Mining the Relationship Between Car Theft and Places of Social Interest in Santiago Chile

Karen Oróstica
Department of Chemical Engineering, Biotechnology and Materials, University of Chile
Santiago, Chile
korostica@ing.uchile.cl

Barbara Poblete
IMFD - Department of Computer Science
University of Chile
Santiago, Chile
bpoblete@dcc.uchile.cl

ABSTRACT

Recent work suggests that certain places can be more attractive for car theft based how many people regularly visit them, as well as other factors. In this sense, we must also consider the city or district itself where vehicles are stolen. All cities have different cultural and socioeconomic characteristics that influence car theft patterns. In particular, the distribution of public services and places attract a large crowd could play a key role in the occurrence of car theft. Santiago, a city that displays drastic socioeconomic differences among its districts, presents increasingly-high car theft rates. This represents a serious issue for the city, as for any other major city, which –at least for Santiago– has not been analyzed in depth using quantitative approaches. In this work, we present a preliminary study of how places that create social interest, such as restaurants, bars, schools, and shopping malls, increase car theft frequency in Santiago. We also study if some types of places are more attractive than others for this type of crime. To evaluate this, we propose to analyze *car theft points* (CTP) from insurance companies and their relationship with *places of social interest* (PSI) extracted from Google Maps, using a proximity based approach. Our findings show a high correlation between CTP and PSI for all of the social interest categories that we studied in the different districts of the Santiago. In particular our work contributes to the understanding of the social factors that are associated to car thefts.

CCS CONCEPTS

• **Social and professional topics** → *Management of computing and information systems*; Systems analysis and design;

KEYWORDS

pattern extraction, car theft, Google Maps

ACM Reference Format:
Karen Oróstica and Barbara Poblete. 2019. Mining the Relationship Between Car Theft and Places of Social Interest in Santiago Chile. In *Companion Proceedings of the 2019 World Wide Web Conference (WWW'19 Companion), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3308558.3316464

WWW '19 Companion, May 13–17, 2019, San Francisco, CA, USA

ACM ISBN 978-1-4503-6675-5/19/05.
https://doi.org/10.1145/3308558.3316464

1 INTRODUCTION

Prior work has established that crimes do not occur randomly in time and space, and that there are specific patterns for each crime class [5]. Through various analysis strategies, such as the mapping of crime hotspots, researchers have been able to identify regions with levels of differential thefts, showing that there are underlying social and spatial characteristics associated with different types of crimes. In particular, some types of places appear to be more attractive than others for certain types of crimes, for example those with a people flow and a large populations. In the case of car thefts, key factors are the number of motivated offenders, of suitable targets, and the lack of security measures, are important variables that could explain this type of crime [8]. However, car theft crimes probably depend on other of factors, such as the place where the crime takes place, the type of car that was stolen and the methods used by criminals [7]. In addition, other factors to be considered are the city or region where the crime takes place, because all cities have different social environments, population, crime statistics, and shopping mall distributions. Shopping malls, in particular, have the characteristic that they have been specifically designed to attract and seduce consumers [6]. In this study, we investigate car thefts in Chile, specifically in its most populated administrative division, the Metropolitan Region of Santiago. Santiago is the economic and social center of the country, being the most populated region, with approximately 7.1 million inhabitants [3]. In recent years, economic growth has been reported in the country, which has been reflected in the increase in the number of motor vehicles that move around the region, which reached 5.19 million in 2017 [2]. In that same year, 12,015 cases of motor vehicle thefts were reported, being one of the regions with the highest number of car thefts in Chile [1]. These statistics are alarming and show that stolen vehicle rates are a problem for Santiago. In particular, there are very few studies that attempt to analyze the problem in the region at a more quantitative level. In this work, we propose to study spatial patterns of car theft crimes in Santiago. Specifically, we focus on analyzing the influence of *places of social interest* (PSI), or of social attraction (e.g., restaurants, schools, and bars) on the concentration of *car theft points* (CTP). We study how places with high car theft frequency relate to nearby social interest points. To achieve this, we study the association between CTP and PSI through statistical and frequency analysis using a proximity-based approach. For this, we model each CTP as a vector in function of their distance to places of social interest, where each dimension represents the frequency of PSIs, for each social category, within a 200 m radius. In the same way, we modelled the PSI as a vector in function of their distance to car theft points, where the first dimension corresponds to the

frequency of CTP, and the second is the social category evaluated. These vectors will be used to compare the distribution of CTP and PSI and evaluate the location the points with high frequencies of CTP and PSIs for each social category.

