On the Importance of Temporal Dynamics in Modeling Urban Activity

Ke Zhang
University of Pittsburgh
kez11@pitt.edu

Qiuye Jin
University of Pittsburgh
qij3@pitt.edu

Konstantinos Pelechrinis
University of Pittsburgh
kpele@pitt.edu

Theodoros Lappas
Stevens Institute of Technology
tlappas@stevens.edu

ABSTRACT

The vast amount of available spatio-temporal data of human activities and mobility has given raise to the rapidly emerging field of urban computing/informatics. Central to the latter is understanding the dynamics of the activities that take place in an urban area (e.g., a city). This can significantly enhance functionalities such as resource and service allocation within a city. Existing literature has paid a lot of attention on spatial dynamics, with the temporal ones often being neglected and left out. However, this can lead to non-negligible implications. For instance, while two areas can appear to exhibit similar activity when the latter is aggregated in time, they can be significantly different when introducing the temporal dimension. Furthermore, even when considering a specific area X alone, the transitions of the activity that takes place within X are important themselves. Using data from the most prevalent location-based social network (LBSN for short), Foursquare, we analyze the temporal dynamics of activities in New York City and San Francisco. Our results clearly show that considering the temporal dimension provides us with a different and more detailed description of urban dynamics. We envision this study to lead to more careful and detailed consideration of the temporal dynamics when analyzing urban activities.

Categories and Subject Descriptors

H.2.8 [**Database Applications**]: Spatial databases and GIS; H.2.8 [**Database Applications**]: Data mining; H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Urban Data Analytics, Urban Computing

Keywords

Urban Activity, Temporal Dynamics, Location-based Social Networks

UrbComp'13, August 11-14, 2013, Chicago, Illinois, USA.

1. INTRODUCTION

The proliferation of location-based services has provided us with an abundant source of data that can serve studies in a number of different disciplines. Typically the information encoded from these systems includes (i) social connections among the users, as well as, (ii) spatial trails of the users. The latter is a timestamped list of locations that the user has visited. Often a LBSN does not continually track locations but relies on users to voluntarily share their whereabouts through *check-ins*. Locations can be either a tuple of latitude-longitude or the actual venue that was visited (or both). Knowledge of the actual place that users have checked-in provides additional semantic information that is crucial in understanding the nature and dynamics of the activity that takes place in a geographically constrainted area. Based on the *type* of check-ins within an area, that is the type of places people visit, we can obtain the area's activity profile. In this paper, we will utilize a dataset from Foursquare, the major LBSN to date, in order to understand (a) the temporal transitions of activity profiles for a given area, as well as, (b) the temporal dynamics of clusters of geographic areas based on their activity profiles.

Having a model that captures the spatio-temporal dynamics of the activity that takes place in an area is of extreme importance for the city management. While there has been a large volume of studies on the statistical properties of urban mobility (e.g., probability of displacement etc.) utilizing cell phone data - e.g., [14, 17, 5, 2, 3, 8, 10] to name only a few - these models, while accurate at describing the actual patterns, cannot capture the reasons for why these patterns emerge in the first place. Knowledge of the latter can facilitate urban planners into understanding how people act and interact in the city. For instance, putting unexpected road congestion instances into the context of urban activities can provide intuition on their causes and possible solutions (e.g., advices on how to change the city layout and the contextual structure of neighborhoods). Hence, it is crucial to place human mobility into the context of activities that people engage in a city and develop models for the spatio-temporal dynamics of urban activity. To reiterate, in this work we seek to emphasize on the importance of the temporal dimension for modeling urban activity.

Recent studies have analyzed similar data sources to identify dynamics of both users and areas activity. For instance, users' check-in properties and mobility patterns have been studied in [4, 13, 14, 17, 5]. More recent research, such as [16, 7, 12, 18, 1], combines different sources of data (e.g., GPS data, cell-phone data, check-in data etc.) in order to study urban dynamics. More closely related to our preliminary study, Noulas *et al.* [15] have examined geographical clustering based on the aggregated activities in an area, while Cranshaw *et al.* [6] *redefine* the notion of a neighborhood by considering the dynamics of the activities that take place in a city. However, the temporal dynamics are absent. In other words, the activity that takes place within a geographic area is considered in aggregate, with no temporal distinction. While this simplifies the data processing, it essentially assumes that the cumulative activity profile of an area is constant during the course of a day or even across days. This leads to an aggregate/average only view of the spatial dynamics and as we demonstrate in this paper a more detailed and realistic view of urban dynamics can be obtained by incorporating the time dimension. To the best of our knowledge, Yuan *et al.* [18] are the only ones to have implicitly incorporate the time dimension in their analysis of functional regions of a city[1]. In particular, they consider in their model the arrival and departure times of people to/from an area.

The rest of the paper is organized as follows. Section 2 describes our urban activity model. Section 3 presents our temporal analysis results, while Section 4 concludes our work and discusses future directions.

2. DATASET AND ANALYSIS SETUP

Dataset: To perform our study we utilize a dataset collected by Cheng *et al.* [4] that includes geo-tagged user generated content from a variety of social media. More specifically, this dataset includes location information that was pushed from a variety of application to Twitter's public feed between September 2010 and January 2011. Each tweet includes location information in the following format: `<userID, tweetID, text, location, time, venue ID>`. There are 22,506,721 tweets in total. From those we initially filter out tweets that have not originated from Foursquare and this provides us with a dataset of 11,726,632 Foursquare check-ins pushed to Twitter.

