The Urban Characteristics of Street Harassment:
A First Look

Mai ElSherief
Dept. of Computer Science
UC Santa Barbara
mayelsherif@cs.ucsb.edu

Elizabeth Belding
Dept. of Computer Science
UC Santa Barbara
ebelding@cs.ucsb.edu

ABSTRACT

Street harassment is a global problem. In this paper, we seek to gain insights into the characteristics of neighborhoods in which street harassment has occurred. We analyze over $7,800$ worldwide street harassment incidents, gathered by the Hollaback project [7], to study the association of street harassment with walkability scores and the number of transit routes in the area surrounding the incident. This unveils a number of key insights. First, we show that more than 50% of the incidents occur in highly walkable areas with walkability scores ranging from 90 to 100, and that nonintuitively, as the walkability score increases, the probability of street harassment events increases. The same result is obtained for areas with high transit scores. Further, the number of transit routes within one mile of the harassment incident has a negative correlation with the number of incidents. The insights gained from our study are a step towards understanding where harassment is likely to occur, which we hope can one day be used for prevention of future incidents.

Categories and Subject Descriptors

H.1.2 [**User/Machine Systems**]: Human information processing

Keywords

Street harassment, Urban analysis, Walkability score, Transit Score, Transit route

1. INTRODUCTION

Street harassment is a worldwide problem; not only is it a frequent occurrence in developing countries, but in many developed countries, such as the U.S., Italy and New Zealand, women are much more likely to feel unsafe on the streets at night than men due to the potential for verbal and/or physical harassment [5]. According to one study [1], 65% of women and 25% of men have experienced street harassment in the United States. This harassment can have numerous

undesirable side effects on victims, such as a reduced sense of safety, anxiety, depression, and refusal to engage in civic life [2]. Efforts to study and address street harassment from a societal point of view include [2] and [5], among others.

Hollaback [7] is a non-profit movement powered by local activists in 92 cities and 32 countries to end street harassment. The Hollaback project collects data on street harassment events worldwide. Through the Hollaback phone app and the online platform, users can report stories of street harassment to share with the Hollaback community. This empowers victims to speak out about everyday harassment and spread the word about the prevalence of these events. In some communities, local governments are informed in real-time about street harassment so that there is a system-wide level of accountability. In addition, the Hollaback app uses GPS to record a data set representing the locations of street harassment events as a means of improving the collective understanding of street harassment and how it can be prevented. As of July 2015, over $7,800$ street harassment incidents have been recorded in their dataset since February 2011. It is on this data set that our work is focused.

In this paper, we use the Hollaback data set to study how users report street harassment stories and analyze the characteristics of the streets where the incidents occur. Our analysis of the data set results in a number of key findings, including:

- The most commonly used words reported in harassment stories are "walking", "man/guy", and "home".
- Street harassment incidents occur more frequently in areas with higher walkability scores [3].
- The most common type of harassment is verbal.
- Street harassment incidents occur frequently along streets with higher transit scores and fewer nearby transit routes (i.e. routes for buses, rail, etc.).

Through our deepened understanding of street harassment events, it is our hope that potential incidents can one day be prevented through, for instance, better route planning to avoid location and time correlations in which events are more likely to occur.

2. RELATED WORK

There are several organizations that fight street harassment by building platforms where users can report incidents, share their stories and interact with others who have gone through similar experiences. Examples include Stop Street Harassment [2], and Hollaback [7], among others. These

Figure 1: Hollaback dataset density map. Darker spots correspond to higher number of incidents.

Figure 2: Histogram of common words in street harassment reports.

platforms aggregate user experiences and some provide a map of harassment incidents.

Our work lies in the area of urban informatics, which is an emerging field that aims to analyze data to understand how cities function and how people behave in response to different issues they face [14]. The field deals with problems related to issues ranging from traffic and morning commute to preparedness for emergencies. Urban informatics data analysis is used to enable more informed planning decisions, which results in more effective city management. For instance, street walkability can have effects on wealth [9] and health [8]. Examples include websites that can be used to learn of neighborhoods with public transit routes, better commutes and healthier lifestyles (e.g., walkscore.com [3] and walkonomics.com). Recommendation of beautiful, quiet and happy routes that can make travel more enjoyable in cities instead of the shortest routes is explored in [11]. Automation of walkability score calculation using social media is presented in [10]. While urban planners are motivated to build walkable streets, [13] shows that adults can be dissatisfied with living on walkable streets due to the association of these streets with more aesthetics-related problems and lower safety.

