



# Residents, Employees and Visitors: Effects of Three Types of Ambient Population on Theft on Weekdays and Weekends in Beijing, China

Guangwen Song<sup>1</sup> · Yanji Zhang<sup>2</sup>  · Wim Bernasco<sup>3,4</sup> · Liang Cai<sup>5</sup> · Lin Liu<sup>1,6</sup> · Bo Qin<sup>7</sup> · Peng Chen<sup>8</sup>

Accepted: 7 November 2021 / Published online: 2 December 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

## Abstract

**Objectives** The residential population of an area is an incomplete measure of the number of people that are momentarily present in the area, and of limited value as an indicator of exposure to the risk of crime. By accounting for the mobility of the population, measures of ambient population better reflect the momentary presence of people. They have therefore become an alternative indicator of exposure to the risk of crime. This study considers the heterogeneity of the ambient population by distinguishing residents, employees and visitors as different categories, and explores their differential impact on thefts, both on weekdays and weekends.

**Methods** We analyze one-year of police recorded thefts across 2104 1 km<sup>2</sup> grid cells in a central area in Beijing, China. Controlling for the effects of attractiveness, accessibility, and guardianship, we estimate a series of negative binomial models to investigate the differential effects of the three groups (residents, employees and visitors) in the ambient population on crime frequencies, both on weekdays and during weekends and holidays.

**Results** Overall, larger ambient populations imply larger theft frequencies. The effect of visitors is stronger than the effects of residents and employees. The effects of residents and employees vary over the course of the week. On weekdays, the presence of residents is more important, while the reverse holds true during weekends and holidays.

**Discussion** The effects of ambient population on thefts vary by its composition in terms of social roles. The larger role of visitors is presumably because in addition to being potential victims, residents and employees may also exercise informal social control. In addition, they spend more time indoors than where risk of theft is lower, while visitors might spend more time outdoors and may also bring about greater anonymity and weaken informal social control.

**Keywords** Ambient population · Residents · Employees · Visitors · Theft · Big data

---

✉ Yanji Zhang  
chairman7up@126.com

Extended author information available on the last page of the article

## Introduction

To monitor, predict and compare crime levels, it is important to know the population at risk of victimization. For spatial units, such as cities or neighborhoods, it has traditionally been assumed that the residential population, the number of people who live in an area, are the population at risk, although this assumption is problematic (Boggs 1965). An alternative indicator of exposure to crime risk is the *ambient population*: the number of people who are present in the during a specific time period (Andresen 2007; Hipp et al. 2019). These people might be the residents of the area, but they could also include people who visit the area but live elsewhere. Measuring the true ambient population at risk has been a key focus in recent studies in criminology (Andresen 2006; Haberman and Ratcliffe 2015; Hipp et al. 2019). The frequency of occurrence of crime depends on an understanding of the population for whom crime is a threat; that is the total number of people who are at the risk of criminal victimization (i.e., the denominator of the crime rate). Traditionally, researchers have used the residential population as the denominator of the crime rate or, in the context of multivariate regression, as one of the predictors of crime frequencies.

However, the local residents are not always the only ones who are at risk of victimization in an area. Most people continuously move beyond their residential area for pursuing important activities (Cagney et al. 2020), and are thus at risk elsewhere (Roman 2005). The activity patterns of residents vary considerably between weekdays and weekends. On weekdays, for example, in many areas large proportions of the resident population work, attend school or pursue other activities away from their homes. During these times, they are not exposed to victimization risk in their own residential area. During weekends, however, large percentages of the population spend much more time at home. At the same time, city centers, generally home to a small proportion of urban residents, are visited by numbers of people that dwarf the number of local residents, in particular on weekends. The residential population thus reflects only part of the population that makes use of the local area, because at predictable days of the week and times of the day they are joined by others who visit the local area for a variety of purposes. Therefore, the size of the local residential population may not necessarily bear a strong relation to the local crime rate, and calculations of crime risk exposure based on residential population may lead to bias (Boggs 1965).

The limited accuracy of residential population as an estimate of risk exposure is aggravated by the way it is usually collected. In most countries residential population measures are based on census data, which are collected at low frequencies. In China, census data are collected every 10 years.

Recent studies have begun to provide empirical evidence against the sole reliance on residential population measures from census data as a useful denominator in the calculation of crime risk. For example, since 2006, Andresen and colleagues carried out a series of studies to explore ambient population estimated from LandScan Global Population Database. Their findings suggest that the ambient population may provide a more reliable estimate of the presence of potential victims than the residential population (Andresen 2006, 2007, 2011; Andresen and Jenion 2010).

In addition to the LandScan approach, three other approaches have been used in the literature to address this challenge. One is to approximate population by the presence of facilities that are known to bring together large crowds, and thereby an increased volume of potential targets and motivated offenders. These facilities have been labeled ‘crime generators’ and ‘crime attractors’ (Brantingham and Brantingham 1995; Kinney et al. 2008). The distinction between both categories is rooted in the nature of the

offender's intention to visit the location. Crime generators are usually visited without a premediated criminal intention, whereas crime attractors are locations that are often visited with a premediated criminal intention. Different types of facilities attract different categories of the population who visit the facilities with varying purposes. Studies following this approach often measure the ambient population by the number of such population-attracting facilities (Bernasco and Block 2011), the co-location of facilities (mixed land use) (Kinney et al. 2008), and the number of employees registered to the facilities (Kim 2018; Wo et al. 2016). By increasing the temporal resolution, recent scholarship finds that facilities exhibit varied magnitudes in its relation to crime in different time periods of a day, thus shedding light on the dynamic nature of the population at risk (Haberman and Ratcliffe 2015).

With recent technological advances, automatically collected geo-referenced and time-stamped data at massive scales ('big location data') emerge as another source for the measurement of the ambient population.

Although not free of potential biases, big location data, such as mobile phone location data and location-based social media data, are much more direct measures of the presence of people than the more indirect proxy measures based on land use. For example, during the COVID-19 pandemic, due to lockdown measures the presence of retail and other facilities is not representative of the size of ambient populations, whereas the locations of smartphones and other mobile devices continue to represent actual locations of their users. Big location data also hold promise for measuring ambient populations at more fine-grained spatial-temporal granularities (Hipp et al. 2019; Malleon and Andresen 2015). LandScan data, for example, provide 24-h estimates per square kilometer (Andresen and Jenion 2010), whereas contemporary big location data can easily provide estimates that are more precise. Some types of big location data can be used not only to automatically measure the presence of people, but also to track their whereabouts over time and thus measure individual mobility in addition to mere presence (Candia et al. 2008; Song et al. 2010).

The third method used to measure ambient population and population mobility is the transportation survey (Boivin and Felson 2017). Compared to survey-sample-based transportation data, the big data approach reports location data with a much wider coverage and refined spatial and temporal resolutions at lower costs.

The abovementioned studies suggest that measures of ambient population bear a stronger relation to crime than traditional measures of residential population, in particular at smaller spatial and temporal scales. However, a limitation of these studies is that they do not distinguish amongst the people that make up the ambient population, and thereby ignore the internal heterogeneity of ambient population. For example, the knowledge of and the strength of attachment to a location will usually vary within the ambient population (Boivin and Felson 2017). Outsiders (who visit the location infrequently for occasional activities) may be less knowledgeable and less attached to the location than employees (who work at the location), and employees may be less knowledgeable and attached than residents (who live at the location). A place with a large proportion of outsiders often features high mobility and anonymity, thereby decreasing local informal social control and potentially increasing crime (Tillyer and Walter 2019). On the contrary, if a place is mainly used by residents or employees, there will be more social control, and potentially less social disorder and less crime (Yu and Maxfield 2014). Few studies have paid attention to the internal heterogeneity of ambient population. As an exception, He et al. (2020) use mobile phone data to distinguish between local and non-local phone users. But they do not make further differentiations based on their activities and attachments, such as between residents, employees, and visitors.

Whereas the residential population of an area is a stable measure that does not change over time in the short run, the ambient population and the composition of the ambient population are likely to vary across weekly cycles, with weekdays displaying different activity levels than weekends. Therefore, to improve our understanding of how ambient populations affect crime, the present study subdivides the ambient population into local residents, employees and visitors, and explores their differential impacts on the distribution of crime during weekdays and weekends<sup>1</sup> in Beijing.

The relation between ambient population and crime will likely depend on the type of crime, because both the motivations and the opportunities for committing crime vary across crime types (Garofalo et al. 1987). In the present study, for two main reasons we focus on theft. First, theft is strongly related to the physical presence of people, which makes it an appropriate type of crime to be linked to ambient population. Second, thefts account for a large part of property crimes in China, and have a great impact on people's daily life. Gaining more knowledge about theft might help develop crime control strategies.

## Literature Review

### Opportunity Theories, Crime Generators and Crime Attractors

Crime is not randomly distributed in space, but tends to be concentrated at different levels of spatial aggregation, such as neighborhoods, street segments (Weisburd 2015) and even addresses (Sherman et al. 1989). To explain these variations, opportunity theories assert that the distribution of crimes is a function of the criminal opportunities. Opportunity theories of crime include routine activity theory, rational choice theory, and crime pattern theory. In the routine activity perspective, the convergence in time and place of suitable targets, motivated offenders, and a lack of capable guardianship leads to the emergence of crime (Cohen and Felson 1979). This perspective highlights the importance of routine activities and the interactions between different stakeholders. The crime pattern perspective suggests that offenders gain their spatial knowledge and awareness of crime opportunities through daily activities. As a result, they often offend in places that they also visit for legal routines activities, such as work, school, shopping or leisure (Brantingham and Brantingham 1995). Rational choice theory argues that offenders take a rational approach to making decisions by balancing the potential benefits, costs, and risks of alternatives when committing crime (Cornish and Clarke 1987). It suggests that offenders prefer to commit crimes where suitable targets are present and capable guardians are absent.

Inspired by these opportunity theories of crime, a sizable criminological literature has been devoted to estimating the frequency of convergences between motivated offenders and potential targets. A major body of empirical work uses *crime generators* as a proxy for the frequency of offender-target convergence, and *crime attractors* for the measurement of predatory activities of potential offenders. Crime generators are “particular areas to which large numbers of people are attracted for reasons unrelated to any particular level of criminal motivation they might have or to any particular crime they might end up committing” (Brantingham and Brantingham 1995: 7), while crime attractors are “particular places, areas, neighborhoods, districts which create well-known criminal opportunities to which

<sup>1</sup> In this paper, we use the word weekends to refer to both weekends and holidays.

strongly motivated, intending criminal offenders are attracted because of the known opportunities for particular types of crime.” (Brantingham and Brantingham 1995: 8). Thus, the main distinction between crime generators and crime attractors is the intention with which potential offenders visit them. Crime generators are visited without a premediated criminal intention, while crime attractors are visited with a criminal purpose in mind.

Crime generators include, but are not limited to, shopping centers, schools, sports and entertainment facilities and public transit hubs. According to crime pattern theory, crime generators bring together crowds and thereby potentially provide a relatively large volume of suitable targets for motivated offenders. As a result, the local crime level is elevated. Indeed, empirical studies demonstrate that land use structure, indicative of the volume of potential targets, is consequential for the local crime volumes (Quick et al. 2019).

As for crime attractors, they attract motivated offenders because they provide known opportunities for particular types of crime. They include bars, prostitution areas, drug markets and youth hangout places. Bernasco and Block (2011) used a subset of shops and businesses with frequent cash transactions and less than 11 employees (like bars, restaurants and liquor stores etc.), and specific places where drug-related, prostitution-related and gambling-related incidents happened, as crime attractors for robberies in Chicago. In practice, however, or in large-scale research, it is difficult to distinguish crime generators from crime attractors. For example, without personal knowledge about the offenders, it is impossible to know whether they committed premediated thefts or visited the crime location without the intention to steal. For example, retail businesses are crime generators if many people come there without any particular criminal motivation, but some potential offenders visit there purposively because there are many easy crime opportunities. In the latter case, retail businesses can also be regarded as crime attractors (Steenbeek et al. 2012; Wilcox et al. 2004). Thus, it is too simple to argue that some types of locations (e.g. bars, or youth hangout places) are attractors and that others (e.g. shopping malls, or stadiums) are crime generators.

Besides crime generators and attractors, in particular with respect to street crimes targeting people, various studies have taken a more direct measurement of the population at risk as criminal opportunities for offenders (Andresen 2011, Hipp et al. 2021). By accounting for population mobility, prior scholarship has made significant progress in measuring ambient population. More recently, the advance of location-aware devices, especially the smartphone, has created opportunities for passively measuring human mobility at fine-grained spatial and temporal resolutions at low costs (Raento et al. 2009). In what follows, we summarize key measurements about ambient population.