In particular, our findings show that the CTP and PSI are highly associated, and also we found that there is a co-location between the most frequent points of theft and PSI in Santiago.

Our work focuses on cars theft and its relationship to places that produce social interest, with the following contributions:

(1) We introduce a methodology for identifying important points of car thefts based on the analysis of frequencies and their proximity to points of social interest.
(2) We show that there is a high correlation between CTP and PSI for all social categories in Santiago.

2 DATASET DESCRIPTION

In this section, we describe the datasets used in our study, which corresponds to car theft reports submitted to car insurance companies and to places that produce social attraction in the city according to Google Map[1].

2.1 Car theft points dataset

This dataset consists of all car thefts reports submitted by users to car insurance companies in Santiago, Chile. The dataset was provided to us by the National Association of Car Insurance Companies through a collaboration agreement[2]. This dataset consists of 23, 557 car theft records of crimes that occurred between 2010 and 2017. The original dataset contained records for all of Chile in which the location of the car theft was written in terms of street names. We geocoded car theft locations using the Google Maps Platform[3]. The geocoding process converted an address (i.e., natural language) to geographic coordinates. In particular, we found a higher concentration of stolen vehicles in the regions of Antofagasta, Valparaiso, Bio Bio and Santiago, where the latter presents the largest number of stolen vehicles.

2.2 Places of Social Interest dataset

This dataset concentrates geographical locations of places in Santiago that are likely to attract people on a day-to-day basis. We refer to these locations as PSI. To obtain the list of such places in the city we used the Google Places API [4] from the Google Maps Platform. This API provides the feature to search for places using keywords or by providing a place type. Currently, the API provides a list of 90 place types[5], such as the airport, restaurant, aquarium, art_gallery, shopping mall, among others. From these categories we selected a subset, which in general produce high people flow or might influence positively or negatively the occurrence of car thefts (e.g., shopping malls and police stations, respectively). Specifically, we used 6 categories that were: restaurant, school, shopping mall, bar, bank, and police. Finally, we filtered the remaining data points to keep only those that were within a 200 meter radius of at least one CTP in the Car theft points dataset. Overall, we remained with 127, 123 data points corresponding to places of social interest for our study. Figure 1 shows the total number of places in the dataset for each category.

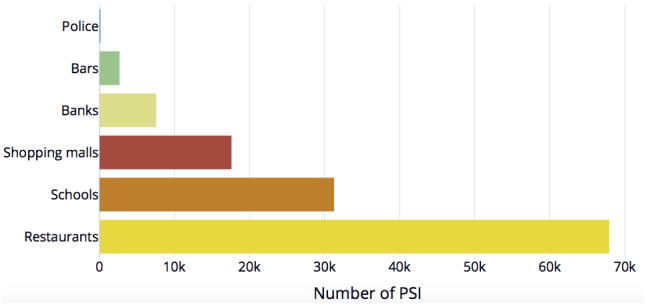

Figure 1: Total number of places of social interest in each category.

3 DATA ANALYSIS

We model each car theft point in function of their distance to places of social interest. This is, each CTP is represented as a vector in which each dimension corresponds to the frequency of social interest points in a particular category within a 200 mt radius.

$$\text{CTP}_{<lat,lon>} =< freq_{category1}, \ldots, freq_{category6} >$$

To model the places of social interest, we decided to represent each PSI as a vector, where the dimensions included to social category and the frequency of car theft points within a 200 mt radius.

$$\text{PSI}_{<lat,lon>} =< freq_{CTP}, Social_{categories} >$$

To obtain the frequencies of CTP and PSI within of specific radius we used the kd-tree algorithm implementation [4] in the R RANN package[6].