3. STREET HARASSMENT DATASET

We analyze a dataset of $7,838$ street harassment stories provided by Hollaback [7]. Our dataset spans the period from February 2011 to July 2015. Figure 1 shows a heat map of reported locations during this period. Cities with the highest number of harassment incidents in this dataset include San Francisco, Los Angeles, New York, Boston, Toronto, Buenos aires, London, Berlin, Paris and Rome. Each street harassment entry is composed of a title, type of harassment, a story, report time, a latitude and a longitude. Reports can be updated after initial entry and only indicate the time of the entry or update, not the time the event actually occurred.

3.1 Preliminary Analysis

The Hollaback data set is chosen because it contains multiple components that can help us better understand street harassment. To understand how and where the harassment events take place, we examine the stories for common situational circumstances. Figure 2 shows the top 15 most fre-
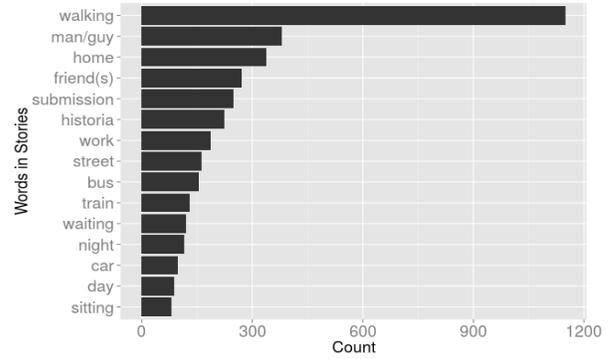
quent words in shared stories. We discard non-descriptive words such as "I, was, a, the, to, my, and, in", among others. We observe that "walking" is the most frequently used word. This leads us to investigate the correlation between street harassment locations and walkability scores, which we discuss in the next subsection. The words "street" and "bus" rank roughly equivalently at eight and nine. This likely indicates that harassment occurs not only along city streets but also on buses. In the next section, we take a closer look at the urban environment surrounding the GPS locations associated with the street harassment reports.

3.2 Urban Analysis

The urban environment around us, whether or not we are consciously aware of it, has a number of effects, both positive and negative. To quantify these effects on human beings, urban informatics researchers have introduced the term "walkability". In his book *Walkable City*, Jeff Speck explains that for a walk to be favorable, it has to be useful, safe, comfortable and interesting [12]. Motivated by "walking" as the most commonly used word in the harassment reports, we pose the following question: **Is street harassment related to walkability?** To answer this question, we use the GPS locations reported in our data set and submit them to the "walkscore.com" web service, which has been used by others in [4, 6]. The "walkscore.com" web service takes a GPS location and returns the walkability score computed for this location. To calculate a walkability score, "walkscore.com" computes the distance to nearby amenities and incorporates pedestrian friendliness and street dimensions.

Figure 3 shows a histogram of the results we obtained. We can draw two important observations from the figure. First, 53.8% of the street harassment events occurred in streets with very high walkability scores, from 90 to 100. Second, the number of street harassment occurrences increases with the increase of walkability score. This suggests that walkability scores are highly correlated with street harassment incidents. We can also observe that there is a slightly greater number of incidents associated with walkability scores from $0 - 10$, which suggests that "unwalkable" streets can be a good medium for harassers, possibly due to the lack of activity/witnesses in these areas.

The significance of this result does not only lie in the positive correlation found between walkability and street harassment. The fact that this data is collected from different cities
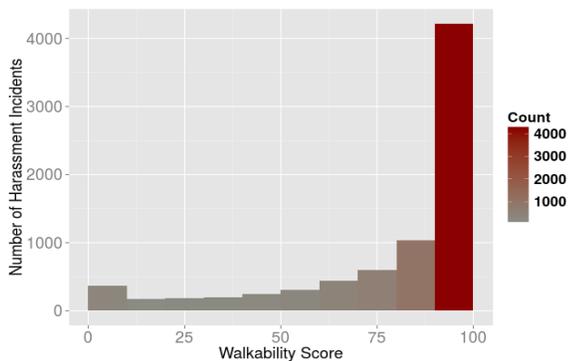
Figure 3: Histogram for Hollaback dataset with respect to walkability scores.

Figure 4: Jitter plot for walkability vs harassment scores.

Harassment Type	Severity Level
assault	5
groping	4
stalking	3
verbal	2
other	1

Table 1: Harassment type mapped to severity level.

Figure 5: Jitter plot for number of routes vs transit scores.

across multiple continents demonstrates the consistency of the results over different parts of the world. Moreover, this result agrees with [13], arguing that walkability is not necessarily positively correlated with adult satisfaction. This indeed opens room for the consideration of other dimensions in the calculation of walkability scores, including safety.