## Ambient Population and Its Measurement

Given the dynamic nature of ambient population, using residential population as a measure of ambient population can hardly advance our understanding of crime, as crime is highly sensitive to its immediate spatial and temporal context. Two alternative strategies for measuring ambient population can be distinguished: survey data and location-based services data.

### Survey Data Measures

One important way to measure ambient populations is to use transportation surveys. Transportation surveys document the mobility of respondents, typically by asking them to keep

a trip diary in which they document the origins, destinations, travel modes and purposes of their movements. The resulting data can capture fluctuations of the ambient population (Boivin 2018). For example, Mburu and Helbich (2016) combine residential population census data and the number of daily commuters to calculate the number of residents in administrative units at any given time as a measure of ambient population. Felson and Boivin (2015) show with a transportation survey of 506 neighborhoods in Canada that the volumes of both violent and property crime in a neighborhood are strongly correlated with the number of trips leading to that neighborhood. Using the same data, Boivin and Felson (2017) further demonstrate that the number of non-crime trips between each pair of neighborhoods positively predicts the volume of crime trips between them.

### Location-Based Services (LBS) Data Measures and Their Time Effects

As compared to conventional approaches described earlier, a location-based services (LBS) approach can much more effectively capture the mobility of large population in detail. There is a growing body of research that utilizes location-based services data, such as mobile phone and social media data, to measure the presence of people throughout the city (Zhang, Zhou and Zhang 2017). One advantage of using such 'big data' is that it provides an unobtrusive measure to capture the spatial patterns of human mobility and social behavior rather than relying on retrospective survey responses (Kounadi et al. 2018; Song et al. 2018a). A series of studies lend support to the effectiveness of LBS-derived indicators for ambient population, finding a positive relationship between ambient population and crime (Hanaoka 2016; Kounadi et al. 2018; Malleon and Andresen 2016, Wang, Gerber and Brown 2012).

Two major types of LBS data, mobile phone and social media data, are increasingly being employed in criminology studies. For example, the mobility of the general population measured by mobile phone trajectories can help predict offenders' crime location choices (Song et al. 2019), and aggregated and anonymized human location data derived from mobile network activity can be used to predict crime levels (Bogomolov et al. 2014). Based on spatially referenced mobile phone data from Xi'an, a large city in China, with demographic characteristics of anonymized users, He et al. (2020) take into account variations in the ambient population by distinguishing between local and non-local phone users and find that besides crime attractors, generators and detractors, the proportion of non-local ambient population is significantly correlated with increased risk of larceny-theft. Another recent advance that emerges as a preferable alternative to static census population data stems from location-based social media platforms such as Twitter. In a study by Lan et al. (2019), counts of Twitter messages ('tweets') are used to analyze the spatial pattern of theft. They find that tweet counts, interpreted as a measure of ambient population, show a significant spillover effect on thefts, and that the total effect based on tweets counts outperforms that by Census population measure. Also using Twitter data from Southern California, Hipp et al. (2019) reveal that the temporal population estimated from social media help explain the level of crime in blocks during corresponding time periods.

However, recent studies also emphasize the limitations of these data. Mobile traces are recorded only when the user makes a call or sends a text message, and location of Twitter users are only recorded when they post a tweet. Less than 10 percent of tweets are geo-located (Anselin and Williams 2016), and multiple counting may bias the estimated distribution of the ambient population (Malleon and Andresen 2015). Both LBS measures may not reliably capture what people do and where they go between observed activities

(Cagney et al. 2020). In addition, big location data are produced by a sample that is not necessarily representative of the population of interest. The LBS traces might also vary between persons, between locations, and between different timeframes in phone and application use. Furthermore, their spatial resolution varies between urban and rural areas for mobile phone traces as a result of the uneven distribution of cell towers (De Montjoye et al. 2013). Consequently, these big location data may produce biased estimations of ambient population.

Data from *generic* location-based application platforms are less sensitive to a single type of phone usage and location traces and show increasing promise. They usually include all locations traces from all applications that use the platform location services. Baidu map is one such example. As the biggest online mapping service provider in China, like Google map in the West, it offers location services to a wide variety of individual users as well as applications. Recently, such data have been applied to the study of urban science after careful desensitization. For instance, Lv et al. (2021) use Baidu map data to uncover polycentric urban development and its determinants in China. Given their high quality and rich contextual information, the application of such embedded big data holds great promise for testing social science theories. This multi-source-based data is likely to catch a wider population and provide a more accurate picture of the population of interest. Nevertheless, to the best of our knowledge, this type of big data has not been used in criminology research.

In terms of time effects, considering the mobility of population, researchers have also used the big data to measure the temporal variation of ambient population. For example, using mobile phone location data, Hanaoka (2016) find that the impact of hourly population density on occurrences of snatch-and-run offenses is negative in the daytime while positive in the nighttime. Song et al. (2018a) find that the best indicators of risk populations for theft from the person, i.e. residential population, subway ridership, taxi ridership, and mobile phone users, varies by the course of a day during weekdays and weekends.

### Varied Effects in the Measurement of Ambient Populations for Crime Prediction

Within the myriad of data approaches in the measurement of ambient population, there is still little consensus on which approach is more appropriate. Some recent studies compare the effectiveness of different measures of ambient population in the prediction of crime. Malleson and Andresen (2016) use correlation coefficients to compare the effectiveness of mobile phone data, census population, and twitter, and identify the Census workday population as the most appropriate population-at-risk measure. However, this study does not control the potential impact of offenders and guardianship. Guided by the routine activity approach, Song et al. (2018a) further test indicators of risk populations (residential population, subway ridership, taxi ridership, and mobile phone users) to explain variations in theft from the person across space and time. Controlling for the potential confounding effects of offender and guardian presence, they show that on both weekdays and weekends, the best indicators of risk population vary over the course of a day.

Other researchers try to construct an improved indicator of the presence of potential victims by integrating multiple data sources. For example, Haleem et al. (2021) introduce the concept of an exposed population-at-risk, defined as the mix of residents and non-residents present in a spatial unit at a given time, and discern a temporally non-linear association between population size and violent crime in public space.

Most existing big data used to measure ambient populations only provide time and location, but do not distinguish between different categories of people in the ambient

population. They lack information on basic demographic attributes such as age and gender, and also on socio-economic attributes. Even if they exist in the original data of the data providers, they are typically removed to protect the privacy of the individuals (Cagney et al. 2020). With respect to cell phone data, an exception is a series of studies on ethnic segregation in Estonia, where the researchers assigned ethnic group membership (Estonian versus Russian) based on the user language settings of the cellphones (Silm and Ahas 2014; Toomet et al. 2015). The contents of Twitter messages may also provide context beyond time and location. Williams et al. (2017) have made an attempt to use a measure of broken windows found in the textual content of tweets communication to explain variance in offline crime patterns. Ristea et al. (2020) demonstrate that geo-tagged tweets, and tweets with violent content in particular, exhibit appreciable utility in predicting the volumes of seven common crime types around sporting events, both for game and nongame days. The strength and weakness of different population measures are summarized in Table 1.

When ambient population is used for indicating the risk of criminal victimization, an important distinction may be based on the functional roles of the people who make up the ambient population, i.e. what are the main activities that motivate their presence at the particular time and location. Prior studies have not established the functional roles of those who make up the ambient population: they do not distinguish between residents, employees or visitors, three groups that differ widely in their activity spaces and their roles in producing social cohesion, and may thereby have a different impact on local crime levels. These differential impacts are still to be assessed. The next section introduces the different roles that residents, employees and visitors play in affecting crimes.

### Residents, Employees and Visitors

Supported by the literature in environmental criminology, we have argued that the ambient population—the number of people who are actually present in a specific area during a specific time period—better reflects exposure to the risk of criminal victimization than the residential population — the number of people who happen to live in that area. The concept of ambient population takes temporal variation into account, and recognizes that people are mobile and do not necessarily spend their time in the immediate environment of their homes.

Nevertheless, the ambient population is not a homogeneous group. The individuals who are present at a given location at a given time, are there to pursue a variety of activities with different purposes: some live at the location, for others it is their workplace or school, and still others are there for leisure activities. Generally speaking, and based on the time of day they spend most of their time in a location, the ambient population can be grouped into *residents*, *employees*, and *visitors*. Different proportions of these three categories in the ambient population might entail different effects on the level of crime.

There have been a few studies that explored the relationship between crime and the presence of residents, employees, and visitors (strangers). Theoretically, opportunity theories of crime posit that in places with more strangers, anonymity and a lack of surveillance generate more crime (Brantingham and Brantingham 1995) (pp.14). Social disorganization theory contends that mobile populations will inhibit the development of local social cohesion and mutual trust, which consequently limits the informal social control against crime (Taylor 1997). According to both theories, the increased presence of transient visitors would lead to decreased guardianship and increased crime. The non-resident ambient population, i.e. visitors, arguably offer more anonymity for offenders, are more vulnerable



**Table 1** The strength and weakness of different population measures

Population measure	Strength	Weakness
Census population	A traditional measure with easy access	Cannot reflect the changes of population; long interval, collected every 10 years
Survey data measures	Rich in personal details, including respondents' social attributes and activity purposes	Sampling with small samples; cannot reflect the changes of population; long collected interval and high cost
Location-based services data measure	Mobile phone location data	User location recorded only when making a call or sending a text message
	Social media location data	Poor representative of general population; Less than 10 percent of tweets being geo-located
	Subway ridership	Reflecting people around subway stations
	Taxi ridership	Reflecting people present on the streets
	Data from generic location-based application platforms	Wide coverage including all locations traces from all applications that use the platform location services
		Has not been used in criminology research

for victimization and have less incentive to contribute to collective social control efforts (Tillyer et al. 2021). Empirically, combining transportation survey data and crime data, recent research by Boivin and Felson (2017) estimates daily population flows into each census tracts for four purposes (work, shopping, recreation, and education). Their findings reveal that crimes by visitors and residents both increase with inflows of visitors to the tract. In particular, recreational trips increase crimes in tracts more significantly than shopping trips.

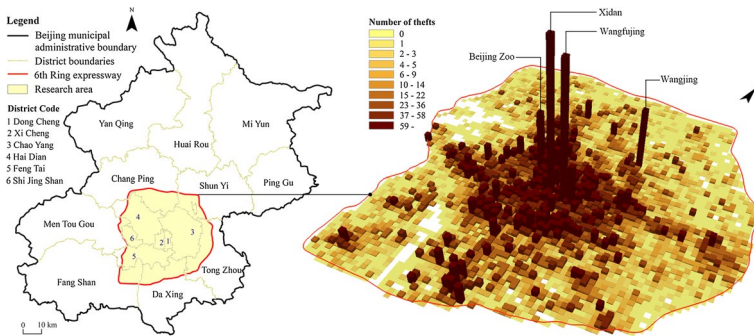
Conversely, informal social control and guardianship against crimes is almost entirely provided by those who frequent the local area the most, notably residents and employees. In the language of Jane Jacobs, the most significant and common form of informal surveillance comes from and is reinforced by, “an intricate, ... unconscious, network of voluntary control and standards among the people [who carry out activities therein]” (Jacobs 1961:31). Residents, shopkeepers, and local employees, for instance, are typically opposed to incivilities and crimes that may threaten their own interests. Lynch (1987) found that the occupational role affects the risk of victimization at work to a much greater degree than demographic characteristics of workers. It further deserves attention that employees on weekends can be a different type of employees on weekdays. In one study, Zeytinoglu and Cooke (2006) find that women, part-time, temporary or seasonal workers, those in the service sector, and those with lower education are more likely to regularly work on weekends in Canada. Consequently, the effects of employees on crimes on weekdays may differ from those on weekends.

In a nutshell, then, the composition of the ambient population observed in an area can vary over time and between areas, and changes in the composition rather than the volume of the ambient population may affect the volume of crime. Residents, employees and visitors vary in their familiarity with the local area, in their risk of falling victim to crime, and in their willingness to intervene against crime. Despite the theoretically presumed differential impact of residents, employees, and visitors on crime, the distinction between these subgroups in the ambient population has hardly been made in the criminological literature.

## Summary

From the literature, we know that more street activity brings together potential offenders and targets/victims, thus increasing criminal opportunities. Ambient population is a better indicator in predicting crimes than static census residential population. However, due to the general unavailability of attributes of the ambient population as measured by big data, prior studies mainly focus on quantifying the size, density, and spatial movement of the ambient population, but fail to distinguish functional roles: who work in the local area, who live there, and who merely pass by? Thus, there is room for improvement in our understanding of how the composition of the ambient population affects crime levels.

