Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media

The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City

Justin Cranshaw jcransh@cs.cmu.edu
Raz Schwartz razs@andrew.cmu.edu
Jason I. Hong jasonh@cs.cmu.edu
Norman Sadeh sadeh@cs.cmu.edu

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

Studying the social dynamics of a city on a large scale has traditionally been a challenging endeavor, often requiring long hours of observation and interviews, usually resulting in only a partial depiction of reality. To address this difficulty, we introduce a clustering model and research methodology for studying the structure and composition of a city on a large scale based on the social media its residents generate. We apply this new methodology to data from approximately 18 million check-ins collected from users of a location-based online social network. Unlike the boundaries of traditional municipal organizational units such as neighborhoods, which do not always reflect the character of life in these areas, our clusters, which we call *Livehoods*, are representations of the dynamic areas that comprise the city. We take a qualitative approach to validating these clusters, interviewing 27 residents of Pittsburgh, PA, to see how their perceptions of the city project onto our findings there. Our results provide strong support for the discovered clusters, showing how Livehoods reveal the distinctly characterized areas of the city and the forces that shape them.

Introduction

The forces that shape the dynamics of a city are multifarious and complex. Cultural perceptions, economic factors, municipal borders, demography, geography, and resources—all shape and constrain the texture and character of local urban life. It can be extremely difficult to convey these intricacies to an outsider; one may call them well-kept secrets, sometimes only even partially known to the locals. When outsiders, such as researchers, journalists, or city planners, do want to learn about a city, it often requires hundreds of hours of observation and interviews. And although such methods offer a way to gather deep insights about certain aspects of city life, they simply do not scale, and so can ever only uncover a partial image of the inner workings of the city.

This work presents a new methodology for studying the dynamics, structure, and character of a city on a large scale. Our approach is fundamentally data-driven. Given geospatial social media data from hundreds of thousands of people, we developed an algorithm that maps the distinct geographic areas of the city depicting a snap-shot of the ever-changing activity patterns of its people. Contrary to traditional organizational units such as neighborhoods that are often stagnant and may portray old realities, our clusters reflect current collective activity patterns of people in the city, thus revealing the dynamic nature of local urban areas, exposing their individual characters, and highlighting various forces that form the urban habitat.

Our work is made possible by the rapid proliferation of smartphones in recent years and the subsequent emergence of location-based services and applications. Location-based social networks such as foursquare have created new means for online interactions based on the physical location of their users. In these systems, users can "check-in" to a location by selecting it from a list of named nearby venues. Their check-in is then broadcast to other users of the system.

To algorithmically explore the dynamics of cities, we use data from millions of check-ins gathered from foursquare. Using well studied techniques in spectral clustering, we introduce a model for the structure of local urban areas that groups nearby foursquare venues into clusters. Our model takes into account both the *spatial proximity* between venues as given by their geographic coordinates, as well as the *social proximity* which we derive from the distribution of people that check-in to them. The underlying hypothesis of our model is that the "character" of an urban area is defined not just by the types of places found there, but also by the people that choose to make that area part of their daily life. We call these clusters *Livehoods*, reflecting the dynamic nature of activity patterns in the lives of city inhabitants.

We take a qualatative approach to evaluating this hypothesis. In a true urban studies tradition, we conducted interviews with 27 residents of different areas of Pittsburgh, Pennsylvania to see how well their mental maps (Milgram 1977) of the city projected onto the Livehood clusters our algorithms discovered. In addition to validating our algorithm, these interviews shed light onto the forces that shape the city, and the ways in which Livehoods can help untangle them. We present these results through a series of case studies exploring the relationships between our Livehood clusters and the municipal neighborhoods.

Our results provide a promising step for the emerging field urban computing,[1] which seeks to provide a compu-

[1]Urban computing, like many new fields, has several semi-

tational perspective to understanding and improving the urban form. In merging computer science research with social science and urban studies, our work also contributes to other growing fields such as computational social science and digital humanities. By rethinking the organizational units that structure a city according to people's movements, our model has many potential applications, both for future automated algorithms that hope to make the city infrastructure smarter, and also for businesses, urban planners, real-estate developers, researchers and end users who are looking to more effectively explore the various areas of a city.

This work offers three main contributions. First, we present a clustering model for discovering the distinct geographic areas of the city, reflecting the collective movement patterns of its people. We then introduce a methodology based on semi-structured interviews for exploring the resulting clusters and the urban dynamics that shape them. Finally, we provide an interactive web-based tool[2] for visualizing the clusters, allowing users to discover new insights about the city.

Background and Related Work

Our results are based on check-in data that was gathered from foursquare. Foursquare is a location-based online social network that was founded in 2009 that provides users a way to share their location with their friends by "checking-in" to the places they visit. As of December 2011, foursquare registered 15 million users, and over 1 billion checkins (foursquare 2011). Although our technique is agnostic to the particular source of such data, foursquare is appealing because it is the first such service to gain a wide user base. This success is due in part to the mix of uses and services that the system provides for its users, as foursquare has many built-in mechanisms that actively encourage users to check-in.

