# Generating Neighborhood Guides from Social Media

**Dan Tasse, Jennifer T. Chou, Jason I. Hong**
Carnegie Mellon University
dantasse@cmu.edu, jtchou@andrew.cmu.edu, jasonh@cs.cmu.edu

### Abstract

What are the neighborhoods in this city like? This is a question that movers and travelers ask all the time, and despite the information sources on the internet, there are few neighborhood guides to help people answer it. We describe formative research, including the Twitter Neighborhood TF-IDF map and an interview-based study of 17 recent movers and travelers. Based on this research, we then describe plans for construction of neighborhood guides based on social media posts in each neighborhood and publicly available data.

## Introduction

How do people who are moving to a new city find the right neighborhood to live in? Similarly, how do travelers find a good area of a city to stay in? In the age of the internet, we can check movie reviews on IMDB, comparison shop for new goods on Amazon, look up restaurants on Yelp, but for this crucial task of finding where to stay or live, we are lacking guidance. We can look up prices and statistics, but what people really want to know before moving to, or staying in, a neighborhood is what it feels like.

As an example, consider the question, "Does this place feel safe at night?" A straightforward approach might be to look up crime statistics, but these statistics certainly do not tell the whole story. Most statistics are overly broad: they lump together bar fights and armed muggings, or random attacks and premeditated violence. Even detailed crime information, though, doesn't answer the question of whether a place *feels* safe. However, a series of photos showing either people enjoying the outdoors at night could easily get across a neighborhood's feeling of safety. Similarly, derelict buildings and abandoned streets could show that a place feels unsafe.

Or consider "Does this place feel like a fun place to go out?" Yelp and Foursquare can show where there are a lot of bars, restaurants, music venues, and other "fun" places, but the same venues that seem fun to a 21-year-old college student could seem boring to a 30-year-old professional. Even within the same age group, some people might want a rowdy sports bar, some might want a DIY punk neighborhood, and others want a swanky area to see and be seen. "Fun" is hard to understand programmatically, but one can often understand exactly whether a place looks "fun" to them by seeing what people are doing and talking about there.

Current guides do not help people answer these questions. Current resources to learn about neighborhoods come in a few forms: guidebooks, point-oriented guides, and statistics. Guidebooks like Lonely Planet [1] and Frommer's [2] focus on the tourist experience with an emphasis on sights to see. Point-oriented guides like Yelp [3] and Foursquare [4] appeal more to locals, but still only point out individual places: restaurants, bars, and other local businesses. Statistics, meanwhile, are growing in popularity: many cities are creating open data portals like the Western Pennsylvania Regional Data Center [5] and SF OpenData [6]. However, these also give partial pictures of the cities they represent. None of these sources can tell people about the "feel" of a neighborhood overall.

One exception is the AirBnB Neighborhood Guides. These guides, produced by the peer lodging site AirBnB, are explicitly designed to reveal the character of each neighborhood as a way to encourage travelers to book an AirBnB guide there. However, the AirBnB model has one major flaw: the guides are manually generated. AirBnB has guides for 25 of its most popular destinations, but they operate in over 30,000 cities; it would be incredibly expensive to write and maintain guides for that many cities.

We aim to create a new model of neighborhood guide by using public social media posts. Every day, users post millions of pictures, tweets, reviews, and other types of social media publicly online. Taken as a whole, these resources can give movers and travelers a reasonable picture of what a neighborhood feels like. This approach will scale better than manual writing, provide a more accurate sense of feel than point-based guides, and be more engaging and informative than raw statistics.

These guides will be useful beyond movers and travelers,



[1] http://www.lonelyplanet.com
[2] http://www.frommers.com
[3] https://www.yelp.com/
[4] https://foursquare.com/
[5] http://www.wprdc.org/
[6] https://data.sfgov.org/

as well. City officials aiming to create entertainment districts can scan summaries of social media to see if the neighborhood feels like they hope. Public safety officials can understand if a neighborhood feels as safe as they hope.

In this paper, we describe our ongoing work to build useful web-based neighborhood guides. We describe our first guide application, the Twitter Neighborhood TF-IDF map; our interviews with 17 participants about how they find neighborhoods to move and travel to; and our results and promising directions for neighborhood guide development.