Based on the results in section 3.1, we next seek to determine **whether the degree of severity of the harassment is related to walkability scores**. To answer this question, we annotate each type of harassment with a number depending on its severity as shown in Table 1. The type of harassment is specified by the user reporting the incident through check boxes and he/she may choose more than one type to include. Figure 4 depicts a jitter plot that graphs harassment severity on the x-axis and walkability scores on the y-axis. The figure shows that the dominant harassment type is verbal, constituting approximately 52% of the entries, and occurs across virtually all walkability scores. The other types of harassment tend to occur more frequently in areas with high walkability scores. At any walkability score, the most likely type of harassment will be verbal, but surprisingly, the risk of harassment events is positively correlated with high walkability scores.

Next, we shift our attention to studying the transit properties of the environment surrounding the street harassment reports. By transit properties, we mean the number of transit routes in an area and the quality of service of these routes. To examine the transit properties, we investigate two metrics: the transit score and the number of transit routes. The website "walkscore.com" defines the transit score of a GPS location as a patented measure of how well a location is served by public transit on a scale from 0 to 100. The number of transit routes is a measure of the number of different routes taken by buses, trains and other transit options within one mile of the specified location. In this section, we
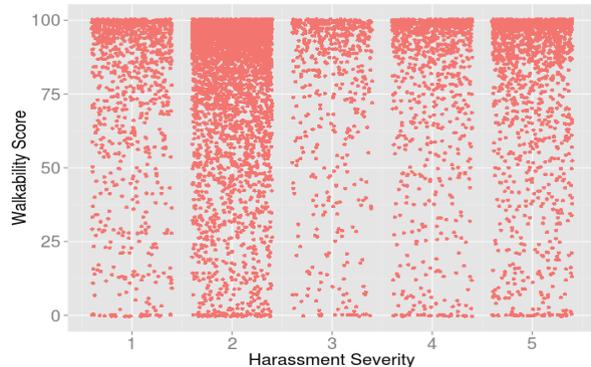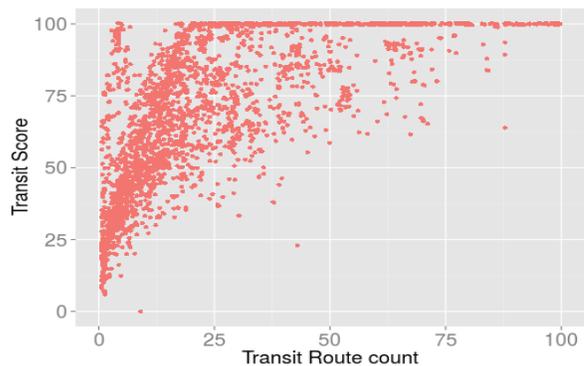
are limited by the cities for which "walkscore.com" has transit information. Thus, our dataset is reduced to 3,289 street harassment entries. It is worth noting that the number of transit routes and transit scores are not directly correlated. An area served by one transit route can have either a very high or low transit score depending on other characteristics such as service level/frequency and the distance to the nearest stop. This is illustrated in Figure 5, which shows that areas with few transit routes can have a wide spectrum of transit scores. However, in general as the number of transit routes increases, so does the transit score.

We then ask the following question: **Is street harassment correlated with transit scores and/or number of transit routes?** To answer this question, we plot a histogram of the transit scores and number of transit routes for our reduced dataset in Figures 6 and 7, respectively. From the trends in Figures 6 and 7, we note that, in general, the better a place is served by public transportation as measured through the transit score metric, the higher the number of street harassment events. Further, locations with fewer transit route options suffer more from harassment.