Car theft frequency analysis We want to identify geographical locations in Santiago where car thefts occurred more frequently than others places. For this, we obtained geographical locations and its number of car theft in Santiago. In Figure 2 we show geographical locations where the car thefts frequently occur in the city. We represent the points with the lowest frequency of car theft (less than 5) as light yellow dots. Geographic locations with an intermediate frequency of car theft (greater than 5 and less than 50) are displayed as yellow dots. In red, we show geographic locations with highest frequency of car theft (high than 50). we noted that geographical locations are not distributed randomly; on the contrary, we observed that there is a trend towards the northeast sector of the region, which is the wealthiest sector of the city. This result is consistent with what we expected since the data set corresponds to insured cars, which have a higher value than the average of cars. The great majority of the points have a frequency of thefts less than 5, but also there are some places with an high frequency of car thefts. We identify 4 points of car thefts with a frequency of over 50, which are located in the southeast, north, middle and East

[1] http://maps.google.com
[2] Project FONDEF project ID16I10222, CONICYT.
[3] https://developers.google.com/maps/documentation/geocoding/intro
[4] https://developers.google.com/places/web-service/intro
[5] https://developers.google.com/places/supported_types?hl=es
[6] https://cran.r-project.org/web/packages/RANN/index.html

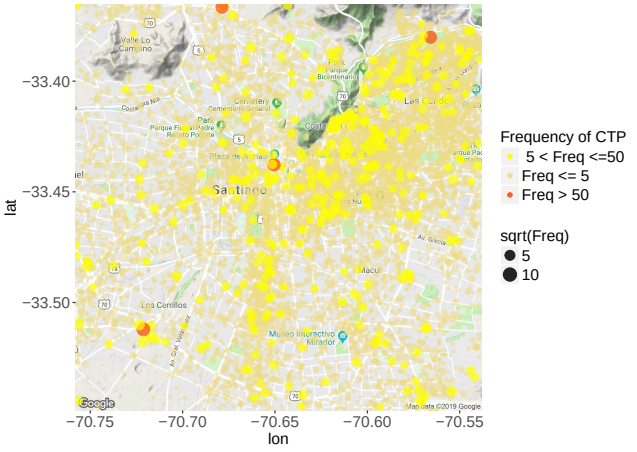

Figure 2: Frequency categories of car thefts: low (light yellow), intermediate (yellow) and high (red) in Santiago.

of the city. These geographical locations are very close to the main shopping centres in Santiago. These shopping centres correspond to *Mall Plaza Oeste*, *Mall plaza Norte*, *Plaza de Armas* and *Mall Alto las Condes*, respectively.

Identification of hotspots based in CTP and PSI vector To understand the relationship between car thefts and their proximity to places that attract large numbers of people, we characterize the car theft points (i.e, geographic locations) were near places of social interest with a high frequency of car thefts. To evaluate this idea, we created histograms for all dimensions of CTP and PSI, where we categorized each CTP and PSI into ranges according to its frequency. We selected CTP with a high frequency of PSI, for all dimension, and also, we choose PSI with a high frequency of car theft for all social categories. Then, we show the selected points on the map of Santiago. To the side of the maps, we display the corresponding histograms of CTP and PSI. The Figure 4 shows these results for police (A), restaurant (B) and school (C) categories, and the Figure 5 display the results for bank (A), bar (B) and shopping mall (C). We found that there is a high correspondence between the neighbourhoods recognized by social categories evaluated in this work and the highlighted geographical locations obtained by our method. For example, we noticed that the main shopping malls of Santiago were identified by our method as hotspots of car thefts based on the frequency of CTP and PSI. In the case of restaurant, we