Firstly, the total population present in a spatial unit, is regarded as a homogeneous group, regardless of their purposes in the area. For example, residents and visitors with different purposes have different impacts on crimes (Boivin and Felson 2017). Depending on the nature of their activities in a particular spatial unit, individuals may have varying degrees of exposure to and leave differential influence on the spatial unit. While the literature generally supports the positive link between the size of ambient population and/or residential population and crime, it is unclear which group of the ambient population—residents, visitors, or employees—better represents the risk population of crime.



**Fig. 1** Research area and theft spatial pattern for the full week

Secondly, though the temporal variation of effects of general ambient population on crimes have been found, are the impact of different ambient population groups on crimes during weekdays different from that during weekends? People's activity patterns on weekdays differ from those on weekends and holidays. Accordingly, the compositions of the ambient population (i.e., visitors, employees, and residents) in a specific area may also differ. Previous studies reveal that the distribution of crime hotspots observed on weekdays is different from that on weekends (Andresen and Malleeson 2015). Little is known, however, about whether such variation is a result of changes in the composition of ambient population during two time periods.

To summarize, this study will contribute to ambient population literature by (1) decomposing the ambient population into residents, employees and visitors and evaluating their influence separately; (2) differentiating between weekdays and weekends.

## Data and Methods

### Study Area and Study Units

As the political and cultural center of China, Beijing has become one of the most populous and developed cities in China. In 2019, the administrative area of Beijing city was about 16,808 km<sup>2</sup> and the permanent population, who lived in Beijing continuously for more than 6 months, reached 21.54 million, of whom 7.46 million were the non-permanent or floating population without a Beijing *Hukou*. The *Hukou* regulation, or the household registration system, provides privileges regarding to education, housing and medical services to those who are registered with *Hukou* status. The system is an important means for the Chinese government to control population migration, by way of separating migrants from native residents (Long et al. 2021; Xiao et al. 2021).

This paper takes the central urban area of Beijing as the study area (see Fig. 1). The central urban area of Beijing refers to the area within the sixth ring expressway, with a total area of 2201 km<sup>2</sup>.

To capture spatial variation in crime, we overlaid the area with a 1 km × 1 km grid raster, creating initially 2198 grid cells in the study area. These grid cells are the spatial units of analysis in the present study. As some of the grid cells are mainly covered by water, forest or farmland, where there are very few people present and where virtually no thefts

can take place, these were eliminated. The final analysis includes 2104 square grid cells of 1 km × 1 km.

Recent work in the criminology of place emphasizes that crime opportunities are concentrated in micro-places rather than larger areas, such as neighborhoods (Weisburd et al. 2009). Although square 1 km<sup>2</sup> grid cells cover a larger surface and may be more heterogeneous than the units that have been distinguished in some studies of crime and place (street segments, or census blocks in the USA, or Lower Layer Super Output Areas in the United Kingdom), they are smaller and likely more homogeneous than traditional spatial units such as neighborhoods or urban districts. An additional advantage of grid cells is that in terms of size and shape they are identical, whereas more ‘natural’ units like neighborhoods can vary widely in size or shape. Square 1 km<sup>2</sup> grid cells were also used in prior spatial crime studies of cities in China (e.g., Song et al. 2018a).

## Data

The current study integrates multi-source data including criminal case data from the police recorded crimes dataset, population data from the Baidu map company whose data is aggregated to Baidu *Huiyan* Platform, population data of the latest available census (2010), Points of Interest data (POIs), global land cover map (FROM-GLC10), and road network data.

## Crime Data

This study focuses on thefts that took place outdoors in the study area. Police reported thefts data, from January 1 to December 31, 2014, were obtained from the Beijing Municipal Public Security Bureau. They include all cases that were reported to the police, regardless of the value of the stolen items. The data contain the geographic coordinates and the date and time stamps for each case. In 2014, there are total 49,280 cases, 33,829 on weekdays (i.e., Monday to Friday workdays excluding national holidays) and 15,451 on weekends (i.e., Saturday, Sunday, and holidays) in the study area.

Systematic empirical research on victims’ willingness to report to the police in China is limited (Wu, Sun and Hu 2021). However, in one of the few exceptions (Zhang et al. 2007), it was estimated that in contemporary urban China, 19 percent of theft victims reports to the police. This percentage is much lower than is common in many Western countries of the world. For example, the estimated victims’ reporting rate of personal property across 16 Western countries was estimated to be 42 percent (Goudriaan et al. 2004, Table 2), which is more than twice the percentage of 19 percent reported by Zhang et al. (2007).

We believe the lower reporting rate in China is unlikely to seriously bias our estimates, because amongst urban citizens in China, neither personal characteristics (age, gender, education, income, marital status, unemployment, prior victimization), nor neighborhood attributes (neighborhood disadvantage, social cohesion, informal control) seem affect the likelihood of reporting theft victimization to the police (Zhang et al. 2007; Zhuo et al. 2008). Like elsewhere (Goudriaan et al. 2004), in China the reporting of theft is mostly affected by the financial value of the stolen items. Nevertheless, although police reported thefts and other property crimes have been widely used in crime research in China (e.g., Peng et al. 2011; Song et al. 2018a), it is important to remain alert to potential issues of selectivity, including the notion that thefts of items with low value may be underrepresented in police data.

**Table 2** Descriptive statistics

	Mean	S.D.	Min	Max
<b>Dependent variables</b>				
Number of thefts during full week (case)	23.42	50.23	0.00	812.00
Number of thefts on weekdays (case)	16.08	33.78	0.00	547.00
Number of thefts at weekends (case)	7.34	17.00	0.00	348.00
<b>Independent variables</b>				
<b>Full week</b>				
Number of residents (1000, per day)	10.73	19.21	0.00	131.60
Number of employees (1000, per day)	5.79	9.84	0.00	116.42
Number of visitors (1000, per day)	16.50	20.73	0.31	247.14
Number of overall ambient population (1000, per day)	33.09	47.35	0.34	675.43
<b>Weekday</b>				
Number of residents (1000, per day)	10.96	19.61	0.00	137.26
Number of employees (1000, per day)	6.96	11.86	0.00	139.18
Number of visitors (1000, per day)	18.07	23.23	0.39	259.59
Number of overall ambient population (1000, per day)	36.06	51.79	0.39	706.11
<b>Weekend</b>				
Number of residents (1000, per day)	10.27	18.37	0.00	129.47
Number of employees (1000, per day)	5.09	8.59	0.00	103.20
Number of visitors (1000, per day)	15.55	19.40	0.22	239.59
Number of overall ambient population (1000, per day)	31.04	44.21	0.27	648.72
<b>Covariates</b>				
<b>Attractiveness</b>				
Proportion of built land area	0.88	0.58	0.03	1.00
Number of attractors	5.82	10.07	0.00	62.00
POI entropy	1.09	2.49	0.00	2.49
Proportion of female residents	0.49	0.12	0.37	0.68
<b>Accessibility</b>				
Proportion of branch roads (km)	0.51	0.35	0.00	1.00
Number of bus stops	6.32	18.52	0.00	43.00
Number of subway stations	0.08	0.34	0.00	5.00
<b>Guardianship</b>				
Distance to nearest police station (km)	1.88	1.93	0.02	7.11
Proportion of natives	0.66	0.72	0.00	0.96
Proportion of the highly educated	0.17	0.12	0.00	0.67

## Population Data

Our ambient population data come from the *Huiyan* platform belonging to the Baidu map company, a database collecting the data from all users of the Baidu Map location-based services (LBS). Baidu map is one of the most popular electronic navigation map suppliers in China, like Google Maps in the West. Over 1.65 million application developers as well as over 650 thousand apps or websites rely on the Baidu LBS, and it services more than 600 million Chinese users (Li et al. 2019). The mean daily location

requests received by the Baidu Map Location-based service surpasses 120 billion times. Furthermore, Baidu LBS data is regarded as a relatively high-quality data source for its integrated information collecting from three independent sources: GPS, WiFi, and cellular network (Chao et al. 2018). It needs to be pointed out that the Baidu users do not represent the complete population in an area, because not everyone uses a smartphone or the location-based services. However, according to Lv et al. (2021), the accuracy of the estimated population distributions based on the location logs of the Baidu Maps is above 90 percent. Baidu LBS data have been widely used in the study of population distributions, migration, urban land use, and other topics (Li et al. 2016; Liu et al. 2020b; Lyu and Zhang 2019).

Based on their trajectory characteristics, all users are classified into three categories: residents, employees, and visitors. For example, from the recorded locations of their mobile phone, if a person is at home, they will be counted as a resident in their home grid cell. But when they go to their place of work, at that location they will be counted as an employee. In grid cells they visit that do not contain their home or their work, they are counted as a visitor. The details of the processing algorithm and the criteria to judge whether a place is the person's home or work place are as follows. The home location and work location of the users are calculated using the Density-Based Spatial Clustering of Applications with Noise algorithm (DBSCAN) (Ester et al. 1996; Lv et al. 2021). It needs to be pointed out that this algorithm was developed and applied by the Baidu company and the result, the aggregated data, subsequently is described and used by us.

*Residents* in this study are defined by the following four criteria: (1) they send their location requests hundreds of times at a particular location over three months; (2) their location requests mainly happen in residential areas; (3) The time of their location requests is usually in the nighttime of weekdays or weekends; (4) they usually use non-public WiFi to send the location requests. *Employees* are defined based on three criteria: (1) the location requests mainly happen in the workplace such as offices or factory buildings; (2) the time of their location requests usually happen in daytime of weekdays; (3) they usually use public WiFi to send the location requests. *Visitors* are those who neither live nor work in that/a certain grid cell but request a location-based service there.

According to the algorithm, one person can be a resident or an employee of only one specific grid cell. If a person has two apartments in Beijing, in the data he will still appear to have only one residential grid cell, the one where he lives most of the time. The residential grid cell and working grid cell can be the same if a person lives and works in the same place. More specifically, if the distance between his residential place and working place is beyond 200 m, he is defined as both a resident and an employee in this grid cell, but if the distance is within 200 m, he is only defined as a resident. The probability for the latter situation is lower than 0.2%. As for the definition of visitor, one person can be recognized as a visitor in multiple grid cells excluding the grid cells in which they are a resident or an employee.

After discerning the types of a person belonging to every grid cell, for each day the sum is taken of the average total numbers of visitors, residents and employees in the ambient population that had been present in a specific grid cell. In this study, we calculate the average daily number of each of these three categories of population from January 2018 to December 2019 in each grid cell. Because the period during which ambient population data is available (2018–2019) does not align with the period over which crime data were collected (2014), we assume that the daily mobility of the population in Beijing did not change a great deal between 2014 and 2018–2019.

## Covariates

As mentioned above, criminal opportunities easily arise when motivated offenders and potential targets converge in space and time in the absence of capable guardianship (Cohen and Felson 1979). These factors are covered by three categories of covariates that are used to explain the frequency of thefts: attractiveness, accessibility, and guardianship.

**Attractiveness** The presence of crime generators and crime attractors is often associated with increased crime (Bernasco and Block 2011; Kinney et al. 2008). A first indicator of theft opportunity is land use. It should be noted that even the central urban area of Beijing, within the Sixth Ring Road, includes farm and forest land, where socio-economic activities are quite limited. Areas with built land attract more activities by the general population as well as by offenders. Therefore, we include the proportion of built land per grid cell as a general indicator of theft opportunity. Built land is measured using the FROM-GLC10 global land cover map. FROM-GLC10 is a 10-m resolution map originated from satellite image and classified by training samples (Gong et al. 2019).

With regard to crime attractors, in the Chinese context, several studies suggest that areas around (inter)net bars, game centers, bars and Karaoke TVs (KTVs) attract offenders and have elevated crime rates (Li et al. 2014; Song et al. 2018b). They are places where mostly young people meet each other and hang out. Therefore, we use the total number of (inter)net bars, game centers, bars and KTVs as an indicator of crime attractors.

As mentioned in the literature review, in most research, including ours, we do not know what the motivation of the offenders is when they visit the locations where they end up committing crimes. Therefore, it is often hard to tell crime attractors from crime generators. Instead of using a specific POI as a crime attractor or generator, following a prior study (Sung et al. 2015), we use the entropy index (EI) of urban function to measure the degree of urban diversity. The higher the entropy index, the more mixed the urban functions, which is found to produce elevated crime levels, exceeding those that would be predicted from population size alone (MacDonald 2015). This effect is presumed to exist because mixed land use, especially in metropolitan cities, often has a higher accessibility and wider service area than the local community, so that clustered facilities attract not only local residents but also visitors who live elsewhere. The anonymity and a lack of surveillance in such busy locations often facilitate offenders' crime decision making and elevate local crime rates (Boivin 2018).