Our methodology for comparing the fabric of the urban environment with the results of our clustering algorithm is grounded in several earlier works in urban studies and urban design. Such works investigate the structure and function of cities (Lynch 1992) as well as people's perception of their local surroundings (Suttles 1973; Jacobs 1992), the importance of social interactions for the creation of the local character (Milgram 1977; Putnam 2000; Oldenburg 1989). Such studies often require long time spans and extended resources to discover meaningful results.[3]

As data from location sharing systems are becoming increasingly available for researchers to analyze, there have been a number of recent results from social science, computer science, and machine learning that are finding new ways to extract various insights on relations between online and offline interactions (Gordon and de Souza e Silva 2011; Cranshaw et al. 2010), large scale urban dynamics (Cranshaw and Yano 2010; Noulas et al. 2011; Chang and Sun 2011) and the effects location technologies have over people's behavior (Lindqvist et al. 2011; Cramer, Rost, and Holmquist 2011; Schwartz 2012). This work also aligns with the ideas offered by Rainie and Wellman (2012) who note that the move towards flexible, mobile, fragmented social systems results in the weakening of traditional boundaries such as neighborhoods.

Clustering Model

Our data comes from two sources. We combine approximately 11 million foursquare check-ins from the dataset of Chen et al. (2011) with our own dataset of 7 million check-ins that were downloaded between June and December of 2011. Foursquare check-ins are by default not publicly visible, however users may elect to share their check-ins publicly on social networks such as Twitter. These 18 million check-ins were all collected from the Twitter public timeline, then were aligned with venue information from the foursquare API. For each check-in, our data consists of the user ID, the time, the latitude and longitude, the name of the venue, and the category of the place.[4]

In this work, we examine the Livehood clusters from Pittsburgh, PA. In the Pittsburgh metropolitan area, our data contained 42787 check-ins of 3840 users at 5349 venues.

Clustering Algorithm: We present a spectral clustering approach to the discovery of local urban areas from geospatial check-in data. Spectral methods for data clustering are a well studied (Shi and Malik 1997; Luxburg 2007), and are popular in practice due to the quality of the clusters that are often produced and the simplicity of implementation.

One of our main contributions is the design of an affinity matrix between check-in venues that effectively blends spatial affinity and social affinity. Suppose that V is a set of n_V foursquare venues and that for each $i, j \in V$ we can compute a geographic distance $d(i, j)$ given their latitude and longitude coordinates. We also have a set U of n_U foursquare users, and a set C of check-ins of these users to the venues in V. Ignoring the temporal aspects, we represent each venue v by the "bag of check-ins" to v. That is, we compute an n_U dimensional vector c_v, where the u^{th} component of c_v is the number of times user u checked-in to v. Under this representation, we can compute a social similarity $s(i, j)$ between each pair of venues $i, j \in V$ by comparing the vectors c_i and c_j. Using cosine similarity for this measure, we get $s(i,j) = \frac{c_i \cdot c_j}{||c_i|| \, ||c_j||}$.

We compute an $n_V \times n_V$ affinity matrix $A = (a_{i,j})_{i,j=1,\ldots,n_V}$ on the venues as follows. First, for a given venue v, we let $N_m(v)$ be the m closest venues to v according to the $d(v, \cdot)$ for some parameter m. Then we let

$$a_{i,j} = \begin{cases} s(i,j) + \alpha & \text{if } j \in N_m(i) \text{ or } i \in N_m(j) \\ 0 & \text{otherwise} \end{cases}$$

where α is a small constant that prevents degenerate venues from having no connections to any others.

interchangeable names. It is also often referred to as urban Informatics, urban analytics, and smart cities.

[2] See http://livehoods.org/.

[3] William Whyte needed to collect thousands of hours of video to study social interactions in the city's public places (Whyte 2001) while Stanley Milgram deployed an army of graduate students to document New Yorkers in their daily routines (Milgram 1977).

[4] Foursquare venues are each labeled from a category hierarchy such as *College & University::College Bookstore*.

