Understanding Venue Popularity in Foursquare

Xuelian Long, Lei Jin and James Joshi
School of Information Sciences
University of Pittsburgh
Pittsburgh, PA 15260
Email: {xul10, lej17, jjoshi}@pitt.edu

Abstract—Recently, social media has become an increasingly important part of business and marketing. More and more businesses use social media as part of their marketing platforms. Moreover, the fast development of the 4th generation mobile network and the ubiquity of the advanced mobile devices in which GPS modules are embedded promote the location-based services. Location-based social networks (LBSNs), as the combination of mobile, location-based service and social media, have been changing the way customers interact with the physical location of a business. Foursquare is one of such popular LBSNs in which a user can check in at his current location, leave tips about the venue, explore discounts around his current location, add other people as his friends and so on. These services provide more information to users on where to eat, shop and go for entertainment, as well as a platform to share their activities with their friends. In this paper, we analyze the Foursquare data pertaining to greater Pittsburgh area to investigate several interesting features that could impact venue popularity. By extracting various information in LBSNs, we investigate which are the popular venues, what kind of venues are popular, and what makes them popular. In particular, we study the local hot spots that indicate users' preferences of venues. We also explore if the special offers and web presence help venues become more popular in general. We also analyze trending venues (*i.e.*, very popular venues at a certain time) to investigate the influence of these features on venue popularity over time. Our quantitative analysis could be used to help business owners to design better marketing strategies in LBSNs.

Keywords—Location-based Social Network; Venue Popularity; Foursquare

I. INTRODUCTION

Recently, social media has become an increasingly important part in business and marketing. Nowadays, more and more businesses use social media as part of their marketing platforms. Moreover, the fast development of the high-speed wireless mobile networks, the ubiquity of the advanced mobile devices in which GPS modules are embedded and the powerful interfaces supporting map services such as *Google Maps*, *Microsoft Bing Maps* and *Yahoo! Maps* have greatly promoted the location-based services. Location-Based Social Networks (LBSNs) such as Facebook Place [1] and Foursquare [2], as the combination of mobile, location-based service and social media, have been growing rapidly and attracting a huge number of users. Users choose such LBSNs because of their functionality of both the social networking and location-based services. Users can explore restaurants, museums, popular bars, department stores with specials (e.g. discounts, coupons), *etc.*, around their current locations, and they can also add people in the nearby venues as their friends. These services provide more information about where to eat, shop and go for

entertainment, as well as a platform to share their activities with their friends. Thus, LBSNs have been changing the way customers interact with venues of a businesses and they are ideal marketing platforms for businesses. Business owners can collaborate with LBNSs by creating specials, providing more information to users in order to attract more customers. At the same time, users can search and find venues with good reputation from LBSNs.

Foursquare is one of the most popular LBSNs today. Users can create and add a venue in Foursquare and they can click "Check in" at the venue where they currently are by using a Foursquare mobile App. Users can choose either public "Check in" or private "Check in", and they earn points by checking in. A user can explore venues around him to find interesting ones and view other people who are checked in there. Users can also easily add friends as they usually do in traditional social networks such as Facebook. They can also choose to post their Foursquare check-ins on Facebook or Twitter to broadcast their locations to more friends. They can choose to be notified of their friends' check-ins. Users can get coupons or discounts by checking in, and this motivates users to check in.

Foursquare is very important for both the owner and the customers of a venue. It is expected that 59% of users will search Foursquare for new local businesses [3], thus a business owner will likely miss good opportunities to reach out to customers if his business is not integrated with Foursquare. He can view his venue statistics including the check-ins in a certain period of time (*e.g.*, yesterday, last week or last 30 days, *etc.*), social reach (*i.e.*, how many users checked in post their check-ins to Twitter or Facebook?), the gender and age of the customers, *etc.* [4]. Such information can help a business owner to design special offers or prepare for rush hours. People who check in at a venue also help advertise the venue to his friends (*e.g.*, Twitter followers or Facebook friends) and other people in the same area. On the other hand, customers can claim specials by checking in at the venue if they are currently there. Customers can also earn the Mayorship of a venue by checking in frequently, and they can claim more specials as some specials are only available for the mayor. Besides, Mayorship helps to improve the popularity of the mayor as it is always visible at the associated venue. Moreover, users can explore the interesting places in Foursquare, which is especially helpful during travel. Users can also find out where their friends are and join them.

However, in Foursquare an owner can only get the statistics related to his own business and cannot see the statistics of the other venues and users' check-ins around his venue. Besides, it is still unknown what helps make venues popular.

Such information can help the business owners promote the popularity of their venues. On the other hand, popularity also implies good reputation, thus it can also help users to find good venues. In this paper, we investigate the venues in Foursquare and analyze the user check-ins at these venues in a quantitative way. In particular, we explore and analyze real data related to greater Pittsburgh area obtained by crawling Foursquare from Feb. 23 to Apr. 24, 2012. Our main contributions are as follows:

- By analyzing venue popularity based on the cumulative number of check-ins and cumulative number of unique users who checked in at over 70 thousand venues in grater Pittsburgh area, we obtain a general view on customers' activities related to local venues in LBSNs. To the best of our knowledge, we are the first to use the cumulative data to study the user behavior in LBSNs.

- We investigate the features in LBSNs that could impact the venue popularity in Foursquare. We study the local hot spots that indicate people's preferences of different venues. We first explore if the special offers and web presence would help venues become more popular in general. Then we focus on trending venues (i.e. venues that are popular at a certain time) to investigate the influence of these features on venue popularity over time. Our quantitative analysis could be used to help business owners improve their business strategies.

- In this paper, we quantitatively investigate the local venues of a specific city and the venues' popularity in Foursquare. Moreover, since the geographic distance will limit the user activities more or less and most users' activities will be within a certain geographic region, the analysis of a local city can tell more about user activities and their relationship to local venues than the analysis of general global data.

The rest of the paper is organized as follows: we describe the terminology, our strategy to crawl Foursquare data and the summary of our dataset in Section II; we present the overall popularity analysis in Section III and analyze the popularity over time in Section IV; we discuss some interesting observations based on our analysis in Section V; we review the related work in Section VI and conclude the paper and discuss our future work in Section VII.

II. TERMINOLOGY & DATASET

In this section, we first describe the terminology used in the paper. Then we present the data collection strategy we used to collect the greater Pittsburgh area data in Foursquare and summarize the dataset used in the rest of the sections.