To measure the different types of urban functions, 1.34 million Points of interest (POIs) in Beijing with latitude and longitude coordinates were also provided by Baidu map company at the end of December 2018. After eliminating POIs which do not reflect the main urban functions such as place names, road names, natural landscapes or indoor facilities, with reference to the classification standards for urban land use in China (GB50137-2011), 0.59 million remaining POIs were divided into 21 categories: administration, cultural facilities, schools, sports facilities, hospitals, welfare institutions, heritages, retail stores, wholesale markets, restaurants, hotels, office buildings, entertainment facilities, gas stations, parks and squares, factories, warehouses, subway stations, bus stations, parking lots, and residential communities. The entropy formula is

$$EI = \sum_{i=1}^n \frac{P_i \times \ln P_i}{\ln(n)} \quad (1)$$

where  $P_i$  is the ratio of the number of  $i$ -type POIs to the total number of POIs in each block and  $n$  is the number of units.

Females are usually considered as the more vulnerable group than males (Augustine et al. 2002). Therefore, based on the census data, female percentage are also regarded as an important aspect of attractiveness for thieves.

**Accessibility** Public transportation is often considered a rational choice by offenders to lower mobility costs (Oliveira, Natarajan and da Silva 2019) and ease escape after committing crimes. As such, communities with high accessibility attract high volumes of outsiders, involving potential offenders (Liu et al. 2020a). Empirically, the number of public transportation stations (e.g., bus stops and subway stations) is used to measure public transportation convenience. Studies confirm that crime rates increase with the presence of bus stops and subway stations (Bernasco and Block 2011; Haberman and Ratcliffe 2015; Liu et al. 2020a). Additionally, the presence of through traffic as well as high walkability may increase crime by increasing street traffic and the presence of strangers (Lee and Contreras 2021). Davies and Johnson (2015) found that road structure, measured by betweenness, influences burglary risk, with burglary risk higher on street segments with higher usage potential. Other studies use the total length of roads to capture usage potential. For example, Stucky and Ottensmann (2009) show that the total length of major streets in a grid cell is significantly positively associated with violent crime in Indianapolis. For accessibility, therefore, we control three variables: the number of bus stops, the number of subway stations, and the total length of streets in the grid cell. Road network data of Beijing was crawled from Baidu Map API in December 2018. It contains urban roads of all levels including main roads, secondary roads and branch roads but not the roads within the gated communities. In general, more public transportation facilities and higher density road networks represent better accessibility.

**Guardianship** According to routine activity theory, the absence of capable guardianship is crucial for crime to take place (Reynald 2009). The presence of strong guardianship can protect attractive victims from falling prey to motivated offenders. In empirical studies, distance to the closest police station as a proxy for accessibility to formal social control is commonly used. The presumed mechanism is that offenders are aware of an increased presence of police on streets near a police station, and therefore avoid offending in the proximity of police stations. Helbich and Jokar Arsanjani (2015) and Song et al. (2018a) found that the distance to police stations significantly impacts nonviolent crime. We, therefore, calculate the shortest road network distance from the grid cell center to the nearest police station as a measure of the level of guardianship. Moreover, we consider the percentage of natives as well as the percentage of highly-educated population as indicators of informal social control (Song et al. 2018b).

Variables of percentages of female, natives and highly-educated population are calculated based on census data. It needs to be pointed out that the boundaries of census units (i.e., subdistrict administration) do not align perfectly with 1 km × 1 km grid boundaries. To estimate the census population per grid, we first intersected grids with census units and calculated for each intersection its proportion in the census unit areas. After that, assuming that the population is homogeneously distributed within census units, the population of the intersection was estimated by multiplying the population of the corresponding census unit and the proportion, and finally these estimates were summed by every grid.



## Analytic Strategy

To describe the spatial patterns of theft and the three categories of ambient population, we visualize their spatial distribution with maps, and calculate their Moran's I indices. Moran's I is a measure of spatial autocorrelation that indicates whether characteristics of nearby areas are correlated. Like Pearson's correlation coefficient, it ranges between -1 (perfect negative) and 1 (perfect positive) and in large sample has an expected value of 0 in the absence of any correlation. The overall level of spatial autocorrelation in theft provides a measure of the extent to which theft frequencies are spatially clustered. In addition, we present Pearson correlation coefficients between all independent variables.

The second step in the analysis is to model the relationship between annual crime frequencies per grid cell as the dependent variable, and three different ambient population measures as the key independent variables, in addition to control variables. Like in other recent crime studies (e.g., Kim and Hipp 2020; Kurland and Johnson 2021), the distribution of theft in our data is over-dispersed: it has a larger variance than mean. To account for this overdispersion, as suggested by (Osgood 2000) and in line with most recent studies on sparse crime frequencies (e.g., Bernasco and Block 2011; Kim and Hipp 2020; Kurland and Johnson 2021), we use negative binomial regression models. Controlling for the effects of attractiveness, accessibility, and guardianship, four alternative indicators of ambient population—number of residents, number of employees, number of visitors, and the sum of these three categories—are included in separate models. In addition, and to determine the relative roles of ambient resident, employee and visitor populations, a model is estimated that contains these three ambient population categories simultaneously. Further, each of these models is estimated for weekdays only, for weekends only, and for complete weeks. Moreover, to ensure that results aren't sensitive to the assumptions of the Negative binomial model, with reference on Wooldridge (2010) and Berk and MacDonald (2008), we also present the results of corresponding Poisson regression models with robust standard errors in the appendix.

We estimated a set of negative binomial regression models. The general form of the negative binomial model is:

$$E(y) = \exp\left(\beta_0 + \beta_1 Pop + \sum_i^n \beta_i Cov + \varepsilon\right) \quad (2)$$

where  $\beta_0$  is an intercept and  $\beta_i$  are the coefficients of variables, where  $i$  is greater than 1. *Pop* represents the ambient population measure (residents, employees, visitors and overall ambient population respectively). *Cov* is the set of all other covariates representing the opportunity structure.  $\varepsilon$  represents unobserved individual heterogeneity.

We report coefficients, standardized coefficients and incidence rate ratios (IRRs) for every independent variable. Standardized coefficients are based on standardization ( $\mu=0$ ,  $sd=1$ ) of the independent variables and therefore can be used to compare the relative effects across variables in a model. Because we measure ambient population per 1000 people, the IRR indicates how many additional annual thefts are associated with a (daily) increase of 1000 people, i.e. localized mobile phone users. For example, if the estimated employee ambient population IRR equals 1.05, it means that an increase of 1000 employees is associated with an increase of thefts by a factor 1.05, i.e. by 5 percent.

To compare model fit between models, we use values of the Akaike Information Criterion (AIC). These measures do not require the comparison models to be hierarchically

nested. Smaller values of AIC indicate better fit. Moreover, a more rigorous comparison involves a bootstrapping method (Lubke and Campbell 2016; Wagenmakers et al. 2004), in which a bootstrap sample (a sample with replacement of 2104 cases from a dataset of size 2104) is taken 1000 times repeatedly from the original sample, and a percentage is generated of the most preferred model across the 1000 bootstrap replications. We compare the five models in a specific period. For example, for the full week, with the same covariates controlled, model 1 includes only residents, model 2 only employees, Model 3 only visitors, Model 4 includes residents, employees and visitors simultaneously, and Model 5 considering the overall ambient population. For each of the five models, we count in how many of the 1000 bootstrap replications the model is the preferred model (has lowest AIC) and report it in the results. These bootstrap results offer a more robust assessment of the relative performance of models than a simple comparison of the AIC values of the five models.

The spatial autocorrelation of dependent variables or the covariates can cause concerns, because spatial autocorrelation in the error terms violates standard statistical techniques that assume independence among observations. Ridgeway et al. (2019) and Bester et al. (2011) suggest that clustering the covariance matrix by larger geographical units (i.e. districts in our study) can allow for arbitrary dependence among the analysis units (grid cells) within these larger units. Moreover, as the number of grid cells within districts grows large, dependence among grid cells on the boundaries becomes negligible, so accounting for spatial dependence within districts becomes comparable to accounting for arbitrary spatial dependence. In line with this argument, we cluster standard errors at district level, 2104 grid cells in 12 districts, and compare the results of the negative binominal models with and without cluster robust standard errors. If the differences are not substantial, then spatial autocorrelation is not a concern. Between both models, the values of the key variables of ambient population are stable, and only the proportion of female residents turns from significant to non-significant. Therefore, spatial autocorrelation does not pose concern to the conclusions drawn from the analysis. In this study, we present only the model results of the negative binominal regressions with cluster robust standard errors.

Diagnostic tests for degrading multicollinearity were conducted by calculating variance inflation factors (VIFs). The largest variance inflation factor (VIF) value of all models is 6.18 (visitors), showing that there are no degrading multicollinearity issues in the models. For a discussion of rules of thumb for interpreting VIF values, see O'Brien (2007).

We used Stata version 15.0 software (StataCorp 2017) to calculate Pearson correlations and VIF values, and to estimate the negative binomial regression models. We used ArcGIS software to create maps and calculate Moran's I values.

## Empirical Findings

### Descriptive Statistics

There are 49,280 thefts with valid geographical coordinates in Beijing in 2014, with an average of 135.0 cases per day and 23.4 cases per grid cell per year (Table 2). During a full year, 33,829 thefts occurred on weekdays and 15,451 on weekends and holidays, which implies a weekday to weekend ratio of 2.19. The total number of weekdays and weekend days (which include national holidays) during the study period are 250 and 115 respectively, which a weekday to weekend ratio of 2.17. As both weekday to

weekend ratios of almost identical (2.19 versus 2.17), daily theft rates are roughly equal on weekdays and weekends.

The maximum values of the number of bus stops and the distance to nearest police station (km) are 43 and 7.11 respectively, which may seem quite large, but are quite plausible. Grid cells with train stations and in business areas often have many public transit hubs like bus stops and subway stations. The grid with longest distance to the nearest police station is in the suburb area in the South.

The distribution of theft exhibits a significant spatial concentration (Moran's  $I=0.265$ ,  $p<0.001$ ). Spatial concentration on weekdays (Moran's  $I=0.299$ ,  $p<0.001$ ) is stronger than on weekend days (Moran's  $I=0.194$ ,  $p<0.001$ ). As can be observed in Fig. 1, the overall density of theft in Beijing gradually decreases from the city center to the outskirts. A number of hotspots of theft emerge in areas where business centers, recreational activity spaces and wholesale markets are concentrated, such as the Wang-fujing commercial pedestrian street, the Xidan shopping center, the Beijing zoo clothing wholesale market, and the Wangjing business center. In the outer suburbs, thefts are relatively infrequent and dispersed, with small clusters in urban villages and in towns where district governments are located.

According to the Baidu big-data-based ambient residential population in 2018 and 2019, the mean daily number of residents in each grid cell in central urban area of Beijing is about 10.96 thousand on weekdays, 10.27 thousand on weekends, and 10.73 thousand on weekdays and weekends combined. The number of residents in Beijing decreases on weekends and holidays as people may go on vacation and leave the city.

The mean daily number of employees in each grid cell in central urban area of Beijing is 5.79 thousand. On average, there are 1.87 thousand more people working on weekdays than on weekends. To calculate visitors, we adopt a person-time measure, which means one person can be counted multiple times if they enter multiple grid cells in which they do not live or work. The mean daily number of visitors in each grid cell is as high as 16.50 thousand person-times during the whole study period; it drops from 18.07 thousand on weekdays to 15.55 thousand on weekends.

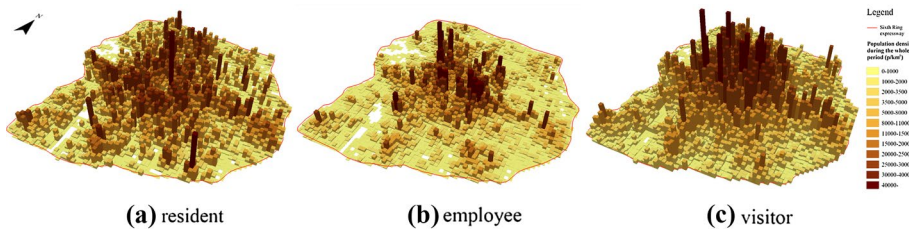
Table 3 shows that the correlations among the three ambient populations (residents, employees and visitors) are moderately high (0.67, 0.68 and 0.72), which indicates that grid cells high on one ambient population category tend to be high on other categories as well. The correlations are not almost perfect, which indicates that they do measure distinct constructs.