Foursquare associates with each venue v a category $c(v)$ (e.g., restaurant, school etc.). This classification is hierarchical, in the sense that an Italian restaurant belongs to the category "Italian restaurant", which can belong to the higher level category "Restaurants", which can itself belong to the category "Food" and so on. At the top level of the hierarchy there are 9 categories, namely, *Arts & Entertainment*, *College & University*, *Food*, *Nightlife Spots*, *Outdoors & Recreation*, *Professional & Other Places*, *Residences*, *Shops & Services* and *Travel & Transport*. Our original dataset did not include the category information for the venues, so we have crawled Foursquare.com and obtained the required mapping.

Analysis setup: In our analysis we focus on two cities, New York City (NYC) and San Francisco (SF)[2]. In particular, we consider all the check-ins in our dataset that took place in a rectangle area of 10 miles^2, centered at the city centers. This corresponds to 277,503 and 82,435 check-ins respectively. We further divide these *city-wide areas* in a grid of 400, equally-sized, *neighborhood areas* (rectangles of 0.5 miles^2 each). The numbering of the areas begins at the bottom left grid point, i.e., area 1, and exactly above this point is area 21. The last area 400 is at the top right grid point.

Similar to [15], each one of these neighborhood areas n, can be associated with an activity profile vector $\overrightarrow{\alpha_n}$, based on the *type* of check-ins that take place within the "neighborhood". Since there are 9 (top level) categories in Foursquare, $\overrightarrow{\alpha_n} \in \Re^9$ and its i^{th} element is:

$$\alpha_{n,i} = \frac{\sum\limits_{v \in n \bigwedge c(v)=i} \#check-ins(v)}{\sum\limits_{v \in n} \#check-ins(v)} \quad (1)$$

Simply put, each element i of the activity profile vector of an area n is essentially the fraction of all the check-ins in n that belong to category i. Note here that, $\overrightarrow{\alpha_n}$ discards information related with the actual volume of activity within the area. However, it captures the *behavior* of the area with regards to the type of check-ins that take place within it. Furthermore, the above activity vector does not include any timing information and considers all the check-ins in aggregate. Hence, by grouping the check-ins based on the time period they took place we can have a time-dependant activity profile vector $\overrightarrow{\alpha_n}(t)$. In our study we consider three coarse-grain time periods, that is morning (4am-12pm; t_m), afternoon (12pm-8pm; t_a), evening (rest; t_e), and hence each area n is associated with three temporal vectors, $\overrightarrow{\alpha_n}(t_m), \overrightarrow{\alpha_n}(t_a), \overrightarrow{\alpha_n}(t_e)$. While we acknowledge that a different time division would probably provide different absolute results, we would like to emphasize on the fact that our goal is to demonstrate the importance of temporal dynamics[3]. In this sense, we pick the above time division, since it represents a fair division in equal time periods, with noon separating morning and afternoon.

3. TEMPORAL ANALYSIS OF AREA ACTIVITY

In this section we will examine various aspects of the temporal dynamics in urban activity, considering both each neighborhood area in isolation, as well as in conjunction with the other parts of the city. In particular, we begin by examining how the volume and the type of activity within an area change across the three coarse-grain time periods we consider. Then using $\overrightarrow{\alpha_n}$ as the feature vector of every area we identify clusters of neighborhoods with similar activity and we study the way that the clusters change over time.

Temporal changes in activity volume: For each one of the neighborhood areas on the grid in our analysis, we first calculate the total number of check-ins in this area. Figure 1 depicts the geographical distribution of the aggregate check-ins in NYC, where the white squares correspond to areas with less than 30 check-ins. The darker the area, the more check-ins have been generated there. As we can see most of the check-ins are in Manhattan, and especially midtown (the area around Time Square has the maximum number of check-ins, 13,927) and downtown areas.

We further examine the variation of the activity volume within areas and across time as captured through the number of check-ins. In particular, we follow exactly the same process as above, but now instead of considering all the check-ins in aggregate, we only consider check-ins that took place in each of the three defined time-periods. As we will see, even this simple aggregate statistic, varies significantly across time, supporting the importance of time in characterizing the *urban pulse*. Figure 2 presents a heatmap for the distribution of the check-in volume in NYC during the three time periods in our setting. As we can see, there are some differences on which areas concentrate the majority of the activity over the three different time-periods. In other words, different parts of the city appear to be more popular over different times. Furthermore, the absolute values of the number of check-ins show that

[1]Of course there exist other studies that have included temporal dynamics in various granularities (e.g., [17] [5]), but not in the context we are examining, that is, urban activity modeling.

[2]For illustrative purposes we will also visualize some of the results, mainly for NYC due to space limitations, on a map.

[3]We further recognize that weekends and weekdays should also be treated separately, since the activity dynamics are expected to be different.