## Related Work

We have organized related work into two categories. The first looks at recommendation systems involving places. The second category involves systems that make sense of places. We explain why our research falls in the second vein and how we will build from it.

### Recommendation Systems involving Places

We first want to address the question, "Why create neighborhood guides? After all, numerous attempts have been made to recommend places and trips, especially to tourists." This is true: given recent research, it is possible to recommend tourism routes (Kurashima et al. 2010; Okuyama and Yanai 2013), points of interest (Majid et al. 2012; Gao et al. 2010), restaurants (Horozov, Narasimhan, and Vasudevan 2006) and shops (Takeuchi and Sugimoto 2006).

However, these are incomplete solutions for multiple reasons. First, they primarily address tourists. House and apartment hunters are not likely to trust a recommendation algorithm unless they know why it has recommended that particular apartment for them. Second, these recommendation systems do not even solve tourists' problems. Given modern tourists' desires to "experience and feel a part of everyday life" (Maitland 2010) and play an active role in co-creating their tourism experiences (Bock 2015), even the best recommendation system would leave a traveler feeling like something was lacking. Instead of recommending a place for a tourist or mover to go, our work aims to help both of these groups to better develop their understanding of multiple areas.

### Understanding Cities with Social Media

Another line of research has aimed to describe areas using social media posts in them. A lot of this work has been in the area of summarizing photos, which is natural because displaying photos is difficult. If a service hopes to show geotagged photos on a map, it must do some summarization, because viewing the whole set would be impossible. This issue has led to projects such as the WWMX (Toyama, Logan, and Roseway 2003), World Explorer (Ahern et al. 2007), and the Geographical Hierarchical Model (Kafsi et al. 2015).

Summarizing textual content, like tweets, is somewhat easier because there is less total information, so one can use a simple method like a word cloud (at least as a supplementary tool) to get a sense of a large corpus of words (McNaught and Lam 2010). More intelligent methods have been used for tweets, for tasks like event detection (Krumm and Horvitz 2015). Importantly for neighborhoods, though, Hao et al (2010) approach high-level neighborhood modeling in another interesting manner, creating Location-Topic Models based on what users write in travelogues.

Other works have envisioned building higher-level concepts to understand areas. People have a natural vocabulary of words like "hipster", "frat", "tourist", or "yuppie" to describe neighborhoods quickly. This approach works reasonably well for Yelp Wordmaps [7], which pull out those and other words from restaurant and other venue reviews.

One high-level way to look at neighborhoods has been to find where the actual neighborhoods are. The flow of people throughout neighborhoods is often not reflected in the official neighborhood divisions, but recent work has been able to find boundaries based on human behavior such as Foursquare checkins (Cranshaw et al. 2012; Zhang et al. 2013) or tweets (Wakamiya, Lee, and Sumiya 2012). We find this a useful and complementary approach to our work. We want to focus on the next stage: after one has determined where the neighborhoods are, what are those neighborhoods like?

Finally, another promising approach to describe cities at a high level is to compare neighborhoods to existing neighborhoods. This approach, popularized in articles and blog posts like (Read 2014), seeks to help people understand neighborhoods in a new city using the neighborhoods framework they already know. (Le Falher, Gionis, and Mathioudakis 2015) presents a more principled approach to this task: instead of just asking people what similar neighborhoods are, they compare vectors of Foursquare venues and checkins. This also limited their approach, however, by reducing the complexity of a neighborhood into just the businesses that happen to be there. Their approach was also limited by their focus on only eight neighborhoods that are easy to characterize ("Fancy shopping neighborhood", "Gay neighborhood"). We appreciate their approach and plan to work further in this direction.

All of this work will play important roles in the guides we plan to develop. We will discuss direct applications in future sections.

## Twitter Neighborhood TF-IDF Map

Inspired by this work, we created our first attempt at a neighborhood guide, the Twitter Neighborhood TF-IDF Map (figure 1). This map shows which terms are used more often in one neighborhood than in others. For example, the Pittsburgh Pirates baseball team's hashtag `#pirates` is tweeted in many neighborhoods in Pittsburgh, but it is most often used near the baseball stadium.