Algorithm 1 *Spectral Clustering for Livehoods*
Input: V, $A = (a_{i,j})$, $G(A)$ the graph of A, k_{min}, k_{max}, τ

1: Compute diagonal degree matrix D with diagonal $(d_1, \ldots, d_{n_V})$ where $d_i = \sum_{j=1}^{n_V} a_{i,j}$.
2: $L := D - A$
3: $L_{norm} := D^{-1/2} L D^{-1/2}$
4: Let $\lambda_1 \leq \ldots \leq k_{max}$ be the k_{max} smallest eigenvalues of L_{norm}. Set $k = \arg\max_{i=k_{min},\ldots,k_{max}-1} \Delta_i$ where $\Delta_i = \lambda_{i+1} - \lambda_i$.
5: Find the k smallest eigenvectors $e_1, \ldots, e_k$ of L_{norm}.
6: Let E be an $n_V \times k$ matrix with e_i as columns.
7: Let the $y_1, \ldots, y_{n_V}$ be the rows of E, and cluster them into $C_1, \ldots, C_k$ with k-means. This induces a clustering on $A_1, \ldots, A_k$ by $A_i = \{j | y_j \in C_i\}$.
8: For each A_i, let $G(A_i)$ be the subgraph of $G(A)$ induced by vertices A_i. Split $G(A_i)$ into connected components. Add each component as a new cluster, removing $G(A_i)$.
9: Let b the area of bounding box containing coordinates in V, and b_i be the area of the box containing A_i. If $b_i/b > \tau$, delete cluster A_i, and redistribute each $v \in A_i$ to the closest A_j under single linkage distance $d(v, A_j)$.

Viewed as a graph, we connect each venue node with an undirected edge to its m nearest neighbors by geographic distance, and we weigh the edges according the cosine similarity of the distributions of check-ins at the two venues. We denote this graph representation as $G(A)$.

Algorithm 1 describes our clustering approach. We use the variation of spectral clustering introduced by Ng, Jordan, and Weiss (2001). Similar to Zelnik-manor and Perona (2004) we introduce a post processing step to clean up any degenerate clusters: step 8 splits each cluster into its disconnected components under $G(A)$, and step 9 removes any clusters that span too large a geographic area. Following many other works, in step 4 we select the number of clusters by searching for the largest gap in consecutive eigenvalues between a minimum and maximum allowable number of clusters.

Related clusters: We also developed a way to compare different clusters based on the similarity of the distributions of users that visit them. Again we use a cosine similarity measure. For each cluster A_i, we represent it as an n_U dimensional vector c_{A_i}, where each component u is the number of check-ins users u had to *any venue* in A_i. We compute the similarity between all pairs of clusters as $s(A_i, A_j) = \frac{c_{A_i} \cdot c_{A_j}}{||c_{A_i}|| \, ||c_{A_j}||}$.

Implementation details: Note that by only connecting each venue v to its m nearest neighbors in geographic distances, it allows us to keep the matrices extremely sparse, enabling us to scale to process hundreds of thousands of venues without any need for parallelization. Large sparse matrices can be efficiently stored, and the first k eigenvectors can be computed quickly with a Lanczos solver. Moreover, we can compute the set of nearest neighbors $N_m(v)$ highly efficiently using k-d trees.

The parameters we used in this work for the Pittsburgh clusters were $m = 10$, $\alpha = 0.01$, $k_{min} = 30$, $k_{max} = 45$, and $\tau = 0.4$.

The Livehoods website: We built an interactive website at http://livehoods.org that visualizes the Livehood clusters of several cities on a map, allowing users to explore various check-in and venue statistics for each Livehood, and the structure of related Livehoods.

Methodology

Between Nov. 17th and Dec. 17th, 2011 we conducted interviews with 27 residents of Pittsburgh. We recruited participants through a webpage that was shared by local businesses and neighborhood organizations on their Facebook and Twitter accounts. To qualify for the study, participants had to be at least 18 years old and must have lived in their neighborhood for at least one year. The interviews took approximately one hour and participants were compensated \$10 for their time.

Of our 27 interviewees 22 were people who responded to the recruitment posting. There were 12 females and 10 males among this group, and they represented a wide age range (mean age 35, min 23, max 62, and standard deviation 11) and had diverse educational backgrounds (1 had completed high school, 2 had some college, 10 had bachelor's degrees, 2 had some graduate school, and 7 had master's degrees). The interviews took approximately one hour, and participants were compensated \$10 for their time. Any names of these 22 recruited participants that appear in this work are pseudonymous. The remaining 5 participants were domain experts with whom we specifically requested interviews: a senior planner from the Pittsburgh Office of City Planning, an independent real-estate developer, and three partners of a large real estate development firm.

The primary goal of these interviews was to validate the clusters discovered by our algorithm. To accomplish this, we developed an interview protocol that explored the similarities and differences between our clusters and the official municipal neighborhood boundaries. We focused on three *dispersion patterns* that explore the intersection between Livehoods and municipal borders: (1) *split* – when a municipal neighborhood contains more than one Livehood, (2) *spilled* – when a Livehood cluster spills over the boundaries of a municipal border, and (3) *corresponding* – when the Livehood cluster and the municipal borders coincide. We use these patterns to identify points of interest to explore in our interviews, assuming that different patterns reflect different local dynamics.