A. Terminology

Venue. A Foursquare venue is a physical location. It can be a place of business office or private residence where Foursquare users can check in.

Venue Category. Foursquare defines a hierarchical list of categories applied to venues. There are 9 top categories in

the hierarchical structure and they are: *Arts & Entertainment*, *College & University*, *Food*, *Professional & Other Places*, *Nightlife Spot*, *Great Outdoors*, *Shop & Service*, *Travel & Transport* and *Residence*. There are some venues in Foursquare which do not have category information and we do not consider these venues in this paper.

Specials. There are eight different types of specials provided by Foursquare and they are: *mayor*, *count*, *frequency*, *regular*, *friends*, *swarm*, *flash* and *other*. They could be specific promotions to get new customers or rewards to most loyal customers [5]. Third parties (e.g. American Express which provides specials like "get $5 off using your American Express card") also can be providers of Specials in Foursquare, but we do not consider them in our paper. The provider of the specials in this paper is Foursquare.

Events. The business owner can create an event on Foursquare [6].

Trending venue. The trending venues in Foursquare are defined as the venues near the user's current location with the most people currently checked in [2]. They are usually the popular venues at a certain time. During our data crawling period we find that a venue in Pittsburgh area is a trending venue if there are at least 5 users currently checked in.

Menu. It is a Foursquare feature that provides the menus in *Food* and *Nightlife Spot* venues to users.

Web Presence (*URL* and *Twitter ID*). In Foursquare, a venue's owner can provide a URL of his website and a Twitter ID that links the owner's tweets in Twitter. Thus, Foursquare users can view the venue's website to get more information about it. They can also tweet messages about the venue or view the tweets about the venue.

B. Data Collection Strategy & Summary

We set Pittsburgh downtown as the center and define a square with sides of around 40 miles. We used Foursquare APIs [7] to discover as many venues as possible in this square. We then obtained 70,390 venues which belong to 9 top categories and 271 second-level categories defined by Foursquare. For each of these venues, we monitored and gathered the following data between Feb. 23 and Apr. 24, 2012:

- Cumulative number of Check-ins: It represents the total number of check-ins made at this venue till a specific date.

- Cumulative number of Users: It represents the total number of users who have ever checked in at this venue till a specific date.

- Specials: When the venue has specials within a period, we collect the information of these specials.

- Menu, URL, Twitter ID: When the venue has any of these resources, we collect the corresponding information.

In order to collect the trending venues we also set Pittsburgh downtown as the center and define a circle with a radius of 25 miles[1]. We then gathered the trending venues per hour in

---

[1] We remove the trending venues that are not in the 70,390 venues.

TABLE I.   SUMMARY OF OUR FOURSQUARE DATASET

# of venues	70,390
# of venues with at least one check-in	66,993
# of trending venue records	10,087
# of distinct trending venues	603
# of specials records	6,501
# of venues with specials	187
# of venues with URL	2,779
# of venues with Twitter ID	2,831
# of venues with menu	1,217

this circle. For a trending venue at a certain time, we obtained the number of users who were currently at the venue.

The summary of our dataset is shown in Table I. For example, there are 66,993 venues with at least one check-in since they were created. The other 3,397 venues have no user checked in at. We get 10,087 records about the trending venues and these records are from only 603 distinct venues. Please note that, in this paper, when we mention venues in Foursquare we mean venues in greater Pittsburgh area in Foursquare, unless mentioned otherwise.

III.   OVERALL POPULARITY ANALYSIS

In this section, we first analyze overall popularity of different venues and different categories by examining the cumulative number of check-ins and cumulative number of users who checked in at these venues. We also investigate if specials, web presence and menu can help improve the popularity of venues.

A. Overall Venue Popularity

We use the following measures to investigate the popularity of a venue: a) cumulative number of check-ins at venue i (CC_i); b) cumulative number of unique users of venue i (CU_i) and c) average user check-in frequency at venue i (UCF_i). We can get CC_i and CU_i from the venue information in our dataset and we calculate UCF_i as:

$$UCF_i = \frac{CC_i}{CU_i} \tag{1}$$

We plot the empirical cumulative distribution function (CDF) of these three measures[2] in Fig. 1 and list the basic statistical features of CC, CU and UCF in Table II. Since the maximum number of cumulative number of check-ins and maximum number of cumulative number of users at a venue are 82,590 and 27,944, respectively, we use log scale in x axis.

In the top part of Fig. 1, we can see that the CC of about 85% of the venues is no more than 100 and the CU of about 96% of the venues is no more than 100. Moreover, about 18% of the venues have only one check-in and about 37% of the venues have only one user who checked in. From Table II, we see that the CC of 50% of the venues is no more than 10 and the CU of 50% of the venues is no more than 2. Therefore, the CC and CU in Foursquare follows the *long tail* distribution [8].

The larger UCF means more check-ins from a user on average. As shown in Table II, the max UCF is 815, which is

---

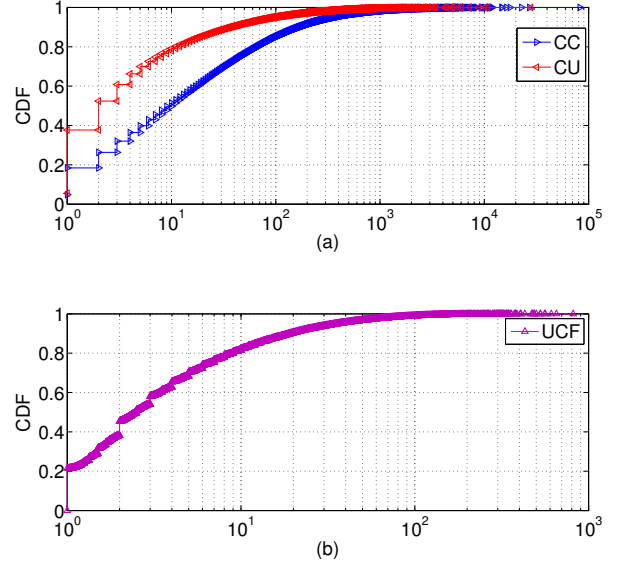[2]We remove the venues with zero users in calculating the UCF.