For our data set, we gathered tweets in Pittsburgh from December 2014 to August 2015. We then used the TF-IDF algorithm to determine which words should be displayed. For each term in a neighborhood, we determined how many times it was used in that neighborhood (TF) and divided by the number of neighborhoods it is used in (IDF, or "inverse document frequency"). The highest-scoring terms were displayed. We omitted words terms that were tweeted by fewer

[7] http://www.yelp.com/wordmap/sf

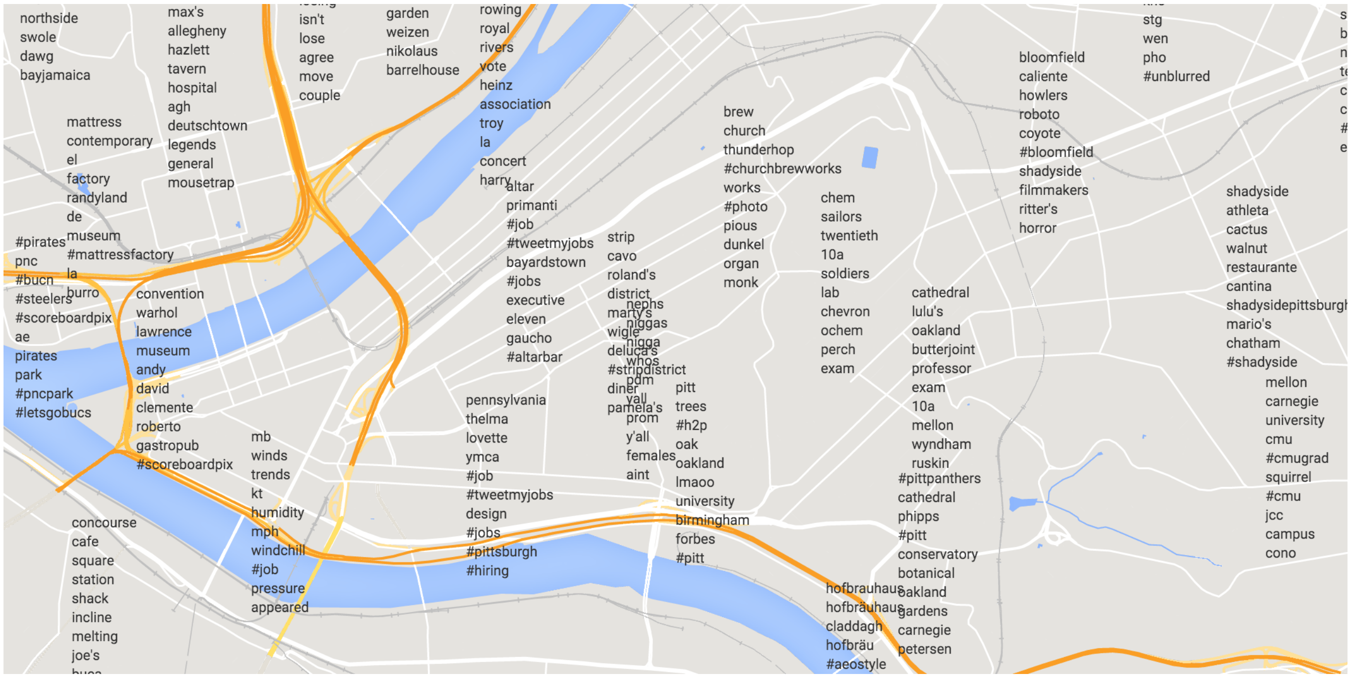


Figure 1: Twitter Neighborhood TF-IDF map. Over each neighborhood, we overlay the 10 words that are used more often in that neighborhood than others. One can easily find popular venues like the Church Brew Works (top center-right) and where the chemistry classes are taught at the University of Pittsburgh (middle center-right).

