Journal of Artificial Intelligence Research 68 (2020) 541-570 Submitted 11/2019; published 07/2020

Simulating Offender Mobility: Modeling Activity Nodes from Large-Scale Human Activity Data

Raquel Rosés RROSES@ETHZ.CH
Cristina Kadar CKADAR@ETHZ.CH
ETH Zurich, D-MTEC
8092 Zurich, Switzerland

Charlotte Gerritsen CS.GERRITSEN@VU.NL
VU University Amsterdam
1008 Amsterdam, Netherlands

Ovi Chris Rouly CHRIS.ROULY@GESS.ETHZ.CH
ETH Zurich, D-GESS
8092 Zurich, Switzerland

Abstract

In recent years, simulation techniques have been applied to investigate the spatio-temporal dynamics of crime. Researchers have instantiated mobile offenders in agent-based simulations for theory testing, experimenting with crime prevention strategies, and exploring crime prediction techniques, despite facing challenges due to the complex dynamics of crime and the lack of detailed information about offender mobility. This paper presents a simulation model to explore offender mobility, focusing on the interplay between the agent's awareness space and activity nodes. The simulation generates patterns of individual mobility aiming to cumulatively match crime patterns. To instantiate a realistic urban environment, we use open data to simulate the urban structure, location-based social networks data to represent activity nodes as a proxy for human activity, and taxi trip data as a proxy for human movement between regions of the city. We analyze and systematically compare 35 different mobility strategies and demonstrate the benefits of using large-scale human activity data to simulate offender mobility. The strategies combining taxi trip data or historic crime data with popular activity nodes perform best compared to other strategies, especially for robbery. Our approach provides a basis for building agent-based crime simulations that infer offender mobility in urban areas from real-world data.

1. Introduction

Criminology is a multidisciplinary research field that aims to explain, predict, and prevent criminal behavior. Although criminals only represent a minority of the overall population, people can come into contact with criminal behavior (either by being criminal or by being a victim) anytime or anyplace. Crime can be an intrusive component in everyday life.

One of the main research interests within criminology is understanding when crime will occur. The most influential theory addressing this challenge is the Routine Activity Theory (RAT) (Cohen & Felson, 1979). This theory states that crime will occur when a motivated offender meets a suitable target without a capable guardian present. Although this theory has shown itself to be very useful in explaining various criminological phenomena, it does not directly address the question of predicting where crime will occur. Based on RAT, the naïve assumption would be that crime is evenly distributed over time and space. However, it is known that the location(s) of criminal behavior are typically not evenly distributed over urban areas (Brantingham & Brantingham, 1993). So how can this uneven distribution be reproduced? Moreover, can we simulate offender mobility patterns reproducing such distributions? What data sources are useful to model strategies for offender mobility and for which types of crime do these perform best? In the current paper we address this question by generating a synthetic population of offender agents and simulating the ways in which they navigate the urban environment.

Previous studies (Brantingham & Brantingham, 1995) have shown us that higher crime concentration rates are found within an offender's awareness space. An awareness space is defined as the area frequently occupied by an offender. The awareness space of an offender can be determined, for example, by his home, workplace, recreation areas, etc., including the routes that lead to them. It follows, therefore, that the area lying 'between' frequently visited activity nodes should be the field of operation for offenders. Hence, to study the spatio-temporal dynamics of crime, we find that it is useful to examine the mobility patterns of offenders in detail and in situ