while nodes in red (more than 20km away) form several paths of varying length. In the black sub-network, there is a strong triad cycle connecting the three check-in hubs at Sam's home, office, and gym. This cycle is a good indication of his daily mobility pattern and is highly regular when time of day and day of the week are taken into account. Moreover, many short trips surround these hubs. These trips are related to daily needs like going out for lunch or dinner. There are also several longer paths that correlate with local travel (i.e. a trip to a local mall). This subnet is a good representation of Sam's regular local check-in activities. For red nodes more than 20km away from Sam's home, one noticeable pattern is several isolated long paths that correlate with Sam's travels in the past 6 months. We can also find some medium length paths around two hubs discussed earlier (his parents' home and a vacation hotel). These trip paths are longer than those near his home. Developers looking to identify popular tourist locations should pay special attention to these red paths, while the black part of the network can be used to identify hot spots for locals.

We analyzed the edge weight distribution in Sam's trajectory network and found that about 74% of edges in Sam's network have a weight equal to one, coinciding with trips that only happened once between two connected locations. About 26% of edges are repeated and about 2% are repeated more than 10 times (forming the triadic cycle between Sam's home-office-gym). Edge weight distribution indicates how many of the place visits are repeated activities and how many represent travel that only occurred once.

Above all, we can draw several conclusions about Sam's activities based on his trajectory network. He has location hubs at his home, office, and gym as well as other places he only visited once. There are repeated trips on a regular basis around where he lives as well as unique trips to other areas. Our analysis indicates that it is possible to use simple network metrics like node degree and edge weight to separate these types of trips.

Next, we applied this method to a large number of Foursquare users to explore similarities and differences between other users with Sam.

2.4 Trajectory Networks for Different Users

2.4.1 Dataset

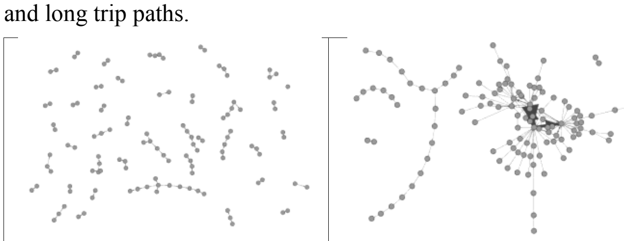
Our dataset was a subset of six months of Foursquare check-ins from Nov. 2011 to May 2012. It was collected using a similar method described in [2] by mining Foursquare-Twitter linked accounts. It is debatable whether Foursquare check-in is preferred when studying spatial behaviors because of varying motivations for using the service [4]. However, Foursquare check-ins is a large dataset with fine grained information that is otherwise not available. We used this data as a test case, limiting our analysis to active users with a high number of check-ins. However, we kept the limitations of this dataset in mind when interpreting our results.

The most interesting question we addressed is whether most users have location hubs in their trajectory networks. We introduced a network metric called Gini Coefficient (GC) to answer the question. GC measures the inequality of degree distribution [10], and in the case of trajectory networks it can be interpreted as the expected difference of degrees between randomly selected places. Assuming the degree distribution of check-ins is sorted in ascending order, we used the same definition as described by Sen [8]:

$$G = \frac{2}{n^2 \bar{x}} \sum_{i=1}^{n} i(x_i - \bar{x})$$

where x_i is the degree of location i, i is the rank of location x, $\bar{x}$ is the mean of the degrees among all locations, and n is the total number of locations. A network with high degree hubs will have a high GC.

Figure 4 shows two similar sized networks with different GCs. User A does not have any location hubs and the trips are mostly disconnected. A detailed review of user A's data reveals that they rarely checked-in when traveling and did not check-in during routine activities such as grocery shopping. User B's network is more similar to Sam's (GCN=0.59) and has several location hubs and long trip paths.



A: GCN=0.19, n=134 B:GCN=0.73, n=136

Figure 4. Two check-in networks with different GCN

Figure 5 shows the distribution of GC of all 38,831 users. We can see that 13.6% users have a GC lower than 0.3, indicating that not all users have location hubs. These users probably only checked-in while traveling. 38.4% of users have a GC higher than 0.5, indicating many users have high degree location hubs. We looked at 1,197 users who checked-in at homes and calculated the distance between their highest degree nodes and home and we that found 82% of the highest-degree nodes are within 3km of home. This indicates that the highest degree node is often a good approximation of home area.

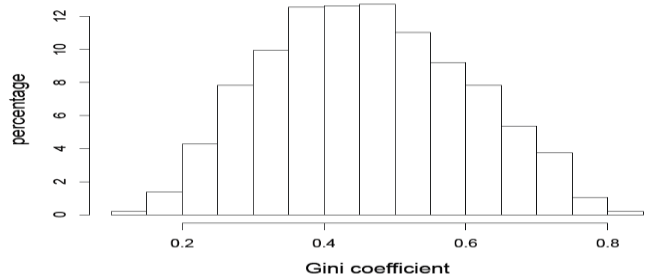


Figure 5: Distribution of Gini Coefficient

Above all, analyzing a large number of active users, we see that there are at least three different types of Foursquare users: those who regularly check-in at local areas and during repeated trips, those who only check-in while traveling and do not record repeated trips, and others who fall somewhere in between. We should consider these type differences into how we identify patterns in check-in data and as we develop location-based services.

3. IMPLICATIONS FOR LOCATION-BASED RECOMMENDATIONS

The network-based method helps us catch important structural and sequential features of people's mobility and gives us a unique perspective in the study of user type and location context. A good understanding of user type and location context has important implications in designing location-based recommendation services. In this section, we illustrate and discuss these implications.