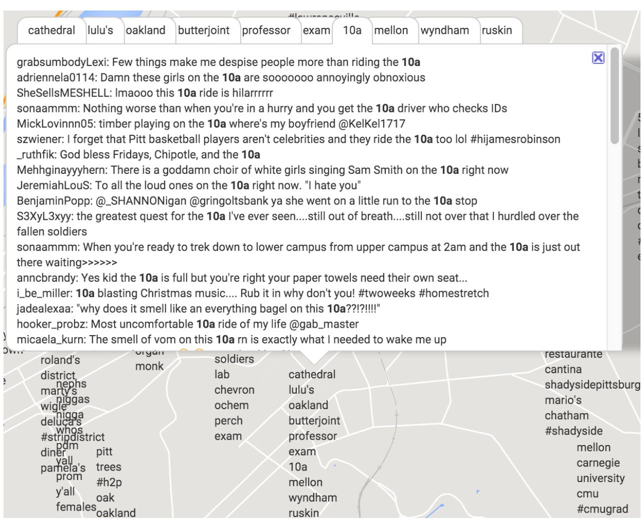


Figure 2: Detail of Twitter Neighborhood TF-IDF map. Clicking on the Oakland neighborhood, where the University of Pittsburgh is located, shows that Twitter users there talk about topics like "professor" and "exam", and even joke about the 10a campus shuttle.

than 5 people to minimize spammers who tweet the same word repeatedly.

This map, along with context popups that displayed on a click (figure 2), provided quick insights into the neighborhoods of Pittsburgh, and showed that prominent neighborhood characteristics (like the University of Pittsburgh in the Oakland neighborhood) were accurately reflected in tweets. While this was a promising start to building useful neighborhood guides, we needed some more guidance for our future developments, so we continued by conducting some research into how people currently understand neighborhoods.

## Build-a-Guide Interviews

To understand which of these data sources will be the most relevant, we conducted interviews with recent movers and travelers. The key questions we wanted to know were:

- What do people want to know about neighborhoods when they're moving?
- What do people want to know about neighborhoods when they're traveling?
- What do people wish travelers and movers knew about their neighborhood?
- What parts of public social media will be most useful?

We recruited 17 participants in Pittsburgh who all recently traveled or moved by posting our study on Reddit, Craigslist, and Facebook. We asked them to describe their experience finding a neighborhood to stay or live. We then asked them to create a paper guide by cutting and taping materials about