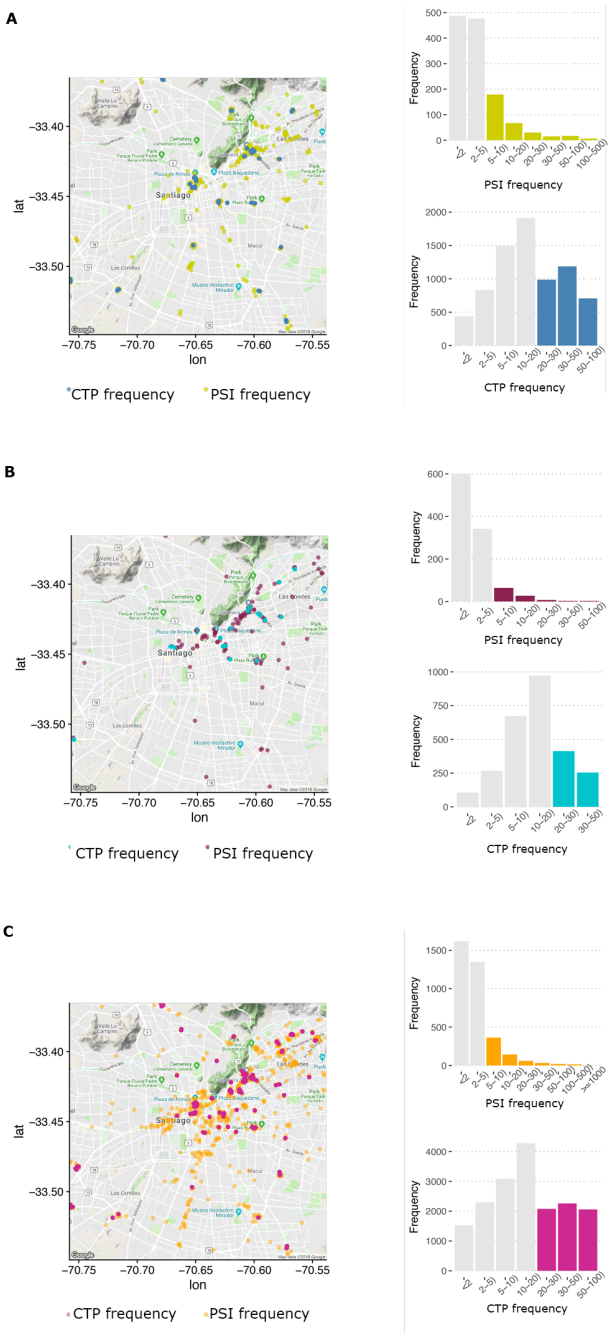

Figure 3: Geographic locations with a high number of PSI and car theft, and the distribution of these points for bank (A), bar (B) and shopping mall (C).

identify neighborhoods of recognized restaurants, such as the*Barrio Italia, Avenida Tobalaba, Barrio Brasil*, with a high frequency of CTP and PSI. The same occurs in the other social categories.

Comparison of the distribution of CTPs and PSIs We obtained the correlation between CTP and PSI agrouped by each

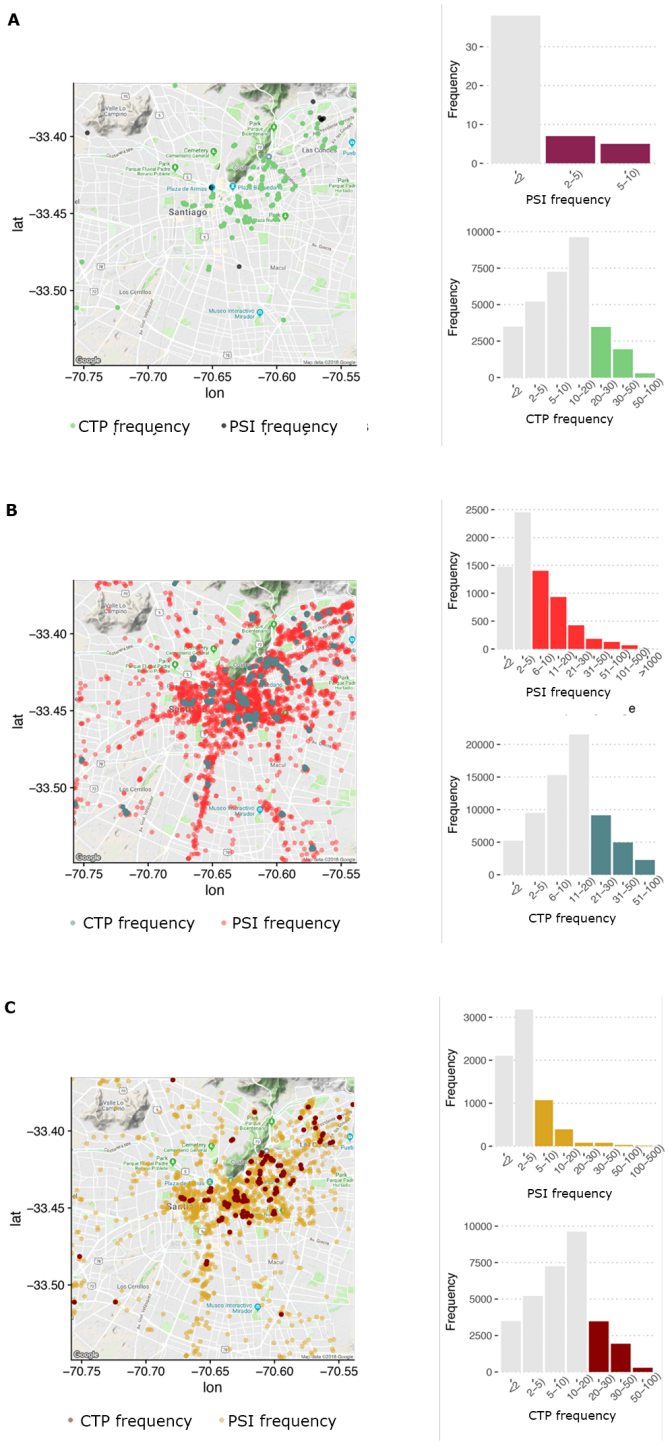

Figure 4: Geographic locations with a high number of PSI and car theft, and the distribution of these points for police (A), restaurant (B) and school (C).