The semi-structured interviews with the participants began with a discussion of their backgrounds in relation their neighborhood. Then without giving specific instructions, we showed them a map of Pittsburgh, and asked them to draw the boundaries of their neighborhood over it. This offered a way to anchor the subsequent interview, and let us explore the differences among the participants' mental perceptions of the area, the municipal neighborhood borders, and the Livehood clusters. Next, we asked the participants whether there are places within the area that they drew where there is a “shift in feel” of the neighborhood. If so, we asked them to mark them on the drawing.

Next we showed a website with an interactive map that had the municipal neighborhood boundaries overlaid on top of it, and we asked them for any comments. Looking at this map, we then asked if there were neighborhoods where the “borders might be shifting or in flux.”

After that, we showed the them a map of the Livehoods clusters, initially explaining that the map shows different “areas of the city” based on an algorithm that looks at “trends of where people go.” The participants were asked to study the map and then to give their feedback. Later, we revealed how the algorithm works, including how we obtained the data. Finally, we showed the participants the “related areas” feature of the website for the areas of the city we discussed.

Results

Our results include three case studies of different areas in the city of Pittsburgh, each reflects one or more of our identified dispersion patterns. We selected these case studies based on the amount of attention they received in the interviews. In each, we characterize and give a short background of the neighborhoods of the area, describe the Livehoods that we found there, and present the interviewees input.

Shadyside and East Liberty: In the fall of 2002, a Whole Foods Market opened in Pittsburgh directly on the border of two very distinct neighborhoods – East Liberty to the north and Shadyside to the south, separated by train tracks and a public busway. East Liberty, once the third-largest retail center in Pennsylvania, has suffered the pains of decades of neglect which led to high crime rates and a demographic population consisting of mainly low income, predominantly black residents. On literally the other side of the tracks is Shadyside, one of the most coveted neighborhoods, characterized in our interviews as a wealthy, predominately white neighborhood (O’Toole 2010). The upscale grocery store that was situated between them was the first component of “East Side,” a multi-phase development project in East Liberty orchestrated by The Mosites Company, a local real-estate firm. Since the opening of Whole Foods, the surrounding area has been massively transformed, consequently affecting patterns of behavior for both local residents and visitors.

Our algorithm discovered two Livehoods in this region. In Figure 1 (Left), LH1 is almost completely contained within Shadyside and encompasses Walnut Street (one of three Shadyside business districts), and the western end of Shadyside, which is mostly residential. On the other hand, LH2 *spilled* across the boundary between East Liberty and Shadyside, containing all of East Liberty and the Whole Foods, in addition to Shadyside’s two other business districts (Ellsworth and Highland) and the eastern residential end of Shadyside.

Two main notions emerged from our interviews that support the way our algorithm clustered this area. First, the high-end national stores of Walnut Street draw an entirely different demographic than the locally owned independent shops of Highland and Ellsworth, supporting the *split* between the eastern end of Shadyside and the western end. Second, the recent developments of East Side, are actively blending the distinction between Shadyside and East Liberty, by connecting the business districts in both neighborhoods, supporting the *spilled* pattern in the region.

Kelley, a 29 year old resident of Shadyside, explained the difference between Walnut and Ellsworth:

> When you go to Walnut Street, that’s where I often see an older demographic. You will see women and men above the age of 50 walking around with shopping bags. I don’t see that demographic on Ellsworth ever shopping around... So I would say that’s a big difference. You are going to see older, straight, richer people on Walnut and you are going to see much younger, more indie looking people on Ellsworth.

The distinction Kelley made between the two areas repeated in many of our interviews with Shadyside residents. In addition to the different demographics visiting the commercial districts, several of our participants noted that the housing stock on each end is different. In the area surrounding Walnut Street one can find large self-owned, well-maintained family houses while on the eastern part, there is much more rental housing, primarily marketed towards students, and young professionals.

The grouping of the eastern portion of Shadyside with East Liberty in LH2 was also supported by many of our participants. For Kelley and many others that live in eastern Shadyside, socializing and using resources in the developed area of East Side feels more natural. As Erin, a 24 year old graphic designer, notes:

> That makes sense to me because I think at one point it was more walled off and this was poor [East Liberty] and this was wealthy [Shadyside] and now there are nice places in East Liberty and there’s some more diversity in this area so they are becoming more the same. And I do think Shadyside is almost shrinking and you only do have a few streets that are really that wealthy and bougie any more.

Just like in our interview with Erin, the blurring of the borders between Shadyside and East Liberty appeared time and time again. Overall, 85% of the interviewees named this area when they were asked the open ended question: “Can you think the neighborhood borders are shifting or in flux anywhere around the city?” For Shadyside residents the East Side development is a natural extension of their neighborhood while for East Liberty people it is clearly part of their territory.

Although we received a great deal of support for our cluster, the mapping was perceived as controversial for several interviewees, mostly older residents of the area. For them, the developments in East Liberty did not blur the lines between the two neighborhoods but rather created neutral grounds where both groups meet. As Donna, a 62 year old resident of East Liberty said in regard to the East Side development: “it doesn’t bring us together. It’s a place where both sides feel comfortable with.”