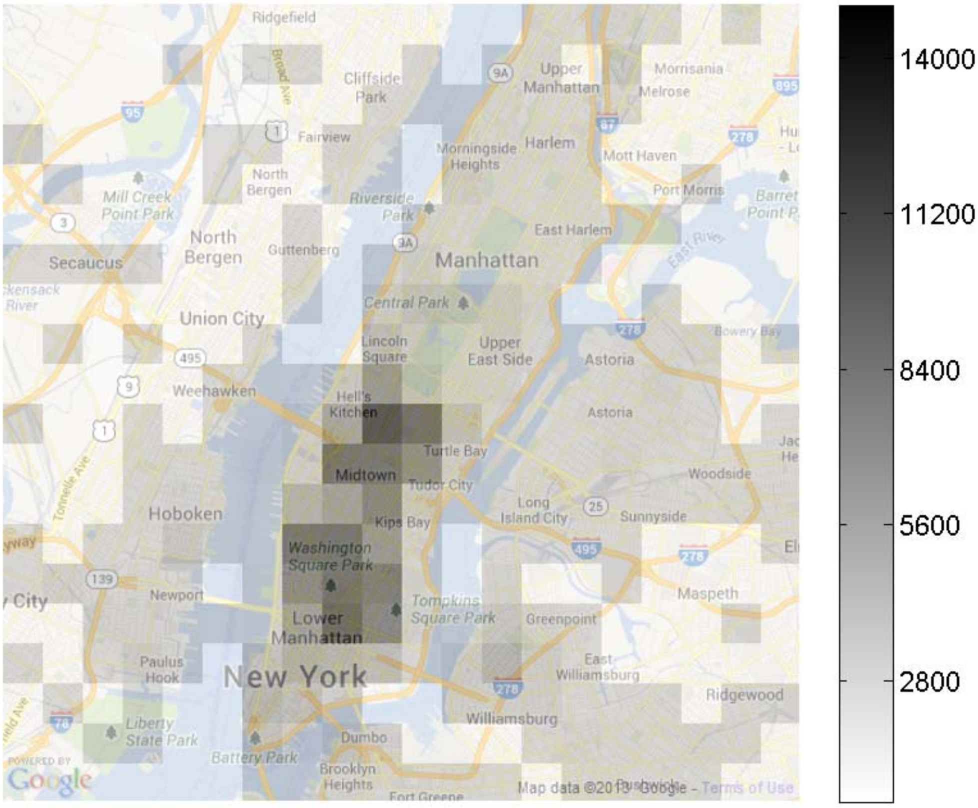


Figure 1: A heatmap of the distribution of the aggregate volume of check-ins in different neighborhood areas of New York City.

people are more actively check-ing during the afternoon.

For each one of the neighborhood areas on the grid, we have also computed the fraction of the check-ins that took place within this area for each one of the time periods we consider. Figure 3 presents our results for both cities (the stacked bar for areas that have less than 30 check-ins is left blank). As we can notice, the volume of the activity is not uniformly distributed across time. Furthermore, this distribution differs across neighborhood areas too. In other words, there exist areas for which the mass of the activity takes place in the morning (e.g., areas 88, 140, 174, 203, 219 in NYC and 17 in SF), in the afternoon (e.g., 24, 36, 59, 76, 87, 162, 191, 211, 224 and 227 in NYC and 3, 4, 10, 28, 43, 55, 66, 79, 91, 111 and 120 in SF) or in the night (e.g., areas 8 and 220 in NYC). We have further overlaid the above information for NYC on a map (Figure 4). These maps are different from the previous heatmaps. Here, the gray scale of each grid point, corresponds to the fraction of check-ins in this grid that took place during the corresponding time (to reiterate, it is simply an overlay of the stacked bars of Figure 3 on a map for better visualization). For instance, let us consider the area in the red ellipse. This area is near the Statue of Liberty and as we can see it attracts more check-ins/visitors during the daytime (morning and afternoon) as compared to night time. Furthermore, areas in the red rectangle are more active during the night, since these are areas that includes mainly bars. While this might have been expected, it strongly reinforces the claim that time is important for characterizing the activity that takes place in a geographical area.

To further delve into the details of the temporal variations in activity volume within an area, we have clustered the grid neighborhoods[4] with regards to the 3-dimensional vectors of their temporal activity volume distribution. We use spectral clustering [11], with the 10-nearest neighbor similarity graph where we use the cosine similarity as the similarity metric. Given two vectors $\overrightarrow{\mu}, \overrightarrow{\nu} \in \Re^n$, the latter is defined as:

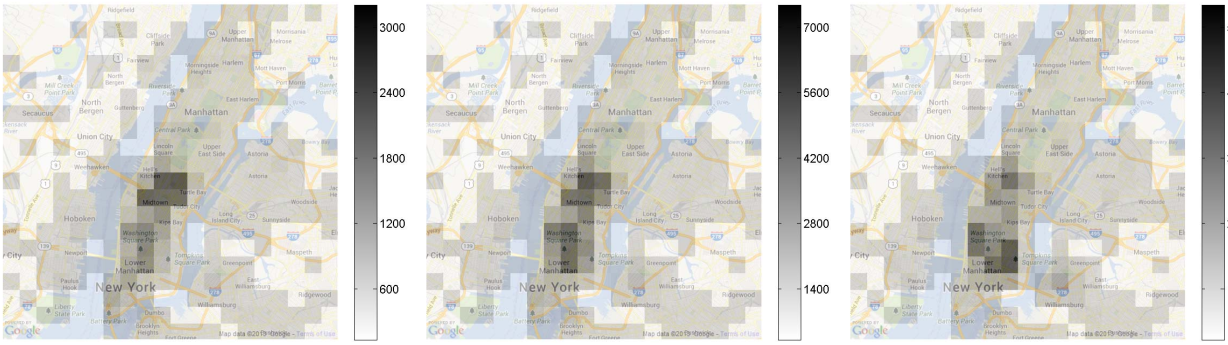
$$\cos(\overrightarrow{\mu}, \overrightarrow{\nu}) = \frac{\overrightarrow{\mu} \cdot \overrightarrow{\nu}}{||\overrightarrow{\mu}|| \cdot ||\overrightarrow{\nu}||} \quad (2)$$

We further apply the eigengap heuristic to decide on the number of clusters. Spectral clustering identified 7 clusters in NYC and 9 in SF. The centroids of each cluster are presented in Table 1. In parenthesis we also provide the number of areas that fall in the corresponding clusters.