The spatial distribution characteristics of thefts mentioned above may be closely related to the spatial structure of various ambient populations. The distributions of residents, employees and visitors are all more concentrated in space than thefts. The strongest concentration is observed for visitors (Moran's  $I=0.641$ ,  $p<0.001$ ), followed by residents (Moran's  $I=0.544$ ,  $p<0.001$ ) and employees (Moran's  $I=0.519$ ,  $p<0.001$ ) (Fig. 2). Affected by the mobility of some residents, the concentration of residents on weekends (Moran's  $I=0.521$ ,  $p<0.001$ ) is slightly lower than that on weekdays (Moran's  $I=0.546$ ,  $p<0.001$ ). Compared with weekends (Moran's  $I=0.618$ ,  $p<0.001$ ), visitors exhibit an even higher clustering in space on weekdays (Moran's  $I=0.657$ ,  $p<0.001$ ). This could be because some socio-economic activities, such as socializing, sports, shopping, traveling, are partially transferred from city center to suburbs on weekends. The agglomeration of employees on weekends (Moran's  $I=0.520$ ,  $p<0.001$ ) is almost the same as that of weekdays (Moran's  $I=0.517$ ,  $p<0.001$ ). Despite that the size of employee population is smaller on weekends than on weekdays, the correlation between the two across grids is as high as 0.95. This indicates that the spatial

**Table 3** Pearson correlation coefficients between independent variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1) Residents	1												
(2) Employees	0.67***	1											
(3) Visitors	0.72***	0.68***	1										
(4) Overall ambient population	0.95***	0.91***	0.97***	1									
(5) Proportion of built land area	0.06**	0.06**	0.08***	0.07*	1								
(6) Number of attractors	0.55***	0.53***	0.55***	0.58***	0.07*	1							
(7) POI entropy	0.53***	0.58***	0.62***	0.68***	0.05*	0.47***	1						
(8) Proportion of female residents	0.20***	0.17***	0.23***	0.22***	0.26***	-0.01	0.20***	1					
(9) Proportion of branch roads	0.29***	0.32***	0.30***	0.32***	0.02	0.28***	0.26***	0.14***	1				
(10) Number of bus stops	0.53***	0.41***	0.42***	0.48***	0.49***	0.01	0.27***	0.11***	0.11***	1			
(11) Number of subway stations	0.40***	0.36***	0.45***	0.43***	0.26***	-0.01	0.27***	0.02	0.10***	0.19***	1		
(12) Distance to nearest police station	-0.14***	-0.21***	-0.27***	-0.22***	-0.12***	-0.02	-0.19***	-0.06**	-0.19***	0.15***	-0.09***	1	
(13) Proportion of natives	-0.12***	-0.11***	-0.14***	-0.14***	-0.13*	-0.04*	-0.15***	-0.11***	-0.13***	-0.03	-0.07***	0.10***	1
(14) Proportion of the highly educated	0.30***	0.38***	0.44***	0.39***	0.32***	0.00	0.29***	0.41***	0.29***	0.00	0.13***	-0.45***	-0.12***

Two-tail tests: \*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$



**Fig. 2** The spatial pattern of different ambient population for the full week

distribution of employees is quite stable on weekdays and weekends at the grid cell level.

## Regression Models

### Models Results of the Full Week

Table 4 shows the model effects of different measures of populations on thefts during the whole study period. From the results of Model 1 to Model 5, with the covariates controlled, all population measures, i.e. ambient population of residents, employees, visitors as well as the whole ambient population, significantly increase thefts. Between Models 1, 2 and 3, the AIC value of Model 3 is the smallest, and it has the highest percentage of being the most preferred model in the bootstrap procedure, which indicates that the visitor ambient population has a stronger relation to thefts than the resident and employee ambient populations. As for the two other categories of ambient population, employees appear to have less impact on thefts than residents do. In terms of the model performance based on AIC values, as well as the percentages in the bootstrap procedure, the overall ambient population models (Model 5) has the strongest relation with the volume of thefts. This implies that our distinction between residents, employees and visitors in the ambient population supports their differential effect on theft rates, but that the distinction itself does not seem to add to the prediction. Compared with the results of Poisson regression model with robust standard errors in the appendix, the values of coefficients do differ from the negative binomial regression models but their directions and significance levels do not change.

In model 4, residents, employees and visitors are considered simultaneously. All of them are significant. A one thousand persons increase in residents, employees and visitors, implies that the number of thefts increases by 1.1 percent, 1.7 percent and 3.8 percent respectively. The standardized coefficients show that visitors have the strongest impact on thefts, then followed by residents. As for the covariates, the area of built land, number of attractors, POI entropy, and the proportion of branch roads are positively and significantly related to thefts regardless of the population measures, while the distance to nearest police station shows a significant negative effect on theft, indicating that more thefts are committed in grid cells closer to police stations, a finding consistent with the findings of previous research (Helbich and Jokar Arsanjani 2015; Song et al. 2018a). The proportion of female residents has no significant effects on thefts. The area of built land has the largest effect according to its standardized coefficient among all the covariates. Regarding accessibility, the number of bus stops and subway stations are not significant, whose effects may be replaced by the ambient population, leading to their insignificance to thefts in most models.

**Table 4** Model results of negative binominal regression with cluster robust standard errors during the full week. N = 2104 square grids

	Model 1			Model 2			Model 3			Model 4			Model 5		
	Coef.	IRR	St.Coef.	Coef.	IRR	St.Coef.	Coef.	IRR	St.Coef.	Coef.	IRR	St.Coef.	Coef.	IRR	St.Coef.
Ambient population															
Residents (1000)	0.029***	1.030	0.559				0.010*	1.011	0.203						
Employees (1000)				0.050***	1.051	0.489				0.017*	1.017	0.169			
Visitors (1000)							0.039***	1.040	0.816	0.038***	1.039	0.788			
Overall ambient population (1000)													0.020***	1.020	1.036
Covariates															
Proportion of built land area	0.093**	1.098	5.466	0.092**	1.096	5.402	0.088**	1.092	5.160	0.088**	1.092	5.166	0.089**	1.093	5.207
Number of attractors	0.050***	1.051	0.504	0.050***	1.050	0.501	0.038***	1.039	0.383	0.038***	1.039	0.381	0.037***	1.038	0.375
POI entropy	0.013	1.013	0.033	0.002	1.002	0.005	0.049*	1.050	0.123	0.050*	1.051	0.125	0.042*	1.042	0.105
Proportion of female residents	1.675	5.337	0.162	2.351	10.498	0.228	2.215	9.162	0.214	2.107	8.225	0.204	2.161	8.679	0.209
Proportion of branch roads	1.352***	3.864	0.478	1.315***	3.724	0.465	1.382***	3.982	0.489	1.376***	3.957	0.487	1.283***	3.608	0.454
Number of bus stops	0.003	1.003	0.050	0.006	1.006	0.111	0.003	1.003	0.053	0.002	1.002	0.045	0.002	1.002	0.031
Number of subway stations	0.141	1.152	0.048	0.098	1.103	0.033	0.177	1.189	0.060	-0.171	0.842	-0.058	-0.079	0.924	-0.027

**Table 4** (continued)

	Model 1			Model 2			Model 3			Model 4			Model 5		
	Coef.	IRR	St.Coef.	Coef.	IRR	St.Coef.	Coef.	IRR	St.Coef.	Coef.	IRR	St.Coef.	Coef.	IRR	St.Coef.
Distance to nearest police station	-0.149***	0.861	-0.287	-0.153***	0.858	-0.296	-0.114*	0.892	-0.220	-0.113*	0.893	-0.219	-0.126**	0.882	-0.242
Proportion of natives	-0.206***	0.814	-0.147	-0.239***	0.788	-0.171	-0.208***	0.812	-0.149	-0.206***	0.814	-0.147	-0.213***	0.808	-0.152
Proportion of the highly educated	-0.569	0.511	-0.069	-0.056	0.945	-0.007	-0.393	0.675	-0.048	-0.318	0.727	-0.039	-0.189	0.828	-0.023
<i>Constant</i>	0.062	1.064	1.795	-0.041	0.960	1.807	-0.274	0.760	1.788	-0.241	0.786	1.788	-0.135	0.874	1.795
AIC	13,974.96			13,988.91			13,898.09			13,889.06			13,861.85		
Bootstrap (%)	0.0%			0.0%			25.2%			30.7%			44.1%		

Two-tail tests: \*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$ . St.Coef. is the coefficient value if all independent variables are standardized. Bootstrap (%) is the percentage of times being the most preferred model, which is with the lowest AIC value, across the 1000 bootstrap replications

## Model Results of Weekdays and Weekends

Separate models are estimated for weekdays (Table 5) and weekends (Table 6). The population distribution is still an important factor in predicting thefts among all the variables throughout all the models. Both on weekdays and weekends, the models of overall ambient population perform better than the others. Looking into different ambient population, the visitor ambient population is the best indicator, while visitors account for less than a half of the overall ambient population.

Nevertheless, the performance of models is some different between weekdays and weekends. Resident ambient population outperforms employee ambient population on weekdays than on weekends. In Model 9, the employee variable is not significant. While during weekends, employees have stronger impact on theft than residents. In Model 14, the residents variable is not significant.

Covariates in weekdays and weekends have similar performance as in models of the full week. Taking the models considering residents, employees, and visitors simultaneously as an example (Model 4, 9 and 14), built land, crime attractors, POI entropy, as well as branch road proportion are significantly related with the increase of thefts, while distance to nearest police station and proportion of natives have negative impact on thefts. There is one exception that in Model 14, proportion of the highly educated will decrease thefts significantly while not significant in Model 4 and 9.

## Conclusions and Discussion

Using big location data of an online map service company (Baidu map), this study decomposes the ambient population into residents, employees and visitors, and explores their impacts on thefts in 1 km × 1 km grids in Beijing, China. With the same set of covariates being controlled, it first evaluates the influence of residents, employees, and visitors on thefts for the full week, and subsequently compares the performance of these different ambient population measures on weekdays and weekends.

The key conclusion of the study is that our findings show that the ambient population is not a homogeneous group, and that the composition of the ambient population affects its influence on thefts. Generally speaking, for the full week, visitors have the largest effect on thefts, followed by residents and employees. Because visitors are outsiders, they may be more likely to fall victim to crimes, given their unfamiliarity with the local area. Moreover, according to social disorganization theory, high population mobility is detrimental for local social cohesion and mutual trust. Large volumes of visitors bring anonymity to the local area (Boivin and Felson 2017), reduce informal social control, and create more criminal opportunities. Possession of valuable belongings, unfamiliarity, and high anonymity all contribute to visitors' elevated risk of theft victimization. In contrast, although residents and employees can also be potential targets, in light of Jacobs' notion of "eyes on the street" (Jacobs, 1961), they may be better surveillants and more likely to intervene for the public good than others (Tillyer et al. 2021). These findings may inform policy makers who may be advised to focus their crime prevention strategies on places that attract large numbers of visitors, rather than places that attract large ambient populations more generally.