Fig. 1.   (a) CDF of CC and CU; (b) CDF of UCF

TABLE II.   STATISTICS OF THE CC, CU AND UCF

Min of CC	0
Max of CC	82,590
Mean of CC	94.70
Median of CC	10
Standard Deviation of CC	627.56
Min of CU	0
Max of CU	27,944
Mean of CU	26.87
Median of CU	2
Standard Deviation of CU	218.19
Min of UCF	1
Max of UCF	815
Mean of UCF	8.68
Median of UCF	2.5
Standard Deviation of UCF	21.99

related to a home address and there is only one person checked in at this venue 815 times. In the bottom part of Fig. 1, the UCF of about 83% of the venues is no more than 10, which means the average check-ins per user at these venues is at most 10. 50% venues have an average check-ins per person of only 2.

The UCF can also be taken as the users' loyalty to a venue. Therefore, UCF of the venues in *Food*, *Shop & Service* and *Nightlife Spots* can help the business owner to understand their customers, as well as help customers to compare similar merchants which offer similar types of services. For example, the average customer loyalty in *Food* category in Foursquare is about 3 and the customer loyalty of Tom's bakery house is 5. Based on this, Tom can observe that there are many returning customers at his venue and he could offer specials to these returning customers to thank for their loyalty. Meanwhile, If a customer wants to find a good nearby bakery house, he can compare the total number of customers and the user check-in

TABLE III. TOP 10 VENUES BASED ON CUMULATIVE NUMBER OF CHECK-INS AND CUMULATIVE NUMBER OF USERS

	CC	CU
1	Pittsburgh International Airport	Pittsburgh International Airport
2	CONSOL Energy Center	PNC Park
3	PNC Park	CONSOL Energy Center
4	Ross Park Mall	Heinz Field
5	Rivers Casino	Rivers Casino
6	Heinz Field	Ross Park Mall
7	Robinson Mall	Hofbräuhaus Pittsburgh
8	Giant Eagle Market District	Robinson Mall
9	Lowes Theatre	IKEA
10	Cathedral of Learning	Lowes Theatre

Fig. 3. Distribution of SC and SU in 9 Top Categories

$$SU_c = \sum_{i}^{N_C} CU_i \qquad (3)$$

Here N_C is the total number of the venues in category *C*. We plot the SC and SU in Fig. 3. We find that the distribution of SC and SU are different from the venue distribution shown in Fig. 2. From Fig. 3, we can see that the *Shop & Service* category has the largest number of user check-ins, and the *Food* category has the second largest number of user check-ins. Although we have the highest number of venues in *Residence* category, the number of check-ins in this category is the third largest.

We also find that the distribution of SC_C is different from that of SU_C as shown in Fig. 3. It can be seen that the *Food* category has the most number of users checked in, followed by that of the *Shop & Service* category. An interesting observation here is that the *Residence* category has the least number of users checked in, although it has the largest number of venues. A possible reason is that *Residence* venues are usually private places, thus people just do not check in at these venues.

We define the average user check-in frequency of category *C* as:

$$\overline{UCF_C} = \frac{\sum_{i}^{N_C} UCF_i}{N_C} \qquad (4)$$

The distribution of $\overline{UCF_C}$ is shown in Fig. 4. The *Residence* category has the largest average user check-in frequency followed by that of the *College & University* and *Professional & Others* category. Zang *et al.*'s work also shows that home and work are the top 2 locations of mobile users [9]. Since we use the cumulative data, our results imply the higher probability of check-ins in the venues where users usually stay.

Fig. 2. Venue Distribution in 9 Top Categories

frequency of the venue with that of the average. If both of the parameters of Tom's bakery house are higher than the average values for the similar other venues, then it indicates that Tom's bakery house is better than the average. However, for this to be true, we need to assume that all the check-ins are honest and are from the customers.

We list the name of the top 10 venues with the largest CC and CU in Table III. The top 10 venues with the largest UCF are all *Residence* venues, so we do not list them for the privacy concerns.

B. Overall Category Popularity

Among more than 70 thousand venues in greater Pittsburgh area, we are interested in what categories people usually check in at. We plot the venue distribution in 9 top categories in Fig. 2, in order to present an overview of the venues and categories of our dataset. We can see that *Residence* has the largest number of venues followed by *Professional & Others* and *Shop & Service*.

Based on the measures for venue popularity, the overall popularity of the category C can be measured by the sum of the cumulative number of check-ins of all the venues in the category (SC_C) and the sum of the cumulative number of users (SU_C) of all the venues in the category:
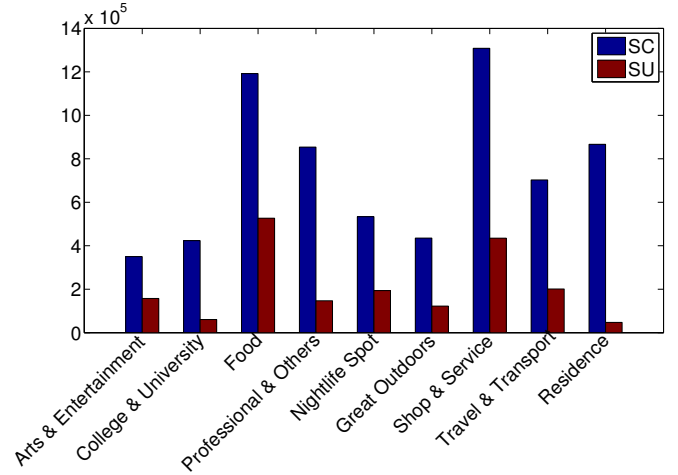
$$SC_c = \sum_{i}^{N_C} CC_i \qquad (2)$$

C. Foursquare Features Promoting Overall Venue Popularity

In this subsection we discuss three features that may help promote overall venue popularity. The following four metrics are used in this subsection:

- Check-ins during a time period (CP) with the first day m and the last day n: $CP = CC_n - CC_m$
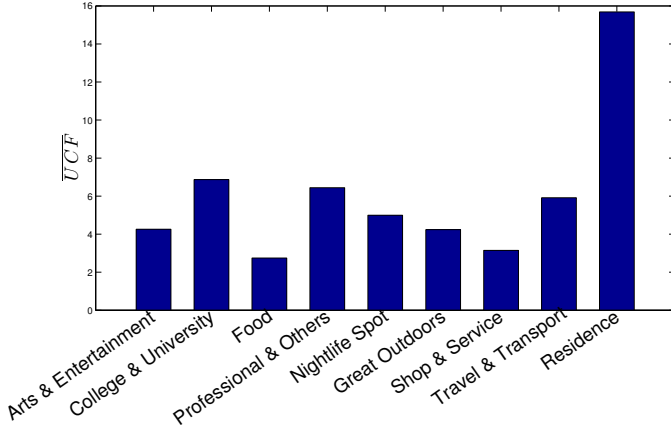
Fig. 4. Distribution of $\overline{UCF}$ in 9 Top Categories