Cluster ID	Centroids NYC	Centroids SF
0	(0.21,0.54,0.25) (42)	(0.13,0.41,0.46) (14)
1	(0.28,0.37,0.35) (38)	(0.30,0.61,0.09) (15)
2	(0.65,0.27,0.07) (18)	(0.08,0.56,0.36) (17)
3	(0.12,0.50,0.37) (33)	(0.19,0.51,0.30) (14)
4	(0.32,0.53,0.15) (44)	(0.51,0.40,0.08) (12)
5	(0.10,0.34,0.56) (30)	(0.31,0.44,0.25) (13)
6	(0.12,0.73,0.15) (24)	(0.14,0.78,0.07) (18)
7	-	(0.23,0.57,0.20) (11)
8	-	(0.11,0.67,0.22) (10)

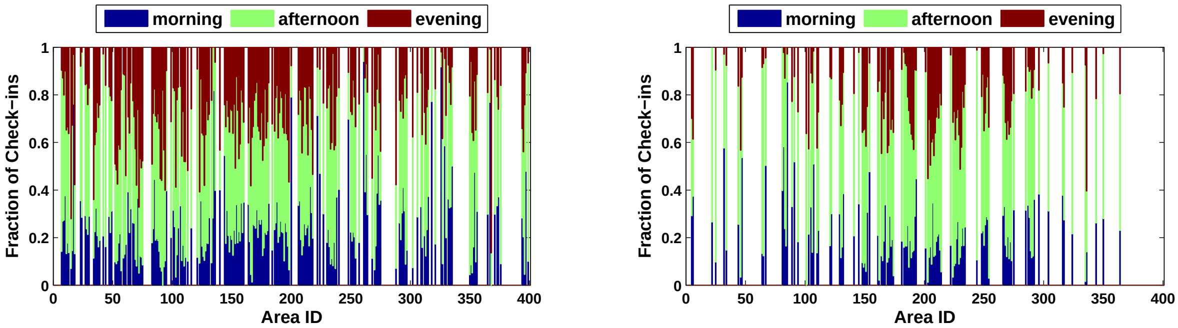
Table 1: Centroids of the different clusters of areas with regards to the temporal activity volume distribution.

[4] We focus on the areas with more 30 check-ins in total.



(a) Morning (b) Afternoon (c) Evening

Figure 2: The "most active" areas of the city change with time.



(a) New York City (b) San Francisco

Figure 3: The activity volume of an area is not uniformly distributed across time and different areas have different activity volume temporal distributions.

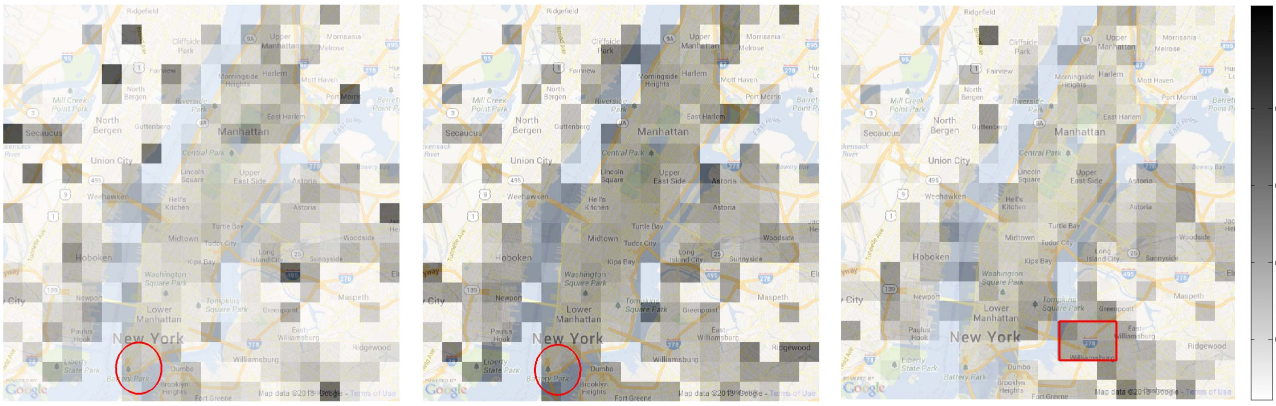
As we can see NYC exhibits seven basic *types* of areas with regards to the time when activity takes place. For instance, among others, there are areas that are active only during morning and afternoon hours (cluster 2) and areas that are active mainly during the afternoon (cluster 6). On the contrary, there is a little larger variety of neighborhoods in SF. For instance, areas that belong to cluster 2 are primarily areas that are active during the afternoon and evening, while areas that belong to cluster 5 have their activity more uniformly distributed across the day. As another example, neighborhoods of cluster 6 exhibit most of their activity during the afternoon hours. Figure 5 visualizes the distribution of clustered areas. As we see, cluster C4 in NYC contains areas that have their activities mainly in the morning and afternoon, which is consistent with the areas in the red ellipse in Figure 4(a) and 4(b). Also, areas in the red rectangle in Figure 4(c) belongs to the same cluster C5 (full of nightlife spots), which exhibits most of their check-ins during the evening. Note here that while we have used a common color-code for the clusters at NYC and SF, the details of the clusters in the two cities are different (see Table 1).

However, note here that this is only one view of the temporal dynamics related with the activity volume. For instance, areas in NYC cluster 2 (C2 for short) are mainly active during the morning, but this does not mean that they all exhibit the same type of activity. Area $x \in$ C2 can be dominated by office buildings, while $y \in$ C2 might be dominated from residence buildings and/or malls. In the following, we will examine the temporal dynamics of the type of activity that takes place in these urban areas.