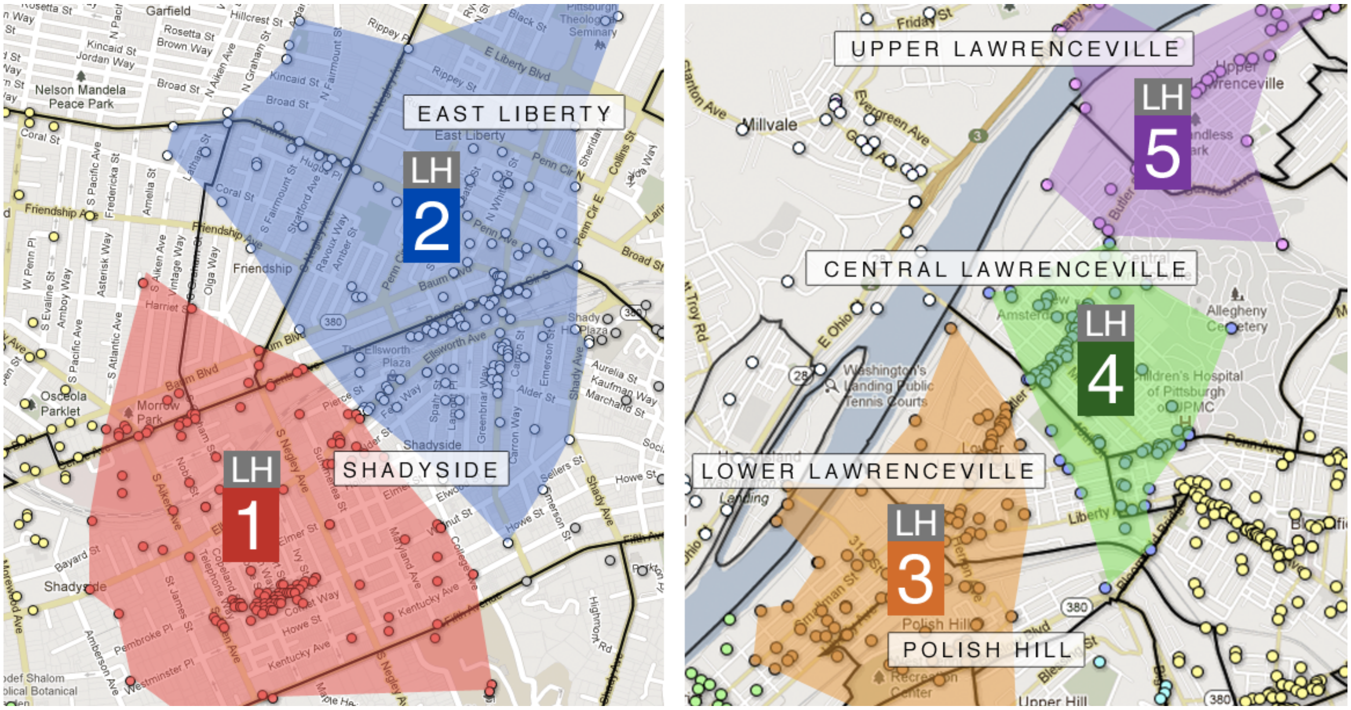

Figure 1: The municipal borders (black) and Livehoods for Shadyside/East Liberty (Left) and Lawrenceville/Polish Hill (Right).

Lawrenceville and Polish Hill: Lawrenceville, one of Pittsburgh's largest neighborhoods, had been going through massive changes and development in recent years. Our interviewees were conflicted about the cohesiveness of the area. For some, it is one big neighborhood encompassing more than 20 blocks whereas others notice distinct subsections carrying different characteristics.

The city itself subdivides Lawrenceville into three different municipal neighborhoods: Upper Lawrenceville, Central Lawrenceville, and Lower Lawrenceville. And although these areas are all connected by Butler street, the character of each of them is different. As Daniel, a 43 year old resident of Lawrenceville, explains:

> The look isn't different, but the vibe and the feel are very different. Middle Lawrenceville from 40th until the cemetery that is where the first people were moving in and fixing up the area... And then, Lower Lawrenceville, is kind of picking up right now and then Upper Lawrenceville it's been like the really rough area with gangs and drugs.

Our algorithm found similar divisions, breaking the area into three Livehoods with boundaries closely *corresponding* to those of the municipal map (see Figure 1 Right). The border between LH3 and LH4 was situated exactly on the 40th St. Bridge, the border between Lower and Cental Lawerenceville. The division between LH4 and LH5 was placed on 48th street, three blocks away from the municipal border between Central an Upper Lawrenceville on 51st street.