Fig. 5. Average Daily Check-ins

- Number of New User during a time period (NUP) with the first day m and the last day n: $NUP = CU_n - CU_m$

- Average Daily Check-ins in a time period ($\overline{DC}$) with the first day m and the last day n: $\overline{DC} = \frac{CP}{n-m+1}$

- Average Daily New Users in a time period ($\overline{DNU}$) with the first day m and the last day n: $\overline{DNU} = \frac{NUP}{n-m+1}$

1) Specials: Within our data collection period, we have 6,501 specials records. There are 187 distinct venues in these records, thus the average special period per venue is about 35 days ($\frac{6501}{187} \approx 35$). Specials are not always continuous. The venue owner can post specials for just one week or he can post such specials for one month, and they can choose when to post specials. We would like to use the average daily check-ins ($\overline{DC}$) and the average daily new users ($\overline{DNU}$) and compare them to see if there is any difference between the venue's promotion periods and the periods without promotions. If the $\overline{DC}$ and $\overline{DNU}$ during the promotion periods are more than that in periods without promotions, it indicates that the specials may help the venue attracting more customers.

Our methodology is as follows: we first pick all the venues which have specials from Mar. 19 to Mar. 25, 2012 in our dataset[3]. Then, we choose the venues which have no specials lasting for one week from the previously picked venues and finally we get 23 venues in total, as we would like to compare these venues in the same temporal duration. Fig. 5 shows the $\overline{DC}$ of the 23 venues and Fig. 6 shows the $\overline{DNU}$ of 23 venues. In Fig. 5, we see that some venues have significant increase in average daily check-ins with specials. The max difference between average daily check-ins with specials and that of one without specials is 4.16 at venue seven. In some other venues, the specials do not seem to attract more check-ins. In Fig. 6, we can also see that some venues have increase in average daily new users with specials, however the increase is as obvious as that of the average daily check-ins. In some other venues, the specials do not attract more new users.

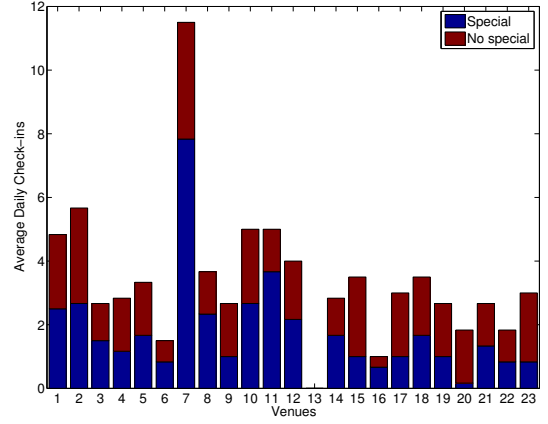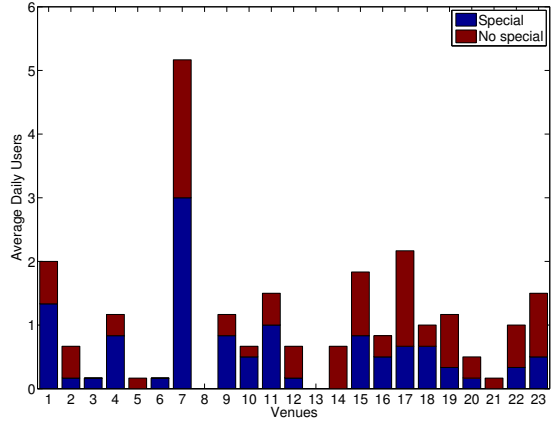
[3]We randomly pick this period as an example.

Fig. 6. Average Daily New Users

We can also see that some venues gain more in terms of both the check-ins and new customers when they offer specials. For example, *Bocktown Beer and Grill* (venue seven) gets both the maximum check-ins and new customers. In *Bocktown Beer and Grill*, the type of the specials provided is frequency and the special message is *Have you gotten 'FRIED' lately? Add an extra topping to an order of our fresh-cut fries today on us!*. Therefore, specials may be one positive factor for improving the popularity of a venue in Foursquare.

2) Web Presence: In this part, we investigate the CP and the NUP related with the venue which has an URL or Twitter ID or both during our data collection period. Table IV summarizes them. We plot the CDF of the CP and the CDF of the NUP of the venues with URL, Twitter ID, both URL and Twitter ID and all venues in Fig. 7 and Fig. 8.

TABLE IV. SUMMARY OF THE CP AND NUP RELATED TO WEB PRESENCE

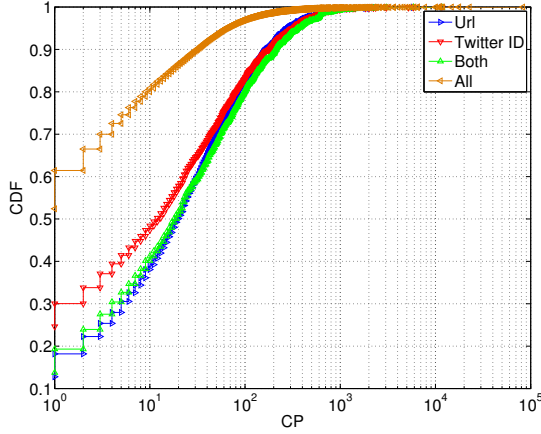|	URL	Twitter ID	both	All
venues	2,779	2,831	1,591	70,390
sum of the CP	205,557	207,312	142,206	1,432,059
sum of the NUP	69,284	64,487	45,634	383,641

Fig. 7. CDF of the CP of the venues with URL, Twitter ID, both URL and Twitter ID and all venues
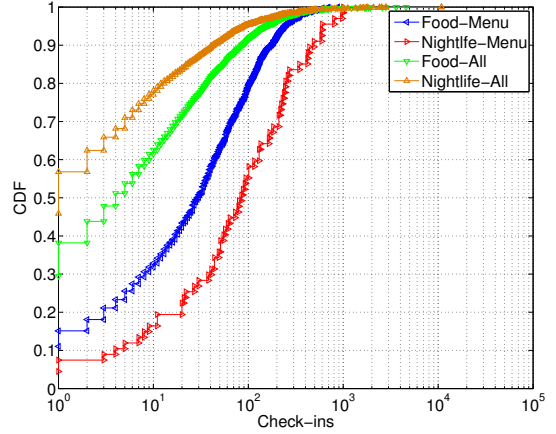