**Table 5** Model results of negative binominal regression with cluster robust standard errors on weekdays. N = 2104 square grids

	Model 6			Model 7			Model 8			Model 9			Model 10		
	Coef.	IRR	St.Coef.	Coef.	IRR	St.Coef.	Coef.	IRR	St.Coef.	Coef.	IRR	St.Coef.	Coef.	IRR	St.Coef.
Ambient population															
Residents (1000)	0.029***	1.030	0.577							0.010*	1.010	0.202			
Employees (1000)				0.042***	1.042	0.493				0.009	1.009	0.115			
Visitors (1000)							0.033***	1.034	0.776	0.032***	1.033	0.664			
Overall ambient population (1000)													0.019***	1.019	0.940
Covariates															
Proportion of built land area	0.101**	1.106	5.926	0.100**	1.105	5.867	0.096**	1.100	5.608	0.095**	1.100	5.603	0.096**	1.101	5.635
Number of attractors	0.047***	1.049	0.477	0.048***	1.049	0.478	0.036**	1.037	0.365	0.035***	1.036	0.354	0.035***	1.035	0.348
POI entropy	0.013	1.013	0.033	0.002	1.002	0.005	0.038*	1.038	0.095	0.043*	1.044	0.109	0.037*	1.037	0.091
Proportion of female residents	1.755	5.785	0.170	2.439	11.464	0.236	2.560	12.931	0.248	2.346	10.447	0.227	2.424	11.288	0.235

**Table 5** (continued)

	Model 6			Model 7			Model 8			Model 9			Model 10		
	Coef.	IRR	St.Coef.	Coef.	IRR	St.Coef.	Coef.	IRR	St.Coef.	Coef.	IRR	St.Coef.	Coef.	IRR	St.Coef.
Proportion of branch roads	1.349***	3.855	0.477	1.321***	3.746	0.467	1.387***	4.003	0.491	1.375***	3.954	0.486	1.288***	3.626	0.456
Number of bus stops	0.003	1.003	0.053	0.006	1.006	0.117	0.005	1.005	0.090	0.004	1.004	0.067	0.003	1.003	0.049
Number of subway stations	0.139	1.149	0.047	0.089	1.093	0.030	-0.176	0.839	-0.059	-0.173	0.841	-0.058	-0.091	0.913	-0.031
Distance to nearest police station	-0.159***	0.853	-0.307	-0.163***	0.849	-0.314	-0.126**	0.882	-0.243	-0.123**	0.884	-0.238	-0.135**	0.874	-0.260
Proportion of natives	-0.203***	0.817	-0.145	-0.239***	0.787	-0.171	-0.224***	0.799	-0.160	-0.215***	0.807	-0.154	-0.220***	0.803	-0.157
Proportion of the highly educated	-0.450	0.589	-0.055	-0.183	0.833	-0.022	-0.741	0.477	-0.090	-0.573	0.564	-0.070	-0.468	0.626	-0.057
Constant	-0.362	0.696	1.367	-0.465	0.628	1.380	-0.718	0.488	1.368	-0.670	0.512	1.366	-0.587	0.556	1.372
AIC	12,629.30			12,646.01			12,563.88			12,557.19			12,551.63		

**Table 5** (continued)

	Model 6			Model 7			Model 8			Model 9			Model 10		
	Coef.	IRR	St.Coeff.	Coef.	IRR	St.Coeff.	Coef.	IRR	St.Coeff.	Coef.	IRR	St.Coeff.	Coef.	IRR	St.Coeff.
Bootstrap (%)	0.2%			0.0%			26.7%			32.9%			40.2%		

Two-tail tests: \*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$ . St.Coeff. is the coefficient value if all independent variables are standardized. Bootstrap (%) is the percentage of times being the most preferred model, which is with the lowest AIC value, across the 1000 bootstrap replications

**Table 6** Model results of negative binominal regression with cluster robust standard errors at weekends, N = 2104 square grids

	Model 11			Model 12			Model 13			Model 14			Model 15		
	Coef.	IRR	St.Coef.	Coef.	IRR	St.Coef.	Coef.	IRR	St.Coef.	Coef.	IRR	St.Coef.	Coef.	IRR	St.Coef.
Ambient population															
Residents (1000)	0.026***	1.027	0.486				0.005	1.005	0.089						
Employees (1000)				0.065***	1.067	0.557				0.021*	1.021	0.182			
Visitors (1000)							0.039***	1.040	0.756	0.038***	1.038	0.730			
Overall ambient population (1000)													0.019***	1.019	0.832
Covariates															
Proportion of built land area	0.108**	1.114	6.338	0.105**	1.111	6.188	0.100**	1.105	5.872	0.100**	1.105	5.862	0.101**	1.107	5.957
Number of attractors	0.052***	1.053	0.521	0.050***	1.051	0.500	0.040***	1.040	0.399	0.039***	1.040	0.394	0.039***	1.040	0.396
POI entropy	0.007	1.007	0.018	0.003	1.003	0.007	0.044*	1.045	0.110	0.043*	1.044	0.108	0.036*	1.036	0.089
Proportion of female residents	1.374	3.953	0.133	2.010	7.464	0.195	1.790	5.990	0.173	1.788	5.978	0.173	1.765	5.840	0.171
Proportion of branch roads	1.216***	3.375	0.430	1.167***	3.212	0.413	1.134***	3.436	0.437	0.039***	3.380	0.431	1.146***	3.145	0.406
Number of bus stops	0.001	1.001	0.013	0.004	1.004	0.068	0.001	1.001	0.002	0.001	1.001	0.002	0.001	1.001	0.013
Number of subway stations	0.180	1.197	0.061	0.139	1.149	0.047	-0.104	0.884	-0.035	-0.105	0.900	-0.036	-0.019	0.981	-0.006

**Table 6** (continued)

	Model 11			Model 12			Model 13			Model 14			Model 15		
	Coef.	IRR	St.Coef.	Coef.	IRR	St.Coef.	Coef.	IRR	St.Coef.	Coef.	IRR	St.Coef.	Coef.	IRR	St.Coef.
Distance to nearest police station	-0.125*	0.882	-0.242	-0.129**	0.879	-0.248	-0.091*	0.913	-0.176	-0.091*	0.913	-0.176	-0.102*	0.903	-0.197
Proportion of natives	-0.180***	0.835	-0.129	-0.205***	0.815	-0.146	-0.168***	0.845	-0.120	-0.171***	0.843	-0.122	-0.180***	0.836	-0.128
Proportion of the highly educated	-0.402	0.630	-0.074	-0.007	0.993	-0.002	-0.124	0.883	-0.015	-0.175	0.839	-0.021	-0.021	0.979	-0.003
<i>Constant</i>	-1.002	0.367	0.591	-1.094	0.335	0.609	-1.302*	0.272	0.600	-1.279*	0.278	0.603	-1.161	0.313	0.607
AIC	10,054.25			10,050.72			9970.788			9961.409			9958.077		
Bootstrap (%)	0.0%			0.0%			28.6%			34.5%			46.9%		

Two-tail tests: \*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$ . St.Coef. is the coefficient value if all independent variables are standardized. Bootstrap (%) is the percentage of times being the most preferred model, which is with the lowest AIC value, across the 1000 bootstrap replications

Another important finding is that the effects of residents and employees vary between weekdays and weekends. Whereas visitors have the largest effect on thefts both on weekdays and weekends, on weekdays the effect of residents is larger than the effect of employees, and on weekends the effect of employees is larger than the effect of residents. A possible explanation is on weekdays residents tend to stay around their residence and have a more flexible schedule. They may carry out outdoor activities, such as grocery shopping and exercising, exposing themselves to motivated offenders to a greater extent. Moreover, some residents, especially the old and the children, a relative vulnerable group to victimization, travel shorter distances during their daily activities, and are more likely to stay close to where they live (Yuan and Raubal 2016). By contrast, employees are predominantly involved in indoor activities in the workspace (e.g., office buildings and factories), making them less likely to fall victim to theft (Song et al. 2018b). Therefore, compared to employees who have to work indoors, residents have a stronger effect on thefts due to their longer stay, higher vulnerability, and greater exposure to risk on weekdays.

But on weekends, for residents, more people perform daily activities near home, more young residents are active in the neighborhood, and more family members are around. This generates stronger guardianship and reduces vulnerability among residents. Although some employees work overtime on weekends, they have a more flexible time budget than on weekdays and might just spend part of the day on working (Sun et al. 2014). For the rest of the time, they can have more outdoor activities around, which increases their exposure to thefts. To summarize, as a result of their distinctive activity patterns—the activity space (i.e. outdoor or indoor) and the way they spend time in a grid cell—different groups of ambient population have varied effects on crime. Consequently, on weekends the presence of employees has a stronger effect on thefts than the presence of residents. The implication of our findings is that for an improved understanding of how ambient populations affect crime rates, it is important to take into account both the different roles of the people that make up the ambient population (e.g. resident, employee or visitor) and the different phases in the weekly time cycle (e.g. weekday and weekend). In the application of these findings, one should generally not favor one ambient population category or one temporal phase over the other, but rather measure and apply them simultaneously.

In line with the literature, we find that indicators of attractiveness—area of built land, number of attractors, and POI entropy—are positively related with crime, consistent with previous findings (Kinney et al. 2008). As for accessibility, the longer the street is, the more thefts there will be. However, unlike the findings of Liu et al. (2020a), we do not find any effect of the presence of bus stops or subway stations in the models considering ambient population. This suggests that these indicators of accessibility (how convenient it is for people to visit the location) do not add to the information already included in the indicators of ambient population (how many people are actually present in the location). Another possible reason is that in Beijing, the bus stops quite evenly distribute around the city, therefore having no significant impact on thefts. The proportion of female residents presents no impact on thefts, indicating that the gender of victims makes no difference to offenders.

When it comes to guardianship, like the findings of Helbich and Jokar Arsanjani (2015) and Song et al. (2018a), distance to nearest police station, potentially indicative of guardianship, has a negative relation with theft when ambient population is not accounted for. A possible explanation is that that proximity to police stations may increase the willingness of victims to report their victimization to the police, resulting in more rather than less thefts being reported closer to police stations. In addition, police stations may be intentionally located in crowded places and in high crime areas to improve accessibility and response to calls for service. Any inhibitive effect of formal guardianship on theft may thus be offset by these two factors. In terms of informal social control, the natives feature strong informal social control and can prevent thefts. Though the highly educated population is not significant in many models, it presents a negative effect, which also indicates their informal social control on thefts, especially on weekends.

As any study, there are limitations of this study that should be noted. First, in terms of the identification of ambient population subgroups, although we are able to differentiate between residents, employees, and visitors, we are unable to further discern the type of activities individuals are engaged in during the study period. The type of employees observed on weekdays and weekends may be different. As argued above, the employee population is likely involved in the public and semi-public domains on weekends than weekdays, which may contribute to the differential effect size of employees on theft on the two timeframes. Relatedly, the population subgroup labels are mutually exclusive, which may bias our estimation of individuals' trip purposes. In our dataset, a person carrying out non-working activities, such as shopping, in the grid cells where one works on the weekend, he or she is still regarded as an employee rather than a visitor. At the same time, different types of employees, such as employees by retail vs. nonretail, are found to have different effects on crimes in the literature. However, this distinction cannot be made in this study because the resident, employee and visitor categories are part of the data made accessible by the Baidu Company, which cannot be further subdivided by us. Another limitation is that the algorithms underlying the distinctions between residents, employees and visitors have been developed, implemented and described by the Baidu company, and are not openly available. Because these data are unique (that is why we are using them to further our understanding of how ambient population affects crime) we have no way of comparing the data with reliable external sources.

In addition, our data is unable to identify people who live and work within the same grid cell. For instance, some owners of groceries live and work in the same units, accessing a unique WiFi, and they may be classified as only residents but not employees. Further, employees without a fixed working place, such as taxi drivers, delivery men and floating vendors, will be treated as visitors. These issues may bias our interpretation of employees, residents and visitors, but only marginally because these groups of city residents probably only constitute a small fraction of the whole population.

Another limitation is that we used crime data from 2014 and population data from 2018–2019. Although the spatio-temporal pattern of population distribution most likely

has not changed a lot from 2014 to 2018, it would have been slightly more precise to analyze both data sources from the same years.

As for the study units, according to the literature of crime at places, 1 km \* 1 km grid may be too large, and potentially raise concerns of environment heterogeneity within the grids. This can be also regarded as a modifiable areal unit problem (Fotheringham and Wong 1991) and could inspire further work that distinguishes multiple spatial units of analysis, provided that more detailed data becomes available.

Our measure of formal guardianship, the distance to the nearest police station, is not a perfect measure of guardianship, as it fails to measure the actual whereabouts and activities of the police, and because it might also be related to the willingness of victims to report to the police. In future work, improved measures of formal guardianship could remedy this limitation. Finally, this study has only discussed one type of crime, theft. Different crimes may have different kinds of crime targets. For example, the ambient population can be a target of thefts, but not burglary which targets on a specific dwelling instead of a person. The effects of visitors, residents, and visitors in the ambient population may be different with respect to other types of crime than thefts.

In the empirical literature, the negative binomial model has been popular in estimating crime count data, probably because it is straightforward, utilizes standard maximum likelihood estimation techniques, can be estimated with standard software packages, and accounts for overdispersion in the data. The negative binomial model itself does not account for spatial autocorrelation. If the differences between results of the negative binomial models with and without cluster robust standard errors are substantial, the technique is not appropriate and other techniques for analyzing spatially autocorrelated count data are needed, such as spatial filtering (Chun 2014; Haining et al. 2009), generalized cross entropy (Roman et al. 2008) and Bayesian methods (Liu and Zhu 2016). A comprehensive overview of methods is provided by (Dorman et al. 2007). In our application, the negative binomial model suffices because it accounts for overdispersion and because the differences mentioned above turn out to be very small.