We found strong evidence from our interviews supporting the Livehood clusters based on factors such as property values, crime rates, business types, and general feel. As Claudia, a 54 year old journalist, notes:

> I think middle [Central] Lawrenceville is the most desirable or well rooted. Where the better housing stock is. LH3 is definately newer. LH5 pretty much was left alone... There are parking lots and convenience stores around 40th that when you hit those you think 'I have left something behind.' And then you are in another part of Lawrencevile because you passed a bridge and there's not a lot of connective tissue at some of these intersections.

Several of the interviewees did not agree with the separation of Lower and Central Lawrenceville. For them, the separation is arbitrary and it is based mainly on local businesses' interests. Since Lawrenceville was perceived as a dangerous area, a group of business owners in Lower Lawrenceville decided to brand the area as "LoLa" and market it as a stand alone destination for unique shops and restaurants.

Another point of interest is in the *spilling* of Lower Lawrenceville into the adjacent neighborhood of Polish Hill in LH3. At first glance, this grouping seems odd and not feasible. Polish Hill is a very small neighborhood that is separated from Lower Lawrenceville by train tracks and a bus way in addition to geographic barrier of being located on an hill. But this grouping seemed natural to Roger, a 47 year old resident of Polish Hill who said:

> I think it's pretty accurate... I think that's how some of our residents identify with Lower Lawrenceville because of their activities and their perception. Where

Figure 2: The municipal borders (in black) and Livehoods for South Side.

> they go to for entertainment, where they go for food, where they go because they enjoy the walk.

The connection the algorithm discovered between these two areas went both ways. As Jessica (a LH3 resident) explains "there are some places in Polish Hill we hang out a lot that feel more like our neighborhood."

The South Side: The South Side Flats neighborhood of Pittsburgh lies along the southern border of the Monongahela River. The main business district in the South Side is along Carson Street, which is one of the top destinations for nightlife in the city, as it has a high density of bars and restaurants. Moreover, occupying a large area on the eastern end of the neighborhood, there is a recently built mixed-use development called South Side Works consisting of an open air shopping mall with national vendors, several office buildings, and luxury condos.

Our clustering algorithm *split* South Side Flats into four Livehoods (see Figure 2). LH7 is the area along Carson between Liberty Bridge and 18th Street, LH8 is the area between 18th and 24th Street, and LH9 is the area east of 24th Street. The fourth area, LH6 is a shopping plaza north of LH8.

In our interviews, we found strong support of the Livehoods clustering for South Side. Particularly strong was the evidence supporting the *split* between the western part of South Side Flats (LH6, LH7 and LH8), and the eastern portion around South Side Works (LH9). We asked every subject who was familiar with South Side to indicate any places where they notice a "shift in feel," and nearly all participants indicated that South Side Works, which begins just to the east of the Birmingham Bridge, is distinctly different from the rest of South Side Flats.

When we showed the municipal borders of South Side to Ashley, a 25 year old who works at a local radio station, she was surprised, commenting "Oh! So that is just all one big neighborhood. I would have definitely thought there is a division near the Birmingham [Bridge]." Later, when we showed the Livehoods mapping and asked her about the boundary between LH8 and LH9, she exclaimed:

> Ha! Yes! See, here is my division! Yay! Thank you algorithm! ...I definitely feel where the South Side Works and all of that is, is a very different feel.

This "different feel" around South Side Works was identified by many of the subjects. Sara, a 30 year old video game designer who lives and works in South Side describes South Side Works as "more up-scale" and having "more chains" than the western part of South Side, which she describes as having more "individual stores."

Although nearly everyone understood and could explain the differences between LH8 and LH9, there was less agreement about whether the *split* between LH7 and LH8 was valid. For instance, Sara mentioned that the difference between LH8 and LH9 made sense to her, but she did not know the difference between LH7 and LH8. On the other hand, Kara, who has lived both on the western end of Carson (LH7) and on the more eastern parts (LH8) noted that it feels "a bit more isolated" around 23rd making her feel "less safe." She elaborated:

> Whenever I was living down on 15th Street [LH7] I had to worry about drunk people following me home, but on 23rd [LH8] I need to worry about people trying to mug you... so it's different. It's not something I had anticipated, but there is a distinct difference between the two areas of the South Side.

As Kara notes, although the difference is not very prominent, the division by the algorithm displays a subtle difference that can be attributed to the type of people and business in each of these parts.

Moreover, those that did notice a shift between LH7 and LH8 described the street as being narrower and the buildings closer together in LH7. Zach, who is a 30 year old technology consultant and who used to be a cab driver in Pittsburgh explained "from an urban standpoint it is a lot tighter on the western part once you get west of 17th or 18th [LH7]." The added density of bars and restaurants west of 18th makes LH7 more appealing to those visiting it for the nightlife.