Fig. 9. CDF of the check-ins of the venues with menu and all the venues in *Food* and *Nightlife Spot* categories
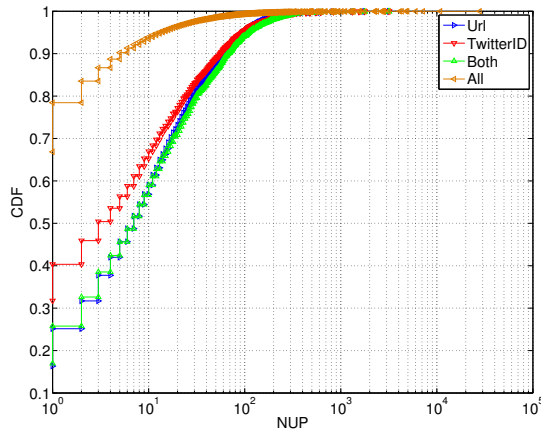

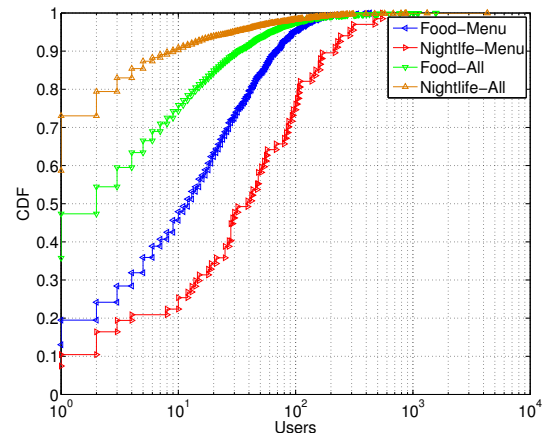Fig. 8. CDF of NUP of the venues with URL, Twitter ID, both URL and Twitter ID and all venues

Fig. 10. CDF of users of the venues with *menul* and all venues in *Food* and *Nightlife Spot* categories

From Fig. 7, we can see that the CP of over 52% of the venues with Twitter ID is more than 10 and the CP of over 60% of the venues with URL or both features is more than 10 in our data collection period. However, the CP of only about 19% of the total venues is more than 10. However, there is no such significant difference for the venues whose CP is more than 1000. Fig. 8 gives similar results about the NUP. NUP of over 32% of the venues with Twitter ID is more than 10; NUP of over 40% of the venues with URL is more than 10; NUP of more than 42% of the venues with both features is more than 10 in our data collection period. However, NUP of only about 7% of the total venues is more than 10. There is still no such significant difference for a venue whose NUP is more than 1000. Thus, venues with good web presence may provide more information about the venue to attract more users in general. Whereas, we observe that good web presence does not help much for the very popular venues.

3) Menu: We explore the CP and NUP of the venues which have menus in these two categories. Table V summarizes the information related to the web *menu* of the venue. We also plot the CDF of the CP and the CDF of the NUP of

the venues with menu and all the venues in *Food* category and *Nightlife Spot* category in Fig. 9 and Fig. 10.

From Fig. 9, the CP of over 67% of the venues with menu in *Food* category is more than 10 and the CP of more than 83% of the venues with menu in *Nightlife Spot* category is more than 10 within our data collection period. However, the CP of only about 37% of the total venues in *Food* category is more than 10 and the CP of only about 22% of the total venues in *Nightlife Spot* category is more than 10. But there is no such significant difference for the venues which have more than 1000 check-ins. In Fig. 10, the NUP of over 52% of the venues with menu in *Food* category is more than 10 and the NUP of more than 75% of the venues with menu in *Nightlife Spot* category is more than 10 within our data collection period. However, the NUP of only about 24% of the total venues in *Food* category is more than 10 and the NUP of only about 9% of the total venues in *Nightlife Spot* category is more than 10. There is still no such significant difference for the venues which have more than 600 new users. Thus, venues with *menus* may provide more information about the venue to attract more new users in general. But menu seems to provide little help to

TABLE V.   SUMMARY OF THE CP AND NUP RELATED TO MENU

	Food	*Nightlife Spot*
venues with menu	1,150	67
all venues	5,921	3,802
sum of the CP in venues with menu	73,739	29,841
sum of the NUP in venues with menu	11,636	5,087
sum of the CP in all venues	101,793	33,164
sum of the NUP in all venues	212,759	88,237

the most popular venues.

IV. POPULARITY OVER TEMPORAL DIMENSION

We first investigate the trending venues based on the spatial-temporal aspect, and then study the influence of the specials on venue popularity over time. After that, we choose the venues in *Food* and *Nightlife Spot* as examples to examine if the attributes such as *URL*, *Twitter ID* and *menu* are helpful in making trending venues. Finally, we explore if the trending venues remain popular over a certain period of time.

A. Spatial-Temporal Analysis of Trending Venues

The geographic locations of the trending venues in our dataset are shown in Fig. 11. We can see that most of the trending venues are around Pittsburgh Downtown area. There are also many trending venues along the main highways or the intersections of the freeways.
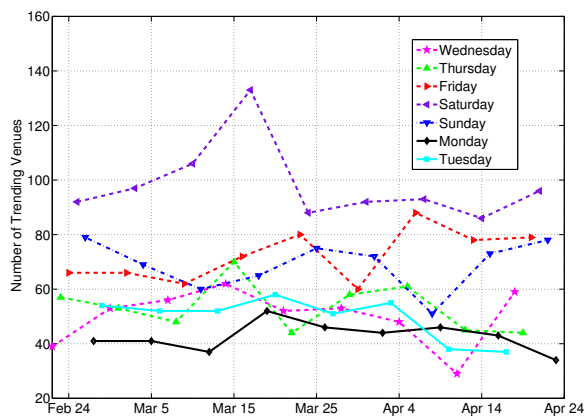

Fig. 12.   Trending Venus VS Time (day)