Despite the aforementioned limitations, this study contributes to the literature by further investigating the differential effects of three main categories of ambient population on thefts on weekdays and weekends, and by using big data with a wider reach. Findings show that visitors and residents have stronger effects on thefts than employees, which presumably is a result of their more varied activity space, length of stay, and possibly of their informal social control capacity too.

## Appendix

See Table 7.



**Table 7** Model results of Poisson regression with cluster robust standard errors

	Full week														
	Weekdays					Weekends									
	Model 1 Coef.	Model 2 Coef.	Model 3 Coef.	Model 4 Coef.	Model 5 Coef.	Model 6 Coef.	Model 7 Coef.	Model 8 Coef.	Model 9 Coef.	Model 10 Coef.	Model 11 Coef.	Model 12 Coef.	Model 13 Coef.	Model 14 Coef.	Model 15 Coef.
<i>Ambient population</i>															
Residents (1000)	0.013***			0.008**		0.013***		0.007*			0.008***			0.003	
Employees (1000)		0.019***		0.007*			0.016***		0.005		0.021***			0.010***	
Visitors (1000)			0.017***	0.020***				0.015***	0.020***				0.017***	0.024***	
Overall ambient population (1000)					0.010***						0.011***				0.014***
<i>Covariates</i>															
Proportion of built land area	0.125***	0.121***	0.120***	0.115***	0.123***	0.126***	0.121***	0.120***	0.117***	0.123***	0.125***	0.121***	0.120***	0.114***	0.124***
Number of attractors	0.027***	0.018***	0.014***	0.017***	0.018***	0.026***	0.017***	0.012**	0.016***	0.015***	0.029***	0.019***	0.016***	0.039***	0.021***
POI entropy	0.031*	0.032**	0.014***	0.011	0.028*	0.033*	0.033* 0.019*	0.015*	0.029* 0.028*	0.029*	0.029*	0.017***	0.017***	0.019***	0.025*
Proportion of female residents	1.885**	2.624***	2.728***	2.872***	2.352***	1.867**	2.610***	2.943***	2.955***	2.472***	1.907**	2.639***	2.577***	2.825***	2.292***
Proportion of branch roads	1.058***	1.006***	1.062***	1.122***	1.022***	1.059***	1.008***	1.075***	1.136***	1.021***	1.055***	1.001***	1.049***	1.106***	1.020***
Number of bus stops	-0.003	-0.002	-0.003	-0.002	-0.003	-0.003	-0.002	-0.002	-0.001	-0.003	-0.003	-0.002	-0.004	-0.002	-0.004
Number of subway stations	0.043	0.008	0.284	0.226	0.162	0.045	0.006	0.260	0.243	0.170	0.040	0.012	0.283	0.209	0.154

**Table 7** (continued)

	Full week			Weekdays					Weekends						
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12	Model 13	Model 14	Model 15
	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.
Distance to nearest police station	-0.198***	-0.192***	-0.185***	-0.151***	-0.207***	-0.207***	-0.200***	-0.186***	-0.162***	-0.215***	-0.177***	-0.174***	-0.172***	-0.132**	-0.187***
Proportion of the highly educated natives	-0.109***	-0.119***	-0.121***	-0.133***	-0.113***	-0.107***	-0.117***	-0.127***	-0.135***	-0.113***	-0.113***	-0.122***	-0.120***	-0.133***	-0.115***
Proportion of the highly educated	-0.652	-0.393	-0.010	-0.106	-0.304	-0.643	-0.382	-0.178	-0.291	-0.281	-0.674	-0.419	-0.194	-0.061	-0.467
Constant	0.640	0.403	0.271	0.111	-0.135	0.286	0.045	-0.172	-0.257	0.101	-0.561	-0.791*	-0.859*	-1.079**	-0.670
AIC	60,008.31	61,694.53	59,671.61	58,466.47	58,066.57	42,297.28	43,687.78	42,031.05	41,636.02	41,077.05	24,520.77	23,818.56	23,297.13	22,693.51	22,297.04
Bootstrap (%)	0.0%	0.0%	16.7%	37.8%	45.5%	0.0%	0.0%	24.3%	34.2%	41.5%	0.0%	0.3%	18.8%	33.6%	47.3%

Two-tail tests: \*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$ . Bootstrap (%) is the percentage of times being the most preferred model, which is with the lowest AIC value, across the 1000 bootstrap replications

**Funding** Funding was provided by National Natural Science Foundation of China (Grant Nos. 41901177, 42171218), Youth Program of the Humanities and Social Science Research of the Ministry of Education (CN) (Grant No. 19YJCZH258), Natural Science Foundation of Guangdong Province (Grant No. 2019A1515011065), Youth Program of National Social Science Foundation of China (Grant No. 21CSH006).

## References

- Andresen MA (2006) Crime measures and the spatial analysis of criminal activity. *Br J Criminol* 46(2):258–285. <https://doi.org/10.1093/bjc/azi054>
- Andresen MA (2007) Location quotients, ambient populations, and the spatial analysis of crime in Vancouver, Canada. *Environ Plan A* 39(10):2423–2444. <https://doi.org/10.1068/a38187>
- Andresen MA (2011) The ambient population and crime analysis. *Prof Geogr* 63(2):193–212. <https://doi.org/10.1080/00330124.2010.547151>
- Andresen MA, Jenion GW (2010) Ambient populations and the calculation of crime rates and risk. *Secur J* 23(2):114–133
- Andresen MA, Malleon N (2015) Intra-week spatial-temporal patterns of crime. *Crime Sci* 4(1):1–11
- Anselin L, Williams S (2016) Digital neighborhoods. *J Urban Int Res Placemaking Urban Sustain* 9(4):305–328
- Augustine MC, Wilcox P, Ousey GC, Clayton RR (2002) Opportunity theory and adolescent school-based victimization. *Violence Vict* 17(2):233–253
- Berk R, MacDonald JM (2008) Overdispersion and poisson regression. *J Quant Criminol* 24(3):269–284
- Bernasco W, Block R (2011) Robberies in Chicago: a block-level analysis of the influence of crime generators, crime attractors and offender anchor points. *J Res Crime Delinq* 48(1):33–57. <https://doi.org/10.1177/0022427810384135>
- Bester CA, Conley TG, Hansen CB (2011) Inference with dependent data using cluster covariance estimators. *J Econom* 165(2):137–151
- Boggs SL (1965) Urban crime patterns. *Am Sociol Rev* 30(6):899–908
- Bogomolov A, Lepri B, Staiano J, Oliver N, Pianesi F, Pentland A (2014) Once upon a crime: towards crime prediction from demographics and mobile data. In: Paper presented at the Proceedings of the 16th International Conference on Multimodal Interaction, Istanbul, Turkey. <https://doi.org/10.1145/2663204.2663254>
- Boivin R (2018) Routine activity, population (S) and crime: spatial heterogeneity and conflicting propositions about the neighborhood crime-population link. *Appl Geogr* 95:79–87
- Boivin R, Felson M (2017) Crimes by visitors versus crimes by residents: the influence of visitor inflows. *J Quant Criminol* 34(2):465–480. <https://doi.org/10.1007/s10940-017-9341-1>
- Brantingham PJ, Brantingham PL (1995) Criminality of place: crime generators and crime attractors. *Eur J Crim Policy Res* 3(3):5–26
- Cagney KA, Cornwell EY, Goldman AW, Cai L (2020) Urban mobility and activity space. *Annu Rev Sociol* 46:623–648
- Candia J, González MC, Wang P, Schoenharl T, Madey G, Barabási A-L (2008) Uncovering individual and collective human dynamics from mobile phone records. *J Phys a: Math Theor* 41(22):224015
- Chao HQ, Cao Y, Zhang JP, Xia F, Zhou Y, Shan HM (2018) Population density-based hospital recommendation with mobile lbs big data. In: 2018 IEEE International Conference on Big Data and Smart Computing (Bigcomp). pp 37–44 <https://doi.org/10.1109/BigComp.2018.00015>
- Chun Y (2014) Analyzing space-time crime incidents using eigenvector spatial filtering: an application to vehicle burglary. *Geogr Anal* 46(2):165–184. <https://doi.org/10.1111/gean.12034>
- Cohen LE, Felson M (1979) Social change and crime rate trends: a routine activity approach. *Am Sociol Rev* 44(4):588–608
- Cornish DB, Clarke RV (1987) Understanding crime displacement: an application of rational choice theory. *Criminology* 25(4):933–948
- Davies T, Johnson SD (2015) Examining the relationship between road structure and burglary risk via quantitative network analysis. *J Quant Criminol* 31(3):481–507
- De Montjoye YA, Hidalgo CA, Verleysen M, Blondel VD (2013) Unique in the crowd: the privacy bounds of human mobility. *Sci Rep* 3:1376
- Dorman CF, McPherson Jana M, Araújo Miguel B, Roger B, Janine B, Gudrun C, Davies Richard G, Alexandre H, Walter J, Daniel Kissling W, Ingolf K, Ralf O, Peres-Neto Pedro R, Björn R, Boris S, Schurr

- Frank M, Robert W (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30(5):609–628
- Ester M, Kriegel H-P, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. *Kdd* 96(34):226–231
- Felson M, Boivin R (2015) Daily crime flows within a city. *Crime Sci* 4(1):31. <https://doi.org/10.1186/s40163-015-0039-0>
- Fotheringham AS, Wong DWS (1991) The modifiable areal unit problem in multivariate statistical analysis. *Environ Plan A* 23(7):1025–1044
- Garofalo J, Siegel L, Laub J (1987) School-related victimizations among adolescents: an analysis of national crime survey (Ncs) narratives. *J Quant Criminol* 3(4):321–338
- Gong P, Liu H, Zhang MN, Li CC, Wang J, Huang HB, Clinton N, Ji LY, Li WY, Bai YQ, Chen B, Xu B, Zhu ZL, Yuan C, Suen HP, Guo J, Xu N, Li WJ, Zhao YY, Yang J, Yu CQ, Wang X, Fu HH, Yu L, Dronova I, Hui FM, Cheng X, Shi XL, Xiao FJ, Liu QF, Song LC (2019) Stable classification with limited sample: transferring a 30-M resolution sample set collected in 2015 to mapping 10-M resolution global land cover in 2017. *Sci Bull* 64(6):370–373. <https://doi.org/10.1016/j.scib.2019.03.002>
- Goudriaan H, Lynch JP, Nieuwebeerta P (2004) Reporting to the police in western nations: a theoretical analysis of the effects of social context. *Justice Q* 21(4):933–969. <https://doi.org/10.1080/0741882040096041>
- Haberman CP, Ratcliffe JH (2015) Testing for temporally differentiated relationships among potentially criminogenic places and census block street robbery counts. *Criminology* 53(3):457–483
- Haining R, Law J, Griffith D (2009) Modelling small area counts in the presence of overdispersion and spatial autocorrelation. *Comput Stat Data Anal* 53(8):2923–2937
- Haleem MS, Do Lee W, Ellison M, Bannister J (2021) The ‘exposed’ population, violent crime in public space and the night-time economy in manchester, Uk. *Eur J Crim Policy Res* 27:335–352
- Hanaoka K (2016) New insights on relationships between street crimes and ambient population: use of hourly population data estimated from mobile phone users’ locations. *Environ Plan B Plan Des*. <https://doi.org/10.1177/0265813516672454>
- He L, Páez A, Jiao J, An P, Lu C, Mao W, Long D (2020) Ambient population and larceny-theft: a spatial analysis using mobile phone data. *ISPRS Int J Geo-Inf* 9(6):342
- Helbich M, Arsanjani JJ (2015) Spatial eigenvector filtering for spatiotemporal crime mapping and spatial crime analysis. *Cartogr Geogr Inf Sci* 42(2):134–148
- Hipp JR, Bates C, Lichman M, Smyth P (2019) Using social media to measure temporal ambient population: Does it help explain local crime rates? *Justice Q* 36(4):718–748
- Hipp JR, Kim Y-A, Wo JC (2021) Micro-scale, meso-scale, macro-scale, and temporal scale: comparing the relative importance for robbery risk in New York City. *Justice Q* 38(5):767–791
- Jacobs J (1961) The death and life of great American cities. Random House, New York
- Kim Y-A (2018) Examining the relationship between the structural characteristics of place and crime by imputing census block data in street segments: is the pain worth the gain? *J Quant Criminol* 34(1):67–110
- Kim YA, Hipp JR (2020) Pathways: examining street network configurations, structural characteristics and spatial crime patterns in street segments. *J Quant Criminol* 36(4):725–752
- Kinney JB, Brantingham PL, Wuschke K, Kirk MG, Brantingham PJ (2008) Crime attractors, generators and detractors: land use and urban crime opportunities. *Built Environ* 34(1):62–74
- Kounadi O, Ristea A, Leitner M, Langford C (2018) Population at risk: using areal interpolation and twitter messages to create population models for burglaries and robberies. *Cartogr Geogr Inf Sci* 45(3):205–220
- Kurland J, Johnson SD (2021) The influence of stadia and the built environment on the spatial distribution of crime. *J Quant Criminol* 37(2):573–604
- Lan M, Liu L, Hernandez A, Liu W, Zhou H, Wang Z (2019) The spillover effect of geotagged tweets as a measure of ambient population for theft crime. *Sustainability* 11(23):6748
- Lee N, Contreras C (2021) Neighborhood walkability and crime: Does the relationship vary by crime type? *Environ Behav* 53(7):753–786
- Li Y, Zhang X, Lu F, Zhang Q, Wang Y (2014) Internet addiction among elementary and middle school students in China: a nationally representative sample study. *Cyberpsychol Behav Soc Netw* 17(2):111
- Li J, Ye Q, Deng X, Liu Y, Liu Y (2016) Spatial-temporal analysis on spring festival travel rush in China based on multisource big data. *Sustainability* 8(11):1184
- Li JG, Li JW, Yuan YZ, Li GF (2019) Spatiotemporal distribution characteristics and mechanism analysis of urban population density: a case of Xi’an, Shaanxi, China. *Cities* 86:62–70. <https://doi.org/10.1016/j.cities.2018.12.008>