Based on these observations, we divide Figure 5 into four quadrants. The upper left quadrant, with high transit scores and a low number of transit routes, can be considered the most dangerous zone for street harassment. The upper right and the lower left quadrants have lower probabilities of harassment as they have either high transit scores or low route
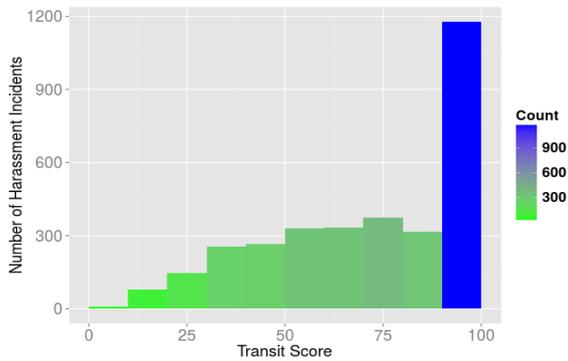
Figure 6: Histogram of $3,289$ incidents with respect to transit scores.
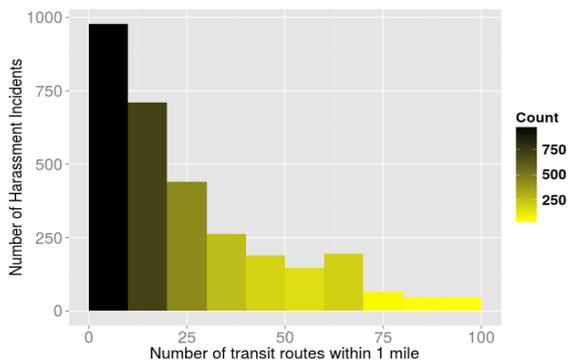

Figure 7: Histogram of $3,289$ harassment incidents based on local transit route availability.

count. The lower right quadrant is considered a safe zone as the probability of experiencing harassment is very low.

4. CONCLUSIONS

In this paper we sought to understand some of the urban characteristics of street harassment incidents. Our analysis shows that street harassment is more common in highly walkable areas with high transit scores and fewer nearby transit routes. On one hand, walkable streets should encourage people to walk more, but on the other hand the sexual harassment rate increases in these areas. While street harassment is considered a crime by law in some countries, other countries have laws that are more tolerant to this behavior. For the countries that criminalize street harassment, the results presented in this paper can be utilized for better targeting of law enforcement. In all cases, we hope that tracking and analyzing street harassment datasets both spatially and temporally can lead to safer route planning for pedestrians.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] Stop Street Harassment. Unsafe and Harassed in Public Spaces: A National Street Harassment Report. `http://www.stopstreetharassment.org/wp-content/uploads/2012/08/2014-National-SSH-Street-Harassment-Report.pdf`, 2014. [Online; accessed July-2015].

[2] Stop Street Harassment. Definitions. `http://www.stopstreetharassment.org/resources/definitions/`, 2015. [Online; accessed July-2015].

[3] Walkscore.com. Walkable Neighborhoods. `https://www.walkscore.com/walkable-neighborhoods.shtml`, 2015. [Online; accessed July-2015].

[4] L. J. Carr, S. I. Dunsiger, and B. H. Marcus. Validation of walk score for estimating access to walkable amenities. *British Journal of Sports Medicine*, 45(14):1144–1148, 2010.

[5] S. Crabtree and F. Nsubuga. Women feel less safe than men in many developed countries. `http://www.gallup.com/poll/155402/women-feel-less-safe-men-developed-countries.aspx`, 2012. [Online; accessed July-2015].

[6] D. T. Duncan, J. Aldstadt, J. Whalen, S. J. Melly, and S. L. Gortmaker. Validation of walk score® for estimating neighborhood walkability: an analysis of four us metropolitan areas. *International journal of environmental research and public health*, 8(11):4160–4179, 2011.

[7] Hollaback. Read and Share Stories. When it comes to street harassment, you are not alone. `http://www.ihollaback.org/share/`, 2015. [Online; accessed July-2015].

[8] R. E. Lee, K. McAlexander, and J. Banda. *Reversing the obesogenic environment*. Champaign, IL: Human Kinetics, 2011.

[9] C. B. Leinberger. Now coveted: A walkable, convenient place. `http://ctmainstreet.org/wp-content/uploads/2012/07/1a-Opinion-Walkable-Place.pdf/`, 2012. [Online; posted May-2012]".

[10] D. Quercia, L. M. Aiello, R. Schifanella, and A. Davies. The digital life of walkable streets. In *Proceedings of the 24th International Conference on World Wide Web*, pages 875–884, May-2015.

[11] D. Quercia, R. Schifanella, and L. M. Aiello. The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, pages 116–125, September-2014.

[12] J. Speck. *Walkable city: How downtown can save America, one step at a time*. NY: Macmillan, 2013.

[13] D. Van Dyck, G. Cardon, B. Deforche, and I. De Bourdeaudhuij. Do adults like living in high-walkable neighborhoods? Associations of walkability parameters with neighborhood satisfaction and possible mediators. *Health & Place*, 17(4):971–977, 2011.

[14] A. Williams, E. Robles, and P. Dourish. Urbane-ing the city: Examining and refining the assumptions behind urban informatics. *Handbook of research on urban informatics: The practice and promise of the real-time city*, pages 1–20, 2009.