The distribution of the number of trending venues for each day is shown in Fig. 12. The number of trending venues on Saturdays is always the largest. This is possibly because more people may hang out on Saturdays. We can see that the largest number of trending venues occurred on Mar. 17, which was St. Patrick's Day, probably indicating that more people would like to hang out during holidays. The number of trending venues on Mondays is almost the fewest. There are no significant differences among the numbers of the trending venues on other days. We also summarize the trending venues every hour in Fig. 13. We can see two peaks in Fig. 13. One is at 2pm and the other is at 9pm. We can also see that the top four hours with largest number of trending venues are 9pm, 8pm, 10pm and 7pm. We capture the trending venues every hour and it

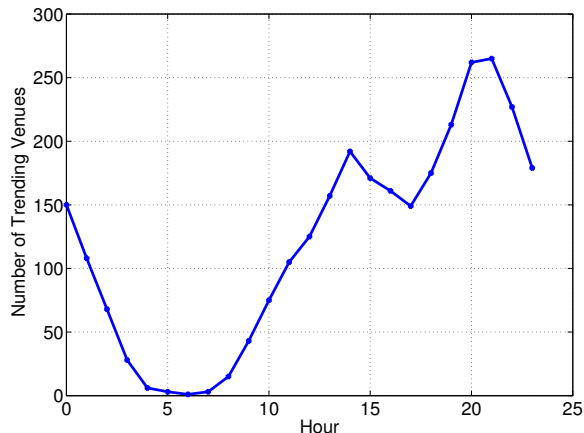indicates that most people usually check in during 1–2pm and 6–10pm.

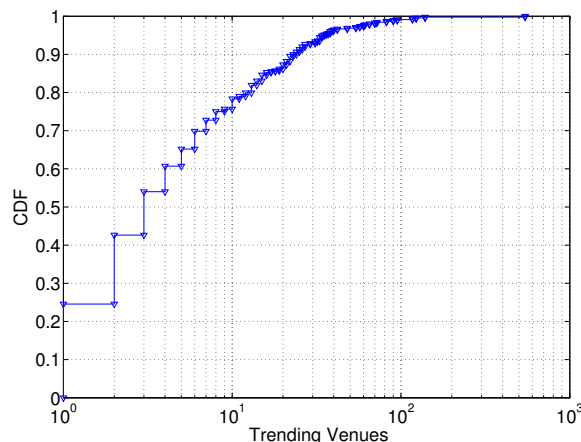Fig. 13.   Trending Venus VS Time (hour)

Fig. 14.   CDF of the trending venue appearance

We also count the number of distinct trending venues and the number of their appearances at each top category and list them in Table VI. The *Food* category has the most number of distinct trending venues and *Nightlife Spot* category venues appear as trending venues most frequently within the data collection period. None of *Residence* category venues appear as a trending venue.

Fig. 14 plots the CDF of the trending venues' frequency of occurrence. It shows that 26% of the trending venues appear only once. About 50% of the trending venues appear no more than three times. However, about 25% of the trending venues appear at least ten times. These trending venues are very popular and they appear on the trending venue list frequently. For example, Pittsburgh International Airport (PIT) appears as a trending venue almost every hour within our data collection period.

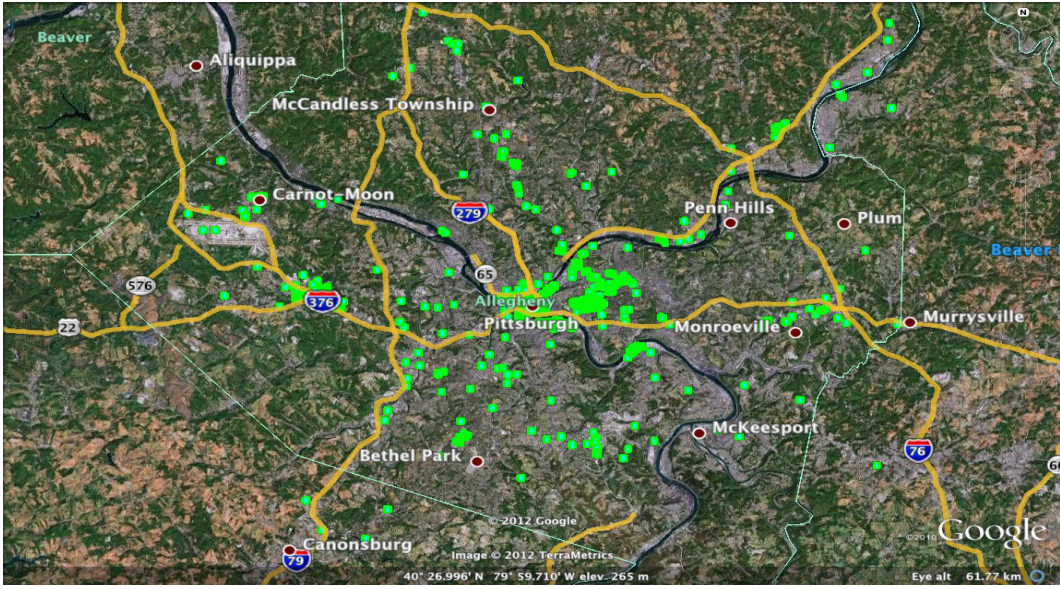### B. Foursquare Features Promoting Trending Venues

Fig. 11. Trending Venus in Pittsburgh Area

TABLE VI.   STATISTICS OF TRENDING VENUES

Category	Venues	Records
Arts & Entertainment	57	1,845
College & University	57	1,243
Food	183	1,135
Great Outdoors	17	94
Nightlife Spot	149	2,103
Professional & Other Places	69	823
Shop & Service	59	1,537
Travel & Transport	12	1,307

TABLE VII.   SUMMARY OF THE VENUES WITH URL, TWITTER ID AND MENU

	Venues	Trending Venues
URL	12.22%	45.18%
$TwitterID$	6.65%	33.13%
$Menu$	12.52%	36.75%

1) Specials: In this subsection, we study the possible features that make the venues become trending venues. We investigate specials and events. During our data collection period, we get 6,501 specials records in our dataset and there are only 406 trending venue records with such specials (6.25%). However, given that there are only 603 trending venues out of more than 70 thousand venues (<1%), we believe specials could be helpful in making venues popular.