Temporal dynamics of activity profile within an area: As aforementioned, a grid neighborhood n is associated with an activity profile vector $\overrightarrow{\alpha_n}$. The actual temporal dynamics of the activity that takes place within an area, can be captured through the changes in the activity vectors with time. A traditional metric that is used for quantifying the similarity between two vectors is the cosine similarity as introduced above in Equation 2. Hence, we can examine the similarity between the aggregate ($\overrightarrow{\alpha_n}$), morning ($\overrightarrow{\alpha_n}(t_m)$), afternoon ($\overrightarrow{\alpha_n}(t_a)$) and evening ($\overrightarrow{\alpha_n}(t_e)$) vectors of area n. Given that the elements of the activity profiles as defined are non-negatives, their cosine similarity will be taking values in the interval [0,1].

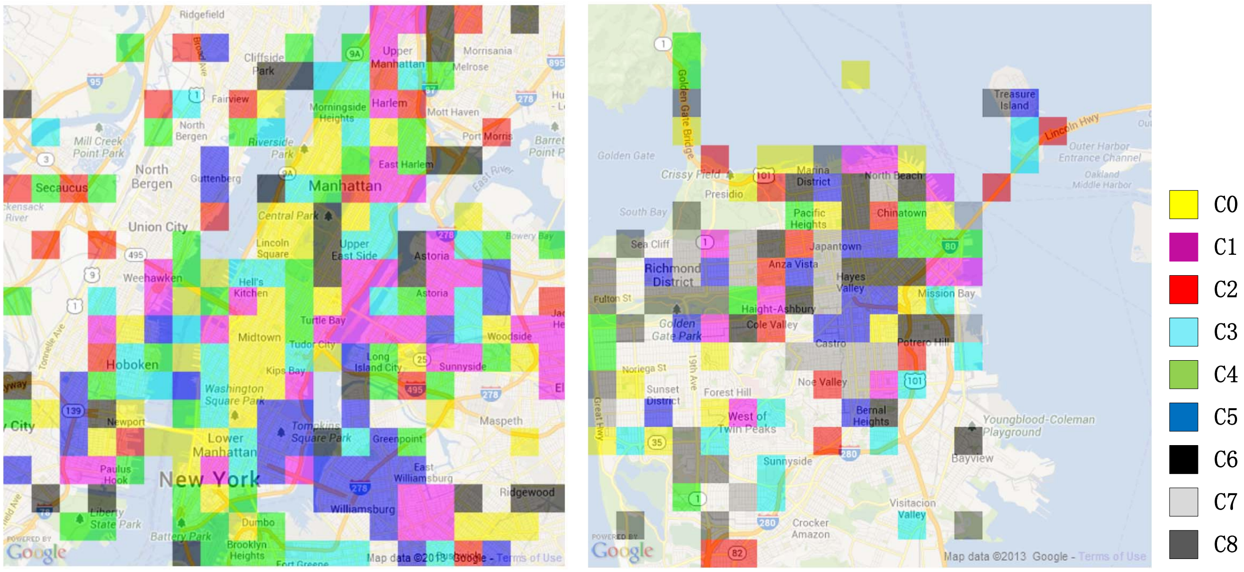
Figure 6 depicts our results for both cities. In particular, for every area with more than 30 check-ins in total, we compute the cosine similarity among all possible pairs of its above four activity profile vectors (six pairs in total). As we can see, while there exist stable areas, that is, large cosine similarity among the activity vectors, there is a significant fraction of areas whose pairwise activity vector similarities exhibit small values. This means that the activity profiles of these neighborhoods change over time and hence, one aggregate vector cannot capture the underlying temporal dynamics.

Spatio-temporal, activity-based clustering of urban areas: In what was presented above, we examined the temporal dynamics of each neighborhood in isolation. However, many applications require to identify areas that exhibit *similar* activity. This knowledge can facilitate services such as better municipal resource allocation and urban development. It can also enable commercial applications



(a) Morning (b) Afternoon (c) Evening

Figure 4: Visualizing the temporal distribution of activity volume in different geographic areas in NYC.



(a) New York City (b) San Francisco

Figure 5: Visualization of the clusters in Table 1. Areas with less than 30 check-ins are not labeled.

such as spatial recommendations. Given that the activity of an area is not stable over time, clusters of similar neighborhoods can also change.

Hence, we study how clustering of areas based on their activity vectors change over time. In particular, using again spectral clustering with the same set up as above, we cluster areas in NYC and San Francisco using as feature vectors their activity profiles. We identify clusters using the aggregate activity vectors and then compare the memberships in the clusters identified, with those that are derived when considering the three temporal activity vectors. Table 2 presents the number of clusters that were identified in every case. We further provide information for these clusters in the Appendix.

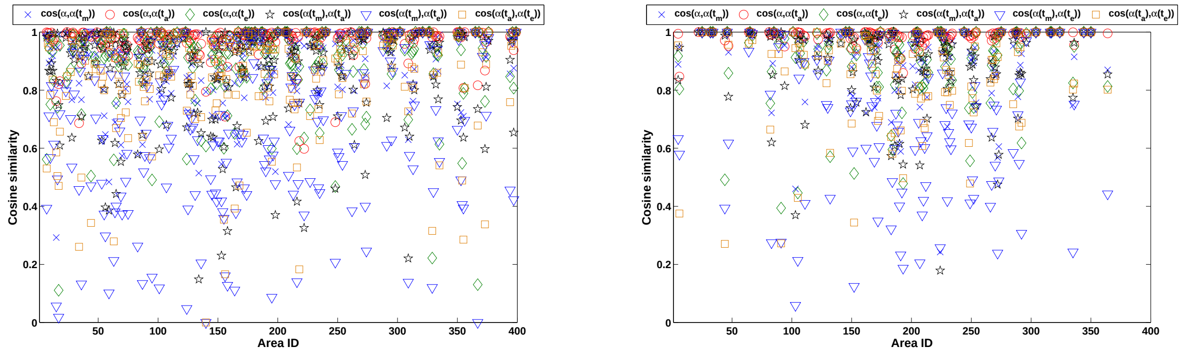
As we can see, even simply the number of groups of similar activity areas that are identified changes depending on the time of the day we are examining. We specifically want to examine pairs of areas that (do not) belong to the same cluster when considering the aggregate activity profiles. In particular, given that two areas x and y, belong to the same "aggregate" cluster C_i^a, we want to calculate how many times they fall into the same temporal clusters (morning, afternoon, evening). If Z is the variable representing the above number then $Z \in \{0, 1, 2, 3\}$. The higher the value of Z, the better the aggregate activity profile vector describes the two areas. Using our datasets we calculate the empirical probability mass function of Z, that is, $Pr\{Z = z\}$.

