Paper Title: Mining Mobile Phone Data to Investigate Urban Crime Theories at Scale

Paper Authors: Martin Traunmueller, Giovanni Quattrone and Licia Capra

Link to Paper: https://link.springer.com/chapter/10.1007/978-3-319-13734-6_29

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.

Summary: In this paper, the authors presented a method to relate urban crime and people dynamics in the city of London. They made use of two data sources: crime data records for the area of Greater London and data from a mobile telecommunication provider for details of people dynamics. Crime data was recorded with latitude/longitude coordinates whereas the telecommunication data was available as footfall in grids of varying sizes (smaller grids in central London as opposed to larger grids in less densely populated areas outside central London). The datasets were cleaned and preprocessed. Inconsistent entries and entries for grid cells outside the Greater London area were discarded. Crime data was broadly classified as street crime or home crime. For each area i, two metrics were considered: Crime Count $CC(i)$ and Crime Activity $CA(i)$. The areas considered were of comparable size so $CC(i)$ was taken as a way of measuring crime normalised by area size. Since areas had varying population densities, $CA(i)$ was found by dividing $CC(i)$ by the estimated population of i, $P(i)$ as obtained by the total footfall in area i from the telecommunication data, and was thereby taken as a way of measuring crime normalised by population density. The following six proxies were defined to link people dynamics with Crime Count and Crime Activity for each area: Diversity of People (age diversity), Ratio of Visitors, Ratio of Residents, Ratio of Workers, Ratio of Female Population and Ratio of Young People. Correlations were calculated for each of these six proxies and crime (both street and home). It was found that more Age Diversity in an area was associated with less crime, a high Ratio of Visitors meant high safety in that area, a high Ratio of Residents in an area showed high crime rates in that area, a high Ratio of Female Population in an area statistically correlated to a higher crime rate in that area and a higher proportion of Young People in an area was associated with more crime in that area. The Ratio of Workers in an area however, was not found to have statistical significance on the crime rate. The authors also performed this analysis on the 32 individual administrative boroughs of London to identify which boroughs were significant contributors to the over all crime rate. There were two major limitations in the study performed. One was that the crime data did not have time stamps and was recorded on a monthly basis whereas the telecommunications data recorded footfall on an hourly basis. Hence, the relationship between crime and people dynamics was determined at a coarser level. The second was that the authors used mobile phone data from a single service provider. Although the provider served a large sector of the population, there was still a considerable portion of the population that was left out (those who were users of other service providers or PayAsYouGo options).