2) Web Presence and Menu: We choose the venues in *Food* and *Nightlife Spot* to investigate the impacts of web presence and menu in making trending venues popular, because these two categories have the most number of trending venues in our dataset. We have 9,723 venues and 332 trending venues from these two categories in our dataset. Table VII summarizes the venues and trending venues with web presence and URL in these two categories. In *Food* and Nightlife Spot categories, 12.22% of all venues post URLs and 45.18% of all trending venues post URLs; 6.65% of all venues have Twitter IDs and 33.13% of all trending venues have Twitter IDs; 12.51% of all venues post Menus and 36.75% of the trending venues have Menus. Comparing these ratios, it shows that web presence and menu could be positive factors in improving venue popularity along temporal dimension.

C. Hotness Status

The previous analysis (Fig. 13) shows two trending venue peaks at 2pm and 9pm, therefore we choose these two two-hour periods (*i.e.*, noon–2pm and 7pm-9pm) from Feb. 23 to Apr. 23 to examine if trending venues will keep their hotness status in the next two hours. We chose the trending venues at noon and 7pm as the reference venues and we examined how many of these venues are still trending venues in the following two hours.

We define the hotness persistence rate (HPR) between time k and time t ($t > k$) as:

$$HPR_{k,t} = \frac{||T_t \bigcap T_{t-1} \bigcap ... \bigcap T_k||}{||T_k||} \quad (5)$$

Here $a \bigcap b$ means the intersection set of a and b. T_t is the trending venues at time t and $||a||$ denotes the number of elements in a.

We plot the HPR in Fig. 15. Generally, trending venues will keep their hotness status with high probability. For example, the average rate is 64.04% at 1PM and 68.10% at 8PM. After another hour, the average rate is still 46.41% at 2PM and 45.46% at 9PM. That is, on average there are more than 60% of trending venues which are able to keep their hotness status each subsequent hour and more than 45% of trending venues remain trending venues after two hours. Moreover, there are eight trending venues which remained trending venues in the next hour. Therefore, trending venues are able to keep their hotness status during rush hours with high probability.
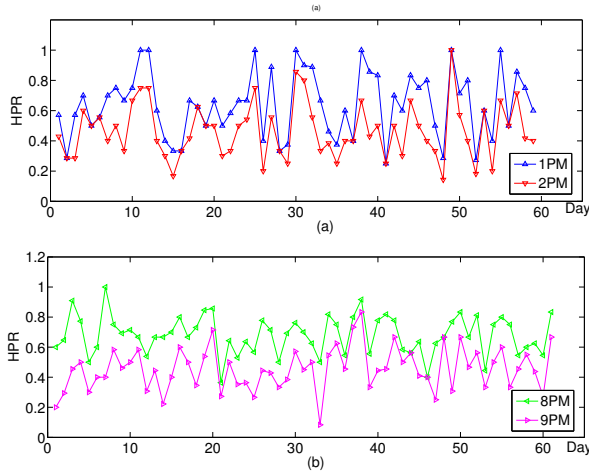
Fig. 15. Hotness Persistence Rates of Trending Venues

V. Discussion

The Always Popular Venue. There is only one venue which is almost always a trending venue in our dataset – Pittsburgh International Airport (PIT). PIT is also the venue with the most number of cumulative check-ins and cumulative users in our dataset. Cheng *et al.*'s work in [10] also show that airports are the most popular venues. We also examined the hometown information of users who checked in at PIT during our crawling period and we found that most users are not from Pittsburgh.

Features Help Make Popular Venues. We examined the features of specials, web presence and menu to see if they could promote overall popularity and popularity at different times of a venue. We find that specials can be helpful in promoting the venue to be a trending venue to some extent. We also observe that web presence can make a venue, whose overall popularity is low, become more popular. However, it has little impact on the very popular venues. Web presence is also a positive factor to make a venue become a trending venue. Lastly, we note that menu has a good contribution for *Food* venues and *Nightlife Spot* venues to promote their overall popularity and popularity over time.

Impact of Events on Venue Popularity. Foursquare allows users to create *events*. It also allows users to associate an *event* with venues. It is possible that an *event* will promote the popularity of a venue during a certain period of time. However, we cannot have a quantitative result on how an *event* can help venues become more popular since we cannot get information about all the *events* through Foursquare APIs. We just manually picked a venue called Console Energy Center, which is the home court of Pittsburgh Penguins (Hockey team). We checked the schedules of Pittsburgh Penguins within our crawling period and we find that Console Energy Center is always a trending venue if there is a home game (*event*). In addition, the venue is the second most popular venue in our trending venue set. Thus, we argue that an *event* may be a positive factor in making venues more popular.

Hotness Status of Trending Venues. Our analysis shows that trending venues are able to keep their hotness status during a certain period of time.

VI. Related Work

Li and Chen's work is the first large-scale quantitative analysis of a real-word LBSN service [11]. They investigate user profiles, update activities, mobility characteristics, social graphs, and attribute correlations. Their work is very general in these aspects and our work is different from theirs in that we focus on venues and users' preferences of the venues.

Cheng *et al.* explore the check-ins to analyze human mobility patterns in the spatial, temporal, social, and textual context [10]. Their work uses the global data collected from Twitter. They do not present the cummulative information and the category information of the venues as we have done. Thus, their analysis about check-ins in spatial-temporal aspect is not as detailed as ours. Moreover, not every Foursquare user has a Twitter account. In our dataset, only 28.44% of the users have Twitter IDs. Thus, many Foursquare users are excluded in their dataset. Their work focus on the human mobility patterns and our work is different from theirs as we focus on the users' preferences of venues.

Gao *et al.* explore social-historical ties in LBSNs by studying the user check-in behavior in [12]. They use both power-law distribution and short-term effect to capture users' historical check-ins. Their experimental results demonstrate how social and historical ties can help location prediction.

Ferrari *et al.*'s work in [13] also uses the local venue dataset (i.e. New York). Their work adopts a probabilistic topic model to identify the human routine behaviors and use this model to extract the urban patterns. Moreover, our previous work [14] also investigate the latent topics through users' check-in data in Foursquare. However, this work doesn't aim to extract the urban patterns or latent topics but presents a more general analysis about the local venues and users' preferences of these venues. We focus more on understanding the reasons that could promote venue popularity.

Ye *et al.*'s work does an initial analysis about Foursquare in [15]. But the dataset in this work is global, not the local one as in our work. Besides, they mainly focus on the recommendation algorithm for LBSNs so they only analyze the common location ratio and the distances among the LBSN users. Another work from them also study the physical distance's influence on the users' check-ins. But their main work is also venue recommendation and their dataset is not local [16]. There are also other work focusing on venue recommendations in LBSNs [17], [18].