	$\overrightarrow{\alpha_n}$	$\overrightarrow{\alpha_n}(t_m)$	$\overrightarrow{\alpha_n}(t_a)$	$\overrightarrow{\alpha_n}(t_e)$
NYC	8	9	13	9
SF	10	5	4	10

Table 2: Number of clusters identified for different times of the day.

Similarly, we consider pairs of areas that do not belong to the same "aggregate" clusters, and we examine how many times they fall in the same temporal clusters. Denoting this variable with W, we see that again $W \in \{0, 1, 2, 3\}$, and we further compute the empirical probability mass function for W as well ($Pr\{W = w\}$).

Figure 7 presents our results. As we can see, there is a signifi-



(a) New York City (b) San Francisco

Figure 6: The activity of a given area is highly dynamic as captured from the cosine similarity between the aggregate and temporal activity vectors.

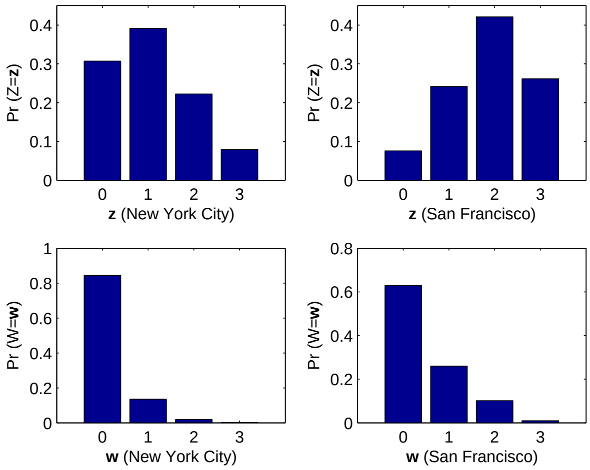


Figure 7: There is a large number of areas similar on aggregate but completely different in smaller time scales (Z small). On the contrary, there is a significantly smaller fraction of areas that are not similar on aggregate, but exhibit similarity in smaller time scales (W large).

cant portion of area-pairs, that exhibit Z=0 and Z=1 - especially for NYC. This means, that these areas, while they appear to be similar on aggregate, they rarely - if at all - obtain the same cluster label during the morning, afternoon or evening hours. Furthermore, while the majority of the pairs of areas that do not belong to the same "aggregate" cluster, they also never lay in the same temporal cluster (i.e., $W = 0$ has the more than 60% of the probability mass for variable W), there is a small fraction of pairs that appear similar in smaller time scales.

While we acknowledge that part of these results might have been affected from the clustering operations (e.g., the accuracy of the eigengap heuristic etc.), they clearly indicate that one should definitely consider the temporal dimension when examining urban activity.

4. DISCUSSION AND FUTURE DIRECTIONS

In this paper, we examine the temporal dynamics of urban activity. In particular, our main finding indicates that when studying urban dynamics we need to consider both space and time dimensions. While this might sound obvious, it has been largely ignored in existing literature and we further support it through a number of different results obtained from a rich Foursquare dataset.

We would like to emphasize here on the fact that similar studies, which utilize check-in information (or in general any other type of social media data) for studying urban dynamics are possible to suffer from a variety of biases. For instance, there can be demographic biases, since similar datasets capture mainly the behavior of a specific population (e.g., "tech savvy" people, who are usually the younger people). Furthermore, the quality of the datasets depends on many other factors. Virtual and real-world rewards can lead to people generate fake check-ins [9, 19], while the voluntary nature of check-ins can provide us with an "under-sampled" dataset of the urban activities.

Despite the above potential biases, check-in data are still a valuable source of information, especially as location-based services are becoming more prevalent. In the future, we plan to investigate and model fine-grain temporal urban activity dynamics in a greater detail. In particular, we seek to provide a generic analytical framework based on time-series clustering and network analysis that could be tuned and applied in a variety of settings/applications.

Acknowledgements

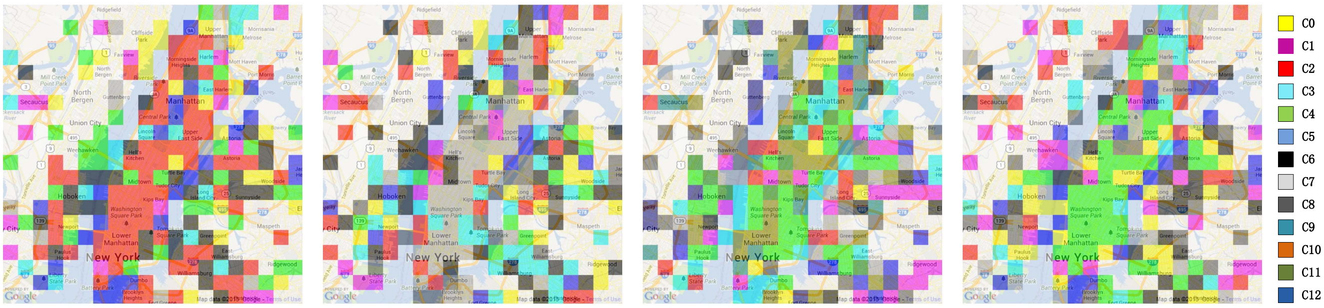
We would like to thank anonymous reviewers for their valuable comments that have helped us improve the quality of the manuscript.