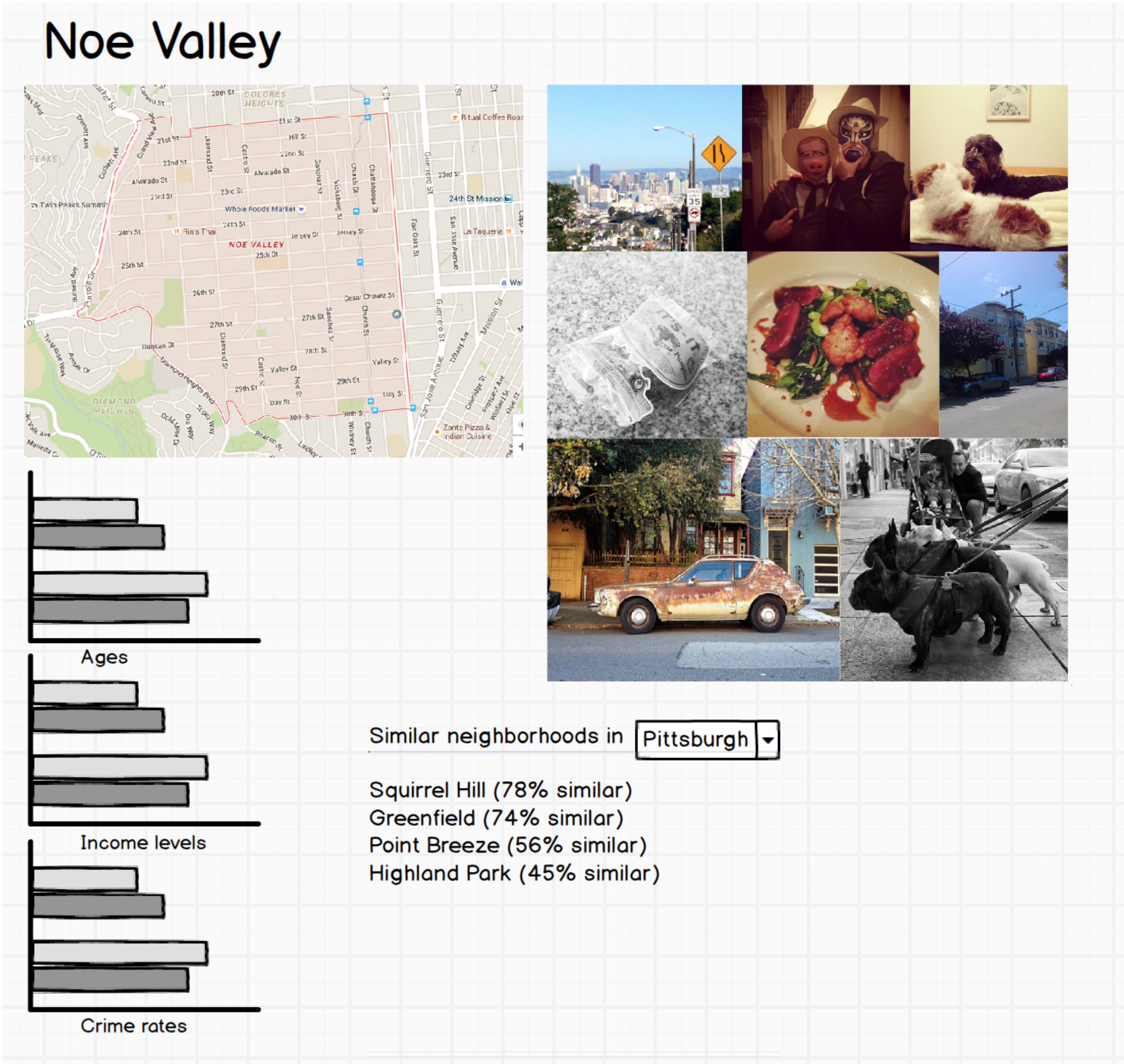


Figure 3: Mock-up of potential neighborhood guide site. Our plan foregrounds three important features: relevant photos, statistics with context, and comparisons to other neighborhoods built from social media posts.

that neighborhood that we printed out ahead of time, showing the most useful and authentic information about that neighborhood. Finally, we asked them to do the same for the neighborhood they currently live in. This study was approved by our university's Institutional Review Board. Analysis of the interview data is ongoing, but we were able to draw out some preliminary findings.

We found that neighborhoods were important to everyone moving or traveling. Most people focused on other constraints first (closeness to a job, school systems, or housing quality), but neighborhood feel was usually almost as important, if not equally important, as these. Both movers and travelers had proximity constraints: movers usually wanted to be close to their jobs, while travelers often wanted something close to downtown or train stations.

Surprisingly, the criteria they used beyond convenience was remarkably consistent. Almost all mentioned safety, as well as "having interesting things around". They wanted an active street life, lots of local businesses, diversity of population, and walkability. The travelers, too, all expressed similar desires: to "live like the locals do" and "get a sense of the place." They all enjoyed visiting "cool" neighborhoods more than seeing tourist sites. While this fits in with modern urban tourism work such as (Bock 2015), we were even surprised that our interviewees were so consistent.

Analysis is ongoing. We will use an open coding approach inspired by grounded theory to have a more principled look at this data. We also plan to recruit around 20 more participants in San Francisco, to get a more diverse cross section of people and talk about neighborhoods in different environments. Despite this, we have still formulated some plans for development of useful neighborhood guides, which we will describe in the next section.

## Neighborhood Guide Development

We plan to build neighborhood guides as a web application. People traveling and moving tend to do some research; not everything is done while mobile. In addition, neighborhood guides could involve a lot of information, which would be difficult to understand on a small screen. As one participant stated, trip planning is inefficient on a phone, compared to a laptop. In this section, we discuss some of the components we hope to include in these neighborhood guides.