Vasconcelos *et al.* 's work investigates the tips, dones and to-dos information in Foursquare [19]. This work still uses a global dataset and it focuses on the user interactions by posting tips and by marking them as done or to-do, which are different from our work.

Noulas *et al.*'s work in [20] also uses a local dataset (i.e. New York) and analyzes the venues according to different categories. However, this work aims to explore the semantic annotations for clustering geographic areas and it still uses the data collected from Twitter. Our work is a quantitative analysis on the venues and check-ins in Foursquare and exploration of the possible features making venues popular

and thus is different from theirs. Another work from the same authors analyzes the spatio-temporal patterns of the check-ins in Foursquare [21]. The check-ins in this work are also collected from Twitter.

Scellato and Mascolo's work measures the user activity in a LBSN [22]. Their work studies how Gowalla users connect with their friends and how users check in at different places. Their work shows that the double Pareto-like distribution can describe the distribution of the number of friends, and the log-normal distribution can describe the numbers of check-ins and places. Their work is mainly based on the number of friends and the number of check-ins of a venue, thus they do not study the venues in as much detail as we do and they do not examine the popular venues.

VII. CONCLUSION AND FUTURE WORK

In this paper, we have quantitatively investigated how location-based social media could impact the local businesses. By analyzing venue popularity based on the cumulative number of check-ins and cumulative number of unique users who checked in at over 70 thousand venues in grater Pittsburgh area, we obtain a general view on customers' activities regarding local venues in LBSNs. We also carefully explore how Foursquare features such as specials, web presence and menu impact the business of local stores, pubs and restaurants. Moreover, we focus on trending venues to study the venue popularity over temporal dimension. We also explore if the trending venues remain popular over a certain period of time. Our results show: 1) menu has a big positive impact on the overall popularity and popularity on temporal dimension of a venue in general; 2) popular venues have good web presence features; 3) Special has a small influence on the venue popularity. To the best of our knowledge, we are the first to quantitatively investigate the local venues of a specific city and the venue popularity in Foursquare. Our findings can be used to better understand the user requirements and help venue owners to design appropriate strategies to attract more customers.

As future work, we plan to analyze the user check-in patterns. Furthermore, we also plan to investigate how the users' check-ins will be impacted by the social relationships (rather than the followers on Twitter) among users.

REFERENCES

[1] Facebook. [Online]. Available: http://touch.facebook.com/

[2] Foursquare. [Online]. Available: https://foursquare.com/

[3] (2011) How important is foursquare to your business? [Online]. Available: http://conversationalmarketinglabs.com/blog/2011/11/social-media-2/how-important-is-foursquare-to-your-business/

[4] (2011) The importance of foursquare for a business. [Online]. Available: http://www.htmlgraphic.com/the-importance-of-foursquare-for-a-business/

[5] What is a special? [Online]. Available: http://support.foursquare.com/entries/195165-What-is-a-special

[6] How to add an event. [Online]. Available: http://support.foursquare.com/entries/22578381-How-To-Add-An-Event

[7] Foursquare api. [Online]. Available: https://developer.foursquare.com/index

[8] Long tail. [Online]. Available: http://en.wikipedia.org/wiki/Long_Tail

[9] H. Zang and J. Bolot, "Anonymization of location data does not work: a large-scale measurement study," in *Proceedings of 17th Annual International Conference on Mobile Computing and Networking (Mobicom)'11*, Las Vegas, USA, Sep. 2011.

[10] Z. Cheng, J. Caverlee, K. Lee, and D. Z. Sui, "Exploring millions of footprints in location sharing services," in *Proceeding of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM)'11*, Barcelona,Spain.

[11] N. Li and G. Chen, "Analysis of a location-based social network," in *Proc. of 2009 International Conference on Computational Science and Engineering (CSE)'09*, vol. 4, Vancouver, Canada, Aug. 2009, pp. 263–270.

[12] H. Gao, J. Tang, and H. Liu, "Exploring social-historical ties on location-based social networks." in *Proceeding of the 6th International AAAI Conference on Weblogs and Social Media (ICWSM)'12*, Dublin, Ireland, Jun. 2012.

[13] L. Ferrari, A. Rosi, M. Mamei, and F. Zambonelli, "Extracting urban patterns from location-based social networks," in *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks (LBSN)'11*, Chicago, USA.

[14] X. Long, L. Jin, and J. Joshi, "Exploring trajectory-driven local geographic topics in foursquare," in *Proceedings of the 14th International Conference on Ubiquitous Computing (Ubicomp'12)*, Pittsburgh, USA, Sep. 2012, pp. 927–934.

[15] M. Ye, P. Yin, and W.-C. Lee, "Location recommendation in location-based social networks," in *Proceedings of the 18th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS)'10*, San Jose, USA.

[16] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee, "Exploiting geographical influence for collaborative point-of-interest recommendation," in *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval (SIGIR)'11*, Beijing, China.

[17] X. Long and J. Joshi, "A HITS-based POI recommendation algorithm for location-based social networks," in *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM'13)*, Niagara Falls, Canada, Aug. 2013.

[18] A. Noulas, S. Scellato, N. Lathia, and C. Mascolo, "A random walk around the city: New venue recommendation in location-based social networks," in *Proceedings of IEEE ASE International Conference on Social Computing (SocialCom'12)*, Amsterdam, The Netherlands, Sep. 2012, pp. 144–153.

[19] M. Vasconcelos, S. Ricci, J. Almeida, F. Benevenuto, and V. Almeida, "Tips, dones and todos: Uncovering user profiles in foursquare," in *Proceedings of the fifth ACM international conference on Web search and data mining (WSDM)'12*, Seattle, USA.

[20] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil, "Exploiting semantic annotations for clustering geographic areas and users in location-based social networks," in *Proceedings of 3rd Workshop Social Mobile Web (SMW'11)*, Barcelona, Spain.

[21] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil, "An empirical study of geographic user activity patterns in foursquare," in *Proceedings of Fifth International AAAI Conference on Weblogs and Social Media (ICWSM'11)*, Barcelona, Spain.

[22] S. Scellato and C. Mascolo, "Measuring user activity on an online location-based social network," in *Proceedings of Third International Workshop on Network Science for Communication Networks (NetSci-Com)'11*, Shanghai, China.