districts of Santiago using Spearman method. We decided to use this administrative division because there is an important economic and social difference between them. We observed a high correlation between CTP and PSI for all social category, where Restaurant reached the highest correlation (r = 0.98) and police got the lowest correlation (r = 0.73) (See Table 1). These results support our idea of insured car thefts are associated with places attractive socially, like restaurants, schools, etc.

Table 1: Correlation between CTPs and PSI

Category of PSIs	Correlation coefficient
Restaurant	r = 0.98 (P = 2.2e-16*)
School	r = 1 (P = 2.2e-16)
Shopping mall	r = 0.97 (P = 2.2e-16*)
Bar	r = 0.97 (P = 2.2e-16*)
Bank	r = 0.96 (P = 2.2e-16*)
Police	r = 0.73 (P =1.3e-0*9)

3.1 Discussion and Conclusion

We study how the thefts of the insured car are mediated by the proximity to places of social interest. In particular, we proposed a new methodology to identify hotspot of car thefts car based on the frequency analysis using information from *car theft points* (CTP) and *Places of Social Interest* (PSI). In summary, we were to able to identify neighbourhoods or streets with a high frequency of CTP and PSI, which are neighbourhoods recognized by social activity, for example, restaurants, schools, bars, among others. In addition, it was seen that there were common hotspots between categories, for example, restaurant and shopping malls.

On the other hand, with the information provided by this new methodology, insurance companies can improve their service, it is even possible to use this information to create a system of recommendation to the user to avoid future thefts of their cars.

ACKNOWLEDGMENTS

Funding support national scholarship conicyt 21182123.

REFERENCES

[1] Carabineros de Chile. 2018. Statistical Report. http://dac.carabineros.cl/Reportes/reporte_noviembre_2018.html. [Online; accessed 22-January-2018].

[2] Instituto Nacional de Estadística de Chile. 2018. Parque de Vehículos. https://www.ine.cl/estadisticas/economicas/transporte-y-comunicaciones?categoria=Anuarios. [Online; accessed 22-January-2018].

[3] Biblioteca del Congreso Nacional. 2018. Región Metropolitana de Santiago. https://www.bcn.cl/siit/nuestropais/region13/index.htm. [Online; accessed 22-January-2018].

[4] Jerome H. Friedman, Jon Louis Bentley, and Raphael Ari Finkel. 1977. An Algorithm for Finding Best Matches in Logarithmic Expected Time. *ACM Trans. Math. Softw.* 3, 3 (Sept. 1977), 209–226. https://doi.org/10.1145/355744.355745

[5] Jerry Ratcliffe. 2010. Crime Mapping: Spatial and Temporal Challenges. In *Handbook of Quantitative Criminology*, Alex R. Piquero and David Weisburd (Eds.). Springer New York, New York, NY, 5–24. https://doi.org/10.1007/978-0-387-77650-7_2

[6] Joel Stillerman and Rodrigo Salcedo. 2012. Transposing the Urban to the Mall: Routes, Relationships, and Resistance in Two Santiago, Chile, Shopping Centers. *Journal of Contemporary Ethnography* 41, 3 (June 2012), 309–336. https://doi.org/10.1177/0891241611434551

[7] Isaac T. Van Patten, Jennifer McKeldin-Coner, and Deana Cox. 2009. A Microspatial Analysis of Robbery: Prospective Hot Spotting in a Small City. *Crime Mapping: A Journal of Research and Practice* 1, 1 (2009), 7–32. http://ivanpatt.asp.radford.edu/Research/Microspatial%2520Analysis%2520of%2520Robbery.pdf

[8] Yuanyuan Mao, Shenzhi Dai, Jiajun Ding, Wei Zhu, Can Wang, and Xinyue Ye. 2018. Space Time Analysis of Vehicle Theft Patterns in Shanghai, China. https://www.mdpi.com/2220-9964/7/9/357/htm