5. REFERENCES

[1] R. Becker, R. Caceres, K. Hanson, J. Loh, S. Urbanek, A. Varshavsky, and C. Volinsky. A tale of one city: Using cellular network data for urban planning. In *IEEE Pervasive Computing*, 2011.

[2] F. Calabrese, M. Colonna, P. Lovisolo, D. Parata, and C. Ratti. Real-time urban monitoring using cell phones: a case study in rome. In *IEEE Transactions on Transportation Systems*, 2011.

[3] F. Calabrese, G. D. Lorenzo, L. Liu, and C. Ratti. Estimating origin-destination flows using mobile phone location data. In



(a) Aggregate (b) Morning (c) Afternoon (d) Evening

Figure 8: Visualization of (temporal) urban activity clusters.

IEEE Pervasive Computing, 2011.

[4] Z. Cheng, J. Caverlee, K. Lee, and D. Sui. Exploring millions of footprints in location sharing services. In *ICWSM*, 2011.

[5] E. Cho, S. A. Myers, and J. Leskovec. Friendship and mobility: User movement in location-based social networks. In *ACM KDD*, pages 279–311, 2011.

[6] J. Cranshaw, R. Schwartz, J. Hong, and N. Sadeh. The livehoods project: Utilizing social media to understand the dynamics of a city. In *AAAI ICWSM*, 2012.

[7] F. Girardin, F. Calabrese, F. D. Fiore, C. Ratti, and J. Blat. Digital footprinting: Uncovering tourists with user-generated content. *Pervasive Computing, IEEE*, 7(4):36–43, 2008.

[8] M. Gonzalez, C. Hidalgo, and A.-L. Barabasi. Understanding individual human mobility patterns. In *Nature, 453, 479-482*, 2008.

[9] W. He, X. Liu, and M. Ren. Location cheating: A security challenge to location-based social network services. In *IEEE ICDCS*, 2011.

[10] S. Isaacman, R. Becker, R. Caceres, M. Martonosi, J. Rowland, A. Varshavsky, and W. Willinger. Human mobility modeling at metropolitan scales. In *ACM MobiSys*, 2012.

[11] A. Ng, M. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In *NIPS*, 2001.

[12] A. Noulas, C. Mascolo, and E. Frias-Martinez. Exploiting foursquare and cellular data to infer user activity in urban environments. *International Conference on Mobile Data Management*, 2013.

[13] A. Noulas, S. Scellato, R. Lambiotte, M. Pontil, and C. Mascolo. A tale of many cities: universal patters in human urban mobility. In *PLoS ONE 7(5): e37027. doi:10.1371/journal.pone.0037027*, 2011.

[14] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil. An empirical study of geographic user activity patterns in foursquare. In *ICWSM (poster session)*, 2011.

[15] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil. Exploiting semantic annotations for clustering geographic areas and users in location-based social networks. In *SMW*, 2011.

[16] J. Reades, F. Calabrese, and C. Ratti. Eigenplaces: analysing cities using the space-time structure of the mobile phone network. *Environment and Planning B: Planning and Design*, 36(5):824–836, 2009.

[17] D. Wang, D. Pedreschi, C. Song, F. Giannotti, and A.-L. Barabasi. Human mobility, social ties, and link prediction. In *ACM KDD*, 2011.

[18] J. Yuan, Y. Zheng, and X. Xie. Discovering regions of different functions in a city using human mobility and pois. In *ACM SIGKDD*, 2012.

[19] K. Zhang, W. Jeng, F. Fofie, K. Pelechrinis, and P. Krishnamurthy. Towards reliable spatial information in lbsns. In *ACM LBSN*, 2012.

APPENDIX

Tables 3-6 show the centroids of the clusters identified both in the aggregate setting as well as in the three defined time-periods, while Figure 8 further visualizes the clusters for NYC - again we have used a common color-code even though for example the details of cluster C0 in the aggregate setting are different compared to C0 for the afternoon. Going back to the two areas, laying in the red ellipse in Figure 4(a) and 4(b), we see that they are classified into the aggregate cluster C5 which is dominated by 66.0% Outdoors activity. They also belong to C8 (64.0% Outdoors) in the morning, and C3 (67.1% Outdoors) in the afternoon. However, in the evening, they are separated into C5 and C7; the latter exhibits totally different activities with Art venues being dominant (64.4%). As another more interesting example, the six areas in the red rectangle of Figure 4(c) are classified into clusters C2 and C6 when considering aggregate activities. From Table 3, we can see C2 is dominated by 38.3% Food and 14.7% Nightlife, while C6 is 14.5% Food and 50.0% Nightlife. The interesting thing is that these areas shift the main activity from Food to Nightlife from daytime to nighttime. In the morning, four of these six areas belong to cluster C3 (68.5% Food, 4.8% Nightlife), one in C4 (12.9% Food, 64.9% Nightlife)[5] and one in C6 (6.3% Food, 1.8% Nightlife). In the afternoon, three of them belong to clusters C4 (44.1% Food, 21.9% Nightlife), two in C5 (76.2% Food, 4.3% Nightlife) and one in C7 (14.3% Food, 65.2% Nightlife). Finally, in the evening, three of them belong to C0 (17.0% Food, 49.6% Nightlife), two in C4 (33.7% Food, 32.0% Nightlife) and one in C6 (2.0% Food, 63.1% Nightlife). It should be clear that considering the temporal dimension can give a much more detailed view of the urban activity dynamics.