3.1 Recommending Hot-Paths for Locals vs. Tourists

Recommending travel itineraries or hot-paths is one of the popular location-based recommendations [5, 11]. However, previous work process and analyze all the users' data (e.g. geo-tagged photos) together to create the recommendation results. The trajectory network analysis allows us to identify various user types with different check-in behaviors, such as people who check-in for daily activities, or people who only check-in when they are traveling. Understanding such difference in check-in behavior is crucial to correctly interpret user generated location data and use it accordingly. For instance, data from regular check-in users are ideal for studying regularities in human mobility and analysis of local customer behavior, while data from traveler is ideal for tourist places.

We developed a simple application to extract "hot paths" at New York City using people's check-in data. We firstly identified two groups of users from our data: New Yorkers and Non-New Yorkers. To extract these users, we selected all the users whose GC is higher than 0.5, and use their highest degree node as their "homes". Then we selected 100 users whose home is in NYC and 100 users who had checked in at NYC but lives outside of NY, NJ and CT (so they are unlikely to live or work at NYC). We then extracted the most common shared edges within these two groups separately. The results are shown in Table 1.

New Yorker group	Non-New Yorker group
Grand Central – Apple store Penn Station – Madison Square Garden Bryant Park – Grand Central	9/11 Memorial preview site - September 11 Memorial at World Trade Center Staten Island Ferry - Statue of Liberty Empire State Building - Empire State Building 86th floor observatory

Table 1: Top trips in New York City

Clearly, we can see that "hot path" recommendations based on data from these two groups are very different. New Yorkers' hot paths are mostly around local hotspots such as an Apple store and public transportation locations such as the Grand Central, while the non-New Yorkers' are around tourist attractions such as the Statue of Liberty. This simple example shows that treating users' location data differently based on their different check-in behaviors is a promising direction to improve current location based recommendations.

3.2 Location Hub as Recommendation Context

The concept of location hub provides an important cue of location context. A location hub is usually a place where many trips start and end, which implies frequent visits and significant staying time at that location. Such places may be home, office, or a regularly visited place such as a gym. Identifying the spatial location and the place type of the hub is important to understand the activities around the hub. On the one hand, the hubs are frequently visited places and there are usually many activities going on at places close to location hubs. Therefore, the hubs can be used as spatial activity center in mobility analysis. On the other hand, the type of the hub gives important context information for activities before or after the visit to the hub. For instance, the activities on the trips departed from home may be very different from the activities on the trips departed from office. Better description of location context leads to better understanding of what people are doing or what they are going to do, which is valuable information for many location-based recommendation systems. We can imagine a system recommending family eating places near home while recommending coffee shops and lunch places near to the office.

4. SUMMARY AND FUTURE WORK

In this paper, we introduced trajectory network analysis on user generated location data. The network-based method helps us catch important structural and sequential features of people's mobility and gives us a unique perspective in the study of user type and location context, which are important factors in designing location-based recommendation systems.

As an early work in this field, there is still much work to be done. In the future, we plan to improve network construction method by developing more precise means of identifying trips rather than using a simple threshold. We also plan to integrate the trajectory network analysis with two other methods we are developing: spatial patch analysis [9] and trade area analysis [8]. We also plan to further develop the "hot-paths" recommendation system and test it with real users.

5. REFERENCES

[1] Cheng, Z., Caverlee, J., Lee, K., and Sui, D. Z. Exploring Millions of Footprints in Location Sharing Services. In Proc. ICWSM 2011.

[2] Colombo, G.B., Chorley, M. J., Williams, M. J., Allen, S. M., and Whitaker, R. M. You Are Where You Eat: Foursquare Checkins as Indicators of Human Mobility and Behavior. In Proc. Percom workshop 2012, 217-222.

[3] Cranshaw, J., Schwartz, R., Hong, J., and Sadeh, N. M.The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. In Proc. ICWSM 2012.

[4] Cramer, H., Rost, M., and Holmquist, L. E. Performing a check-in: emerging practices, norms and conflicts in location-sharing using foursquare. In Proc. MobileHCI 2011, ACM Press (2011), 57-66.

[5] De Choudhury, M., Feldman, M., Amer-Yahia, S., Golbandi, N., Lempel, R., and Yu, C. Constructing travel itineraries from tagged geo-temporal breadcrumbs. In Proc. WWW 2010, 1083-1084.

[6] González, M. C., Hidalgo, C. A., and Barabási, A. L. Understanding individual human mobility patterns. Nature, vol 435, 779-782.

[7] Lindqvist, J., Cranshaw, J., Wiese, J., Hong, J., and Zimmerman, J. I'm the mayor of my house: examining why people use Foursquare: a social-driven location sharing application. In Proc. CHI 2011, ACM Press (2011), 2409-2418.

[8] Qu, Y. and Zhang, J. 2013. Trade Area Analysis Using User Generated Mobile Location Data. Proceedings of the 22nd International World Wide Web Conference (WWW2013), May 13 - 17, 2013, Rio de Janeiro, Brazil.

[9] Qu, Y. and Zhang, J. 2013. Regularly Visited Patches in Human Mobility. Proceedings of Conference on Human Factors in Computing Systems (CHI2013), April 27- May 2, 2013, Paris, France.

[10] Sen, A. On Economic Inequality. Oxford: Clarendon Press 2011.

[11] Zheng, Y., Zhang, L., Xie, X., and Ma, W. Y. Mining interesting locations and travel sequences from GPS trajectories. In Proc. WWW 2009, 791-800.