Mining Mobile Phone Data to Investigate Urban Crime Theories at Scale

Martin Traunmueller[1], Giovanni Quattrone[2], and Licia Capra[1]

[1] ICRI Cities, Dept. of Computer Science, University College London
[2] Dept. of Computer Science, University College London
Gower Street, WC1E 6BT, London, UK
{martin.traunmueller.11,g.quattrone,l.capra}@ucl.ac.uk

Abstract. Prior work in architectural and urban studies suggests that there is a strong correlation between people dynamics and crime activities in an urban environment. These studies have been conducted primarily using qualitative evaluation methods, and as such are limited in terms of the geographic area they cover, the number of respondents they reach out to, and the temporal frequency with which they can be repeated. As cities are rapidly growing and evolving complex entities, complementary approaches that afford social scientists the ability to evaluate urban crime theories at scale are required. In this paper, we propose a new method whereby we mine telecommunication data and open crime data to quantitatively observe these theories. More precisely, we analyse footfall counts as recorded by telecommunication data, and extract metrics that act as proxies of urban crime theories. Using correlation analysis between such proxies and crime activity derived from open crime data records, we can reveal to what extent different theories of urban crime hold, and where. We apply this approach to the metropolitan area of London, UK and find significant correlations between crime and metrics derived from theories by Jacobs (e.g., population diversity) and by Felson and Clarke (e.g., ratio of young people). We conclude the paper with a discussion of the implications of this work on social science research practices.

Keywords: Urban crime, telecommunication data, open data, data mining.

1 Introduction

In modern society we are experiencing two phenomena: on one hand, there is a rapid population shift of people moving from rural areas into urban environments, with an annual growth of 60 million new city dwellers every year [29]. On the other hand, crime activities are on the rise (e.g., [5]), especially in densely populated areas [13]. Being able to understand and quantify the relationship between people presence and crime activity in an area has thus become an important concern, for both citizens, urban planners and city administrators.

The relationship between *people dynamics* and *crime* in urban environments has been researched extensively in architectural and urban studies over the last decades, with theories that sometimes appear to conflict with each other. Most influential theories lead back to the 1960's and 1970's: Jacobs [12] suggests that population diversity, activity

L.M. Aiello and D. McFarland (Eds.): SocInfo 2014, LNCS 8851, pp. 396–411, 2014.

and a high mix of functions lead to less crime for an area, whereas Newman [15] hypothesizes the opposite, supporting clear separation of public, semi-public and private areas towards urban safety. Each theory has been evaluated, and indeed supported, by means of qualitative research methods that enable in-depth investigations into the reasons behind certain phenomena. However, such methods are very expensive and time-consuming to run, so that studies are usually restricted to a rather small number of people (relative to the overall urban population) and constrained geographic areas (e.g., a neighbourhood); furthermore, they are almost never repeated over time, to observe potential changes. It becomes thus very difficult to collect sufficient evidence to explain under what conditions a certain theory holds.

In this paper we propose a new method to quantitatively investigate urban crime theories at scale, using open crime data records and anonymised mobile telecommunication data. From the former, we extract quantitative information about crime activity, as it happens across different urban areas of very fine spatial granularity. From the latter, we extract metrics that act as proxies for previously developed urban crime theories that link people presence in an area with crime. We then use correlation analysis between crime data and our defined metrics to validate urban crime theories at scale. We apply this method to data obtained for the city of London, UK, and find that, in this city and at the present time, Jacobs' theory of 'natural surveillance' [12] holds: we discover that age diversity, as well as the ratio of visitors in a given area, are significant and negatively correlated with crime activities; furthermore, Felson and Clarke theory [9] that links a higher presence of young people with higher crime is also confirmed. We believe the proposed method to be a powerful tool in the hands of social science researchers developing urban crime theories, as they can now complement qualitative investigations with quantitative ones: while the former afford them deep insights into the causality of certain phenomena, the latter afford them the ability to scale up findings in terms of population reach, geographical spread, and temporal evolution.

The remainder of the paper is structured as follows: we first provide a brief overview on background theories from architectural and criminological studies, and state-of-the-art follow-up research that has been grounded on them. We then present our method, in terms of the datasets we leverage, the pre-processing and data manipulation we have conducted, and the metrics we have extracted as proxies for urban crime theories. We discuss the results obtained when applying our method to data for the city of London, UK, and finally conclude by discussing implications, limitations and future steps.

2 Related Work

2.1 Background

Most well known architectural theories about the relationship between people dynamics, the urban environment and crime lead back to the studies of Jacobs [12] and Newman [15], with two different schools of thought. Jacobs [12] defines urban population as 'eyes on the street', a natural policy mechanism that supports urban safety through 'natural surveillance'. An open and mixed use environment supports this concept by enabling diversity and activity within the population using the area at different times. While Jacobs suggests that a high diversity among the population and a high ratio of