### Flickr photo summary

Every participant, when asked to create guides of their neighborhood, included at least one photo from the Flickr photos we printed. This is not surprising; images are quicker to understand and more likely to evoke feelings than text. People often had strong feelings, though, about *which* photos to include. Participants said things like "This octopus sculpture feels very Friendship, and it wouldn't be right for Squirrel Hill", referring to two different Pittsburgh neighborhoods. One participant, who loves living in the Lawrenceville neighborhood, went so far as to say, "If you had shown me these photos of Lawrenceville, I never would have moved there!" Therefore, this cannot be a simple collage of images; they must be selected carefully to accurately represent the essence of the neighborhood. We will refer to summary algorithms as in (Jaffe et al. 2006) and (Ahern et al. 2007).

### Statistics with comparisons

Statistics are important ways to view neighborhoods. Many participants used statistics in their guides to describe some part of the city (such as "it's mostly middle-aged people" or "yeah, low incomes, mostly students"). Also, many people described wanting to know how safe and how walkable a neighborhood is. Safety could be reasonably estimated from city crime statistics, while walkability could be shown using data from Walkscore[8]. Population density would also be a good way to show whether the place will feel more like Midtown Manhattan, Pittsburgh, or a small town, as our participants mentioned this being an important factor.

Of course, appropriate context is important, so we would compare these statistics with nearby neighborhoods and with the city as a whole. Interviewees often had trouble comprehending statistics such as population density, because it is not a statistic that most people interact with in their daily lives. Comparing neighborhoods to other nearby neighborhoods, other neighborhoods in other cities, and the city as a whole would be worthwhile.

### Comparisons to other neighborhoods

Many of our participants talked about neighborhoods by comparing them to other neighborhoods they knew. For example, "Lawrenceville is the Williamsburg of Pittsburgh" or "St. Laurent is like part Squirrel Hill, part Shadyside." This seemed a powerful metaphor and an easy way for people to conceptualize a neighborhood.

We would have to use a different method than (Le Falher, Gionis, and Mathioudakis 2015), though, because neighborhoods cannot be described completely by their venues. Participants in our preliminary research did add some Yelp venues to their guides, but they were not a primary focus, and few added venues that they were not prompted to add by the printout.

Furthermore, transparency is an important part in any neighborhood comparison. Neighborhoods are multidimensional entities, so two neighborhoods that are similar in one way might be quite different in another. Knowing that Lawrenceville, Pittsburgh, and the Mission, San Francisco, are similar would leave a user guessing why: is it because of the trendy bars and restaurants? The recent cost of living increases? The large Hispanic population? The flatness and relative sunniness? Any neighborhood comparison would have to show the reasons for the comparison. In figure 4, we show one way these similarities might be conveyed.

## Conclusion

In this paper, we have described work, both complete and ongoing, in our development build web-based neighborhood guide system. We have described our first attempt, the Twitter Neighborhood TF-IDF map, which has shown us that

[8]http://www.walkscore.com

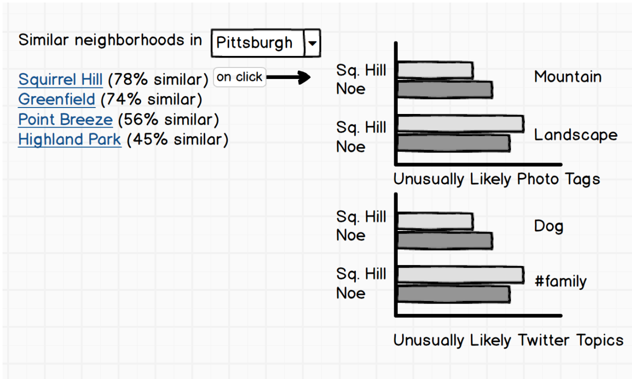


Figure 4: Detail of Neighborhood Comparison subsection. When a user clicks on a term, we will show the most similar neighborhoods and reasons that these neighborhoods are similar.

social media can provide useful descriptions of neighborhoods. We have also described an interview-based study that has led us to develop three key features for neighborhood guides: photo summaries, statistics with context, and neighborhood comparison. These features will help people quickly understand new neighborhoods by providing images, statistics, and references to places they understand. By implementing these guides, we hope to inspire travelers to visit new exciting places, movers to find and become part of great neighborhoods, and city officials to understand the changes that are going on in the smart cities of tomorrow.