LH6 has a completely different story to it. This area contains the only grocery store (Giant Eagle) in South Side. The Giant Eagle is located in a medium sized strip-mall that attracts a demographic that, as noted by our subjects, is distinct from the rest of South Side. As Sara explains:

> There is this interesting mix of people there I don't see walking around the neighborhood. I think they are coming to the Giant Eagle from lower income neighborhoods...I always assumed they came from up the hill.

Kara also expressed the same sentiment. When asked who it is that visits LH6, she said that it is "people that live up on the slopes maybe even towards Carrick," which is another municipality to the south. The related Livehoods for LH6 verified their assumptions, showing a wide area spanning several communities in the hills to the south.

Discussion

In this work we present a clustering model for mapping a city based on the collective behaviors of its residents. By analyzing patterns of people's movements through the city, our approach offers a way to visualize and investigate the on-the-ground dynamics, structure, and character of a city on a large scale. Assuming that both people and places define the character of an area, our results portray a dynamic, almost live, view of the social flows of people throughout the different parts of a city–the Livehoods.

We identify three dispersion patterns that describe the relationship between city neighborhoods and Livehoods: *split*, *spilled* and *corresponding*. Based on our interviews, we find different local dynamics that each of the patterns could possibly represent. *Split* patterns often show the different demographics or different functions that operate in a certain area. *Spilled* patterns typically reveal areas that are in transition, or borders that are in flux. Finally, *corresponding* patterns indicate the strong influence municipal borders and geography have over local social interactions. In the following section we will examine some of the factors that shape the city and show how they translate to our mapping and dispersion patterns.

Municipal Neighborhoods Borders: Contrary to the strict and largely fixed neighborhood borders set by the city government, Livehoods are dynamic, and evolve as people's behaviors change. City neighborhoods borders predominately serve as a way to make order in the chaos of the urban ecosystem. As Justin Miller, a senior planner in Pittsburgh City Planning office explains:

> I need things organized because we have a functional role here...We have to allocate resources and there are a lot of dollars attached to those boundaries...in a lot of the cases, one side of the street is going to qualify for CDBG and the other side is not.[5]

[5] A program of the US government that provides Community Development Block Grants to local communities in need.

These arbitrary borders, set by the city urban planners based on census tracts and geographic landmarks such as roads and bridges, play an important role in the allocation of resources and the planning of local development. But as can be seen from our results, these borders only partially represent the different areas of the city.

In several cases, the Livehoods boundaries *corresponded* perfectly with the municipal borders indicating the strong role that neighborhoods do play in shaping people's activity (e.g. between LH3 and LH4 at 40th Street). However, in some cases, Livehoods *spilled* across the borders between two or more neighborhoods. For example we can see LH2, which *spilled* across the border between East Liberty and Shadyside. In this case, the crossover indicated a shift in peoples' behaviors and perceptions of that area, due to a concerted effort of developers to blur the lines between what were once two very different neighborhoods. In other cases, a single neighborhood may be *split* into several Livehoods. As we saw with LH6, LH7, LH8, and LH9, each had their own character as defined by the demographic mix of local residents and visitors.

Demographics: In our interviews, we found strong evidence that the demographics of the residents and visitors of an area often played a strong role in explaining the divisions between Livehoods. As mentioned above, South Side was *split* by the algorithm into 4 different Livehoods. Our interviewees characterized each of these differently based on the type of people who visit them. For example, LH9 was described as a newly developed area harboring national chain stores in contrast to the more local, mostly night-life oriented area of LH7–each attracting different demographics.

In addition, the lack of both users and venues data for certain areas provides another way of tracing its demographics. For example, The Hill District, one of Pittsburghs poorest neighborhoods did not appear at all in the mapping although it occupies a large area in the heart of the city. The area, that is mainly inhabited by low income, predominately black residents, lacks any representation in our mapping thus implying a the low rate of smartphone usage, and providing a possible depiction of the digital divide.

Development and Resources: Economic development can affect the character of an area. The *spilled* mapping of LH2 captured the social effects the developments of East Side had over the neighboring areas of East Liberty and Shadyside. By visualizing the flow of people between the two once conflicting areas, the algorithm identifies the implications that the economic development had for residents and visitors of the place.

Similarly, the resources (or lack there of) provided by a region has a strong influence on the people that visit it, and hence its resulting character. The *split* area of LH6 in the South Side, which serves as a grocery shopping hub for the communities south of Pittsburgh highlights the distinct single-purpose of that area, and therefore distinguishes it from the surrounding Livehoods.

Geography and Architecture: The flow of people through the streets of a certain area is shaped by the geography and the architecture of the place. We discovered that Livehoods can reveal this influence and the effects it has over peoples visiting patterns. For example, the algorithm created a division between LH7 and LH8 in South Side. This division was unclear to some participants, but others noticed the subtle change in building density on either side of the division. This subtle change effects the kinds of businesses that are on either side, and in turn effects the behavior of people visiting these areas.