- Liu H, Zhu X (2016) Exploring the influence of neighborhood characteristics on burglary risks: a bayesian random effects modeling approach. *ISPRS Int J Geo Inf* 5(7):102
- Liu L, Lan M, Eck JE, Kang EL (2020a) Assessing the effects of bus stop relocation on street robbery. *Comput Environ Urban Syst* 80:101455
- Liu W, Hou Q, Xie Z, Mai X (2020b) Urban network and regions in China: an analysis of daily migration with complex networks model. *Sustainability* 12(8):3208
- Long D, Liu L, Xu M, Feng J, Chen J, He L (2021) Ambient population and surveillance cameras: the guardianship role in street robbers' crime location choice. *Cities* 115:103223. <https://doi.org/10.1016/j.cities.2021.103223>
- Lubke GH, Campbell I (2016) Inference based on the best-fitting model can contribute to the replication crisis: assessing model selection uncertainty using a bootstrap approach. *Struct Equ Modeling Multi-discip J* 23(4):479–490
- Lv Y, Lan Z, Kan C, Zheng X (2021) Polycentric urban development and its determinants in China: a geo-spatial big data perspective. *Geogr Anal* 53(3):520–542
- Lynch JP (1987) Routine activity and victimization at work. *J Quant Criminol* 3(4):283–300
- Lyu F, Zhang L (2019) Using multi-source big data to understand the factors affecting urban park use in Wuhan. *Urban Forestry & Urban Greening* 43:126367
- MacDonald J (2015) Community design and crime: the impact of housing and the built environment. *Crime Justice* 44(1):333–383
- Malleson N, Andresen MA (2015) Spatio-temporal crime hotspots and the ambient population. *Crime Sc* 4(1):1–8
- Malleson N, Andresen MA (2016) Exploring the impact of ambient population measures on london crime hotspots. *J Crim Just* 46:52–63. <https://doi.org/10.1016/j.jcrimjus.2016.03.002>
- Mburu LW, Helbich M (2016) Crime risk estimation with a commuter-harmonized ambient population. *Ann Am Assoc Geogr* 106(4):804–818. <https://doi.org/10.1080/24694452.2016.1163252>
- O'Brien RM (2007) A caution regarding rules of thumb for variance inflation factors. *Qual Quant* 41(5):673–690. <https://doi.org/10.1007/s11135-006-9018-6>
- Oliveira E, Natarajan M, da Silva B (2019) Bus robberies in belo horizonte, Brazil: solutions for safe travel. *Crime Delinq* 0011128719871547
- Osgood DW (2000) Poisson-based regression analysis of aggregate crime rates. *J Quant Criminol* 16(1):21–43. <https://doi.org/10.1023/A:1007521427059>
- Peng C, Xueming S, Hongyong Y, Dengsheng Li (2011) Assessing temporal and weather influences on property crime in Beijing, China. *Crime Law Soc Chang* 55(1):1–13
- Quick M, Law J, Li G (2019) Time-varying relationships between land use and crime: a spatio-temporal analysis of small-area seasonal property crime trends. *Environ Plan B Urban Anal City Sci* 46(6):1018–1035
- Raento M, Oulasvirta A, Eagle N (2009) Smartphones: an emerging tool for social scientists. *Sociol Methods Res* 37(3):426–454
- Reynald DM (2009) Guardianship in action: developing a new tool for measurement. *Crime Prev Community Saf* 11(1):1–20. <https://doi.org/10.1057/cpcs.2008.19>
- Ridgeway G, Grogger J, Moyer RA, MacDonald JM (2019) Effect of gang injunctions on crime: a study of Los Angeles from 1988–2014. *J Quant Criminol* 35(3):517–541
- Ristea A, Al Boni M, Resch B, Gerber MS, Leitner M (2020) Spatial crime distribution and prediction for sporting events using social media. *Int J Geogr Inf Sci* 34(9):1708–1739
- Roman CG (2005) Routine activities of youth and neighborhood violence: spatial modeling of place, time and crime. In: Fahui W (ed) *Geographic information systems and crime analysis*. Hershey, PA, USA: IGI Global pp 293–310
- Roman CG, Reid SE, Bhati AS, Tereshchenko B (2008) Alcohol outlets as attractors of violence and disorder: a closer look at the neighborhood environment. The Urban Institute, Washington, DC
- Sherman L, Gartin PR, Buerger ME (1989) Hot spots of predatory crime: routine activities and the criminology of place. *Criminology* 27(1):27–55
- Silm S, Ahas R (2014) The temporal variation of ethnic segregation in a city: evidence from a mobile phone use dataset. *Soc Sci Res* 47:30–43. <https://doi.org/10.1016/j.ssresearch.2014.03.011>
- Song C, Koren T, Wang P, Barabási A-L (2010) Modelling the scaling properties of human mobility. *Nat Phys* 6(10):818–823
- Song G, Liu L, Bernasco W, Xiao L, Zhou S, Liao W (2018a) Testing indicators of risk populations for theft from the person across space and time: the significance of mobility and outdoor activity. *Ann Am Assoc Geogr* 108(5):1370–1388
- Song G, Liu L, Bernasco W, Zhou S, Xiao L, Long D (2018b) Theft from the person in urban China: assessing the diurnal effects of opportunity and social ecology. *Habitat Int* 78:13–20

- Song G, Bernasco W, Liu L, Xiao L, Zhou S, Liao W (2019) Crime feeds on legal activities: daily mobility flows help to explain thieves' target location choices. *J Quant Criminol* 35(4):831–854
- StataCorp (2017) Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC.
- Steenbeek W, Volker B, Flap H, Oort FV (2012) Local businesses as attractors or preventers of neighborhood disorder. *J Res Crime Delinq* 49(2):213–248
- Stucky TD, Ottensmann JR (2009) Land use and violent crime. *Criminology* 47(4):1223–1264
- Sun K, Yan D, Hong T, Guo S (2014) Stochastic modeling of overtime occupancy and its application in building energy simulation and calibration. *Build Environ* 79:1–12. <https://doi.org/10.1016/j.buildenv.2014.04.030>
- Sung H, Lee S, Cheon S (2015) Operationalizing Jane Jacobs's urban design theory: empirical verification from the great city of Seoul, Korea. *J Plan Educ Res* 35(2):117–130. <https://doi.org/10.1177/0739456x14568021>
- Taylor RB (1997) Social order and disorder of street blocks and neighborhoods: ecology, microecology, and the systemic model of social disorganization. *J Res Crime Delinq* 34(1):113–155. <https://doi.org/10.1177/0022427897034001006>
- Tillyer MS, Walter RJ (2019) Busy businesses and busy contexts: the distribution and sources of crime at commercial properties. *J Res Crime Delinq* 56(6):816–850. <https://doi.org/10.1177/0022427819848083>
- Tillyer MS, Wilcox P, Walter RJ (2021) Crime generators in context: examining 'place in neighborhood' propositions. *J Quant Criminol* 37(2):517–546
- Toomet O, Silm S, Saluveer E, Ahas R, Tammaru T (2015) Where do ethno-linguistic groups meet? How copresence during free-time is related to copresence at home and at work. *PLoS ONE* 10(5):e0126093. <https://doi.org/10.1371/journal.pone.0126093>
- Wagenmakers E-J, Ratcliff R, Gomez P, Iverson GJ (2004) Assessing model mimicry using the parametric bootstrap. *J Math Psychol* 48(1):28–50
- Wang X, Gerber MS, Brown DE (2012) Automatic crime prediction using events extracted from twitter posts. In: Yang S, Greenberg A, Endsley M (ed) *Social Computing, Behavioral - Cultural Modeling and Prediction*, Vol. 7227, Lecture Notes in Computer Science: Springer Berlin Heidelberg Pp. 231–38
- Weisburd D (2015) The law of crime concentration and the criminology of place. *Criminology* 53(2):133–157
- Weisburd D, Bernasco W, Bruinsma GJN (eds) (2009) *Putting crime in its place: units of analysis in geographic criminology*. Springer, New York
- Wilcox P, Quisenberry N, Cabrera DT, Jones S (2004) Busy places and broken windows? Toward defining the role of physical structure and process in community crime models. *Sociol Q* 45(2):185–207
- Williams ML, Burnap P, Sloan L (2017) Crime sensing with big data: the affordances and limitations of using open-source communications to estimate crime patterns. *Br J Criminol* 57(2):320–340
- Wo JC, Hipp JR, Boessen A (2016) Voluntary organizations and neighborhood crime: a dynamic perspective. *Criminology* 54(2):212–241
- Wooldridge JM (2010) *Econometric analysis of cross section and panel data*. MIT Press, London
- Wu Y, Sun IY, Rong Hu (2021) Chinese reluctance to report crime: political efficacy, group care and hukou. *Criminol Crim Just*. <https://doi.org/10.1177/17488958211017370>
- Xiao L, Ruiter S, Liu L, Song G, Zhou S (2021) Burglars blocked by barriers? The impact of physical and social barriers on residential burglars' target location choices in China. *Comput Environ Urban Syst* 86:101582
- Yu S-SV, Maxfield MG (2014) Ordinary business: impacts on commercial and residential burglary. *Br J Criminol* 54(2):298–320
- Yuan Y, Raubal M (2016) Analyzing the distribution of human activity space from mobile phone usage: an individual and urban-oriented study. *Int J Geogr Inf Sci* 30(8):1594–1621. <https://doi.org/10.1080/13658816.2016.1143555>
- Zeytinoglu IU, Cooke GB (2006) Part Iv: quality, efficiency and inequalities: Who is working at weekends? Determinants of regular weekend work in Canada. *Decent Working Time New Trends New Issues*, pp 395–416
- Zhang L, Messner SF, Liu J (2007) An exploration of the determinants of reporting crime to the police in the city of Tianjin, China. *Criminology* 45(4):959–984
- Zhang P, Zhou J, Zhang T (2017) Quantifying and visualizing jobs-housing balance with big data: a case study of Shanghai. *Cities* 66:10–22
- Zhuo Y, Messner SF, Zhang L (2008) Criminal victimization in contemporary China: a review of the evidence and challenges for future research. *Crime Law Soc Chang* 50(3):197–209

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

Guangwen Song<sup>1</sup> · Yanji Zhang<sup>2</sup>  · Wim Bernasco<sup>3,4</sup> · Liang Cai<sup>5</sup> · Lin Liu<sup>1,6</sup> · Bo Qin<sup>7</sup> · Peng Chen<sup>8</sup>

<sup>1</sup> Center of GeoInformatics for Public Security, School of Geography and Remote Sensing, Guangzhou University, Guangzhou, China

<sup>2</sup> Department of Sociology, School of Humanities and Social Sciences, Fuzhou University, Fuzhou, China

<sup>3</sup> Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Amsterdam, The Netherlands

<sup>4</sup> Department of Spatial Economics, School of Business and Economics, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

<sup>5</sup> Department of Sociology, University of Chicago, Chicago, IL, USA

<sup>6</sup> Department of Geography, University of Cincinnati, Cincinnati, OH, USA

<sup>7</sup> School of Public Administration and Policy, Renmin University of China, Beijing, China

<sup>8</sup> Institute for Information and Cyber Security, People's Public Security University of China, Beijing, China