[5]This cluster includes a large percentage of nightlife venues, but many of the venues that are classified as such are "cafe/bars" that are also possibly open during the morning serving coffee and breakfast.

Cluster ID	Arts	College	Food	Nightlife	Outdoors	Professional	Residence	Shops	Travel
0	1.6%	0.5%	19.3%	4.1%	3.3%	4.1%	3.2%	56.5%	7.3%
1	2.6%	0.3%	7.8%	3.5%	0.9%	60.1%	6.7%	7.4%	10.6%
2	5.7%	7.5%	38.3%	14.7%	5.3%	8.3%	4.3%	11.0%	5.0%
3	60.0%	1.3%	8.7%	4.0%	6.0%	10.2%	1.7%	4.3%	4.1%
4	1.2%	0.5%	7.8%	1.5%	5.8%	3.9%	5.2%	6.1%	68.0%
5	6.8%	0.1%	9.3%	3.1%	66.0%	5.6%	3.2%	2.1%	3.9%
6	9.4%	0.2%	14.5%	50.0%	5.8%	3.1%	3.7%	4.0%	9.2%
7	3.7%	0.0%	8.0%	6.4%	2.3%	8.4%	67.3%	5.4%	1.8%

Table 3: Centroids of the different activity clusters with regards to the aggregate activity vectors.

Cluster ID	Arts	College	Food	Nightlife	Outdoors	Professional	Residence	Shops	Travel
0	0.3%	0.7%	14.1%	0.2%	1.8%	4.0%	4.5%	68.4%	6.0%
1	66.5%	0.1%	3.8%	3.2%	16.3%	4.4%	0.9%	2.7%	2.1%
2	2.5%	3.1%	4.3%	0.5%	2.6%	66.4%	3.4%	6.3%	10.9%
3	3.4%	1.3%	68.5%	4.8%	4.5%	2.1%	1.5%	7.8%	6.3%
4	4.1%	0.0%	12.9%	64.9%	2.6%	7.8%	4.4%	1.0%	2.3%
5	0.2%	1.4%	11.8%	1.4%	12.3%	7.4%	53.9%	9.5%	2.0%
6	1.9%	5.6%	6.3%	1.8%	4.7%	2.0%	3.0%	8.4%	71.4%
7	6.9%	11.7%	24.9%	2.9%	6.2%	24.9%	1.7%	15.8%	5.0%
8	2.0%	0.1%	6.3%	1.4%	64.0%	6.2%	3.4%	5.0%	4.0%

Table 4: Centroids of the different activity clusters with regards to the morning activity vectors.

Cluster ID	Arts	College	Food	Nightlife	Outdoors	Professional	Residence	Shops	Travel
0	3.1%	54.9%	18.6%	3.7%	2.5%	4.5%	3.1%	5.2%	4.4%
1	1.6%	0.5%	20.1%	7.5%	2.8%	8.6%	3.1%	12.4%	43.5%
2	61.8%	0.3%	4.7%	10.5%	11.1%	6.7%	1.2%	1.5%	2.0%
3	7.1%	0.4%	7.6%	3.4%	67.1%	5.1%	2.0%	1.7%	5.8%
4	3.2%	1.0%	44.1%	21.9%	5.2%	5.1%	3.7%	12.0%	3.8%
5	0.6%	0.1%	76.2%	4.3%	2.7%	3.4%	5.8%	2.8%	4.0%
6	0.8%	1.1%	5.7%	1.2%	4.6%	4.7%	2.2%	75.5%	4.2%
7	7.0%	0.0%	14.3%	65.2%	1.6%	3.9%	3.8%	2.6%	1.6%
8	2.2%	0.8%	1.4%	1.3%	6.1%	2.1%	3.2%	2.9%	80.1%
9	3.5%	0.5%	18.3%	1.3%	2.1%	50.5%	6.2%	12.2%	5.5%
10	5.5%	1.4%	33.7%	6.2%	7.7%	4.9%	3.5%	32.5%	4.6%
11	31.7%	2.7%	31.1%	7.1%	3.4%	5.4%	2.5%	13.2%	3.0%
12	0.2%	0.0%	8.2%	4.3%	4.7%	5.0%	65.9%	5.2%	6.4%

Table 5: Centroids of the different activity clusters with regards to the afternoon activity vectors.

Cluster ID	Arts	College	Food	Nightlife	Outdoors	Professional	Residence	Shops	Travel
0	1.2%	0.3%	17.0%	49.6%	3.0%	1.7%	11.8%	4.1%	10.7%
1	2.0%	0.0%	6.1%	1.2%	5.3%	2.0%	2.8%	4.6%	75.8%
2	1.2%	0.0%	13.3%	2.5%	1.7%	18.9%	5.1%	54.5%	3.1%
3	0.3%	1.2%	6.7%	4.1%	2.6%	3.5%	73.5%	1.8%	6.3%
4	8.6%	4.1%	33.7%	32.0%	3.4%	3.0%	5.0%	6.3%	4.4%
5	3.1%	1.3%	8.3%	3.8%	65.2%	3.6%	3.7%	6.2%	6.1%
6	9.3%	0.7%	2.0%	63.1%	2.0%	0.7%	1.7%	2.9%	2.4%
7	64.6%	0.2%	8.8%	7.1%	6.1%	5.1%	1.6%	3.0%	3.4%
8	2.7%	0.1%	63.8%	2.0%	3.8%	3.3%	10.0%	7.3%	7.1%

Table 6: Centroids of the different activity clusters with regards to the evening activity vectors.