## References


Ahern, S.; Naaman, M.; Nair, R.; and Yang, J. H.-I. 2007. World explorer: Visualizing aggregate data from unstructured text in geo-referenced collections. In *Proc. 2007 conference on Digital libraries - JCDL '07*, 1.

Bock, K. 2015. The changing nature of city tourism and its possible implications for the future of cities. *European Journal of Futures Research* 3(1):20.

Cranshaw, J.; Schwartz, R.; Hong, J. I.; and Sadeh, N. 2012. The livehoods project: Utilizing social media to understand the dynamics of a city. *ICWSM*.

Gao, Y.; Tang, J.; Hong, R.; Dai, Q.; Chua, T.-S.; and Jain, R. 2010. W2go: a travel guidance system by automatic landmark ranking. *Proceedings of the international conference on Multimedia - MM '10* 123.

Hao, Q.; Cai, R.; Wang, C.; Xiao, R.; Yang, J.-M.; Pang, Y.; and Zhang, L. 2010. Equip tourists with knowledge mined from travelogues. *Proc. of the 19th International World Wide Web Conference* 1–10.

Horozov, T.; Narasimhan, N.; and Vasudevan, V. 2006. Using location for personalized poi recommendations in mobile environments. *Proceedings - 2006 International Symposium on Applications and the Internet, SAINT 2006* 2006:124–129.

Jaffe, A.; Naaman, M.; Tassa, T.; and Davis, M. 2006. Generating summaries and visualization for large collections of geo-referenced photographs. *Proceedings of the 8th ACM international workshop on Multimedia information retrieval - MIR '06* 89.

Kafsi, M.; Cramer, H.; Thomee, B.; and Shamma, D. a. 2015. Describing and understanding neighborhood characteristics through online social media. In *WWW*.

Krumm, J., and Horvitz, E. 2015. Eyewitness : Identifying local events via space-time signals in twitter feeds. *Proceedings of the 23nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*.

Kurashima, T.; Iwata, T.; Irie, G.; and Fujimura, K. 2010. Travel route recommendation using geotags in photo sharing sites. *Proc. 19th ACM international conference on Information and knowledge management* 579–588.

Le Falher, G.; Gionis, A.; and Mathioudakis, M. 2015. Where is the soho of rome? measures and algorithms for finding similar neighborhoods in cities. In *ICWSM*.

Maitland, R. 2010. Everyday life as a creative experience in cities. *International Journal of Culture Tourism and Hospitality Research* 4(3):176–185.

Majid, A.; Chen, L.; Chen, G.; Mirza, H. T.; Hussain, I.; and Woodward, J. 2012. A context-aware personalized travel recommendation system based on geotagged social media data mining. *International Journal of Geographical Information Science* 27(May 2013):1–23.

McNaught, C., and Lam, P. 2010. Using wordle as a supplementary research tool. *Qualitative Report* 15(3):630–643.

Okuyama, K., and Yanai, K. 2013. A travel planning system based on travel trajectories extracted from a large number of geotagged photos on the web. *The Era of Interactive Media* 657–670.

Read, M. 2014. This is the williamsburg of your city: A map of hip america.

Takeuchi, Y., and Sugimoto, M. 2006. Cityvoyager : An outdoor recommendation system based on user location history. *Ubiquitous Intelligence and Computing* 4159(Figure 1):625–636.

Toyama, K.; Logan, R.; and Roseway, A. 2003. Geographic location tags on digital images. *Proceedings of the eleventh ACM international conference on Multimedia - MULTIMEDIA '03* (November):156.

Wakamiya, S.; Lee, R.; and Sumiya, K. 2012. Crowdsourced cartography: Measuring socio-cognitive distance for urban areas based on crowds movement. *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* 935–942.

Zhang, A. X.; Noulas, A.; Scellato, S.; and Mascolo, C. 2013. Hoodsquare: Modeling and recommending neighborhoods in location-based social networks. In *SocialCom*, 1–15.