Limitations and Biases: First, by aggregating the behaviors of many people, the algorithm itself may be prone towards a "majority" bias consequently misrepresenting or hiding behaviors in the minority. Moreover, our data are based on a limited sample of check-ins shared on Twitter and are therefore biased towards the types of places that people typically want to publicly share. The demographic of foursquare users, which is usually characterized as young professionals in the ages between 25 and 35, owners of smartphones and urban residents can also influence our results. Tuning the clusters is non-trivial and may lead to experimenter bias which joins the possible "confirmation bias" of the interviewees. This mapping might also contribute to local segregation and create a "geo-fencing" effect that will perpetuate separation between demographics.

Despite the biases and limitations, the results we present in this work hold strong. We therefore feel the majority of these limitations are artifacts of the data and not limitations in our methodology. We believe that applying the techniques we introduce to a larger less biased sample in future research may provide an even more accurate representation of the urban landscape.

Conclusion

Livehoods help us reconceptualize the dynamics of a city based on the social media its people generate. We show how the Livehood mappings not only present known divisions but they also reveal subtle changes in local social patterns and the effects they have on the character of the city. Livehoods allow us to investigate and explore the various factors that come together to shape the local dynamics of a city, including municipal borders, demographics, development, resources, geography and architecture. Each Livehood tells a different story of how these factors manifest themselves on peoples' behaviors.

Acknowledgements

This work has been supported by NSF grants CNS-0627513, CNS-0905562, and CNS-1012763, by CyLab at Carnegie Mellon under grants DAAD19-02-1-0389 and W911NF-09-1-0273 from the Army Research Office, the Alfred P. Sloan Foundation, and by CMU/Portugal ICTI. Additional support has been provided by Google, Nokia, and Pitney Bowes.

References

Chang, J., and Sun, E. 2011. Location3: How users share and respond to location-based data on social. In *ICWSM*.

Chen, Z.; Caverlee, J.; Lee, K.; and Su, D. Z. 2011. Exploring millions of footprints in location sharing services. In *ICWSM*.

Cramer, H.; Rost, M.; and Holmquist, L. E. 2011. Performing a check-in: emerging practices, norms and 'conflicts' in location-sharing using foursquare. In *MobileHCI '11*. ACM.

Cranshaw, J., and Yano, T. 2010. Seeing a home away from the home: Distilling proto-neighborhoods from incidental data with latent topic modeling. In *Proceedings of the NIPS Workshop on Computational Social Science and the Wisdom of Crowds*.

Cranshaw, J.; Toch, E.; Hong, J.; Kittur, A.; and Sadeh, N. 2010. Bridging the gap between physical location and online social networks. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, Ubicomp '10, 119–128. New York, NY, USA: ACM.

foursquare. 2011. Foursquare year-end statistics. http://foursquare.com/about.

Gordon, E., and de Souza e Silva, A. 2011. *Net Locality: Why Location Matters in a Networked World*. John Wiley & Sons.

Jacobs, J. 1992. *The death and life of great American cities*. New York: Random House.

Lindqvist, J.; Cranshaw, J.; Wiese, J.; Hong, J.; and Zimmerman, J. 2011. I'm the mayor of my house: examining why people use foursquare. In *CHI 2011*.

Luxburg, U. 2007. A tutorial on spectral clustering. *Statistics and Computing* 17:395–416.

Lynch, K. 1992. *The image of the city*. MIT press.

Milgram, S. 1977. *The Individual in a Social World: Essays and Experiments*. London: Longman Education.

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

Noulas, A.; Scellato, S.; Mascolo, C.; and Pontil, M. 2011. Exploiting semantic annotations for clustering geographic areas and users in location-based social networks. In *Workshop on the Social Mobile Web at ICWSM 2011*.

Oldenburg, R. 1989. *The great good place*. New York: Paragon House.

O'Toole, C. H. 2010. *Slumbering Pittsburgh Neighborhood Reawakens*. The New York Times.

Putnam, R. D. 2000. *Bowling Alone: The Collapse and Revival of American Community*. Simon & Schuster.

Rainie, L., and Wellman, B. 2012. *Networked: The New Social Operating System (forthcoming)*. The MIT Press.

Schwartz, R. 2012. The networked familiar stranger: An aspect of virtual and local urban anonymity. In Cumiskey, K., and Hjorth, L., eds., *Seamlessly Mobile? Mobile Media Practices, Presence & Politics*, volume (under review). Routledge.

Shi, J., and Malik, J. 1997. Normalized cuts and image segmentation. In *CVPR '97*.

Suttles, G. 1973. *The social construction of communities*. Studies of Urban Society. University of Chicago Press.

Whyte, W. H. 2001. *The Social Life of Small Urban Spaces*. Project for Public Spaces Inc.

Zelnik-manor, L., and Perona, P. 2004. Self-tuning spectral clustering. In *NIPS*, 1601–1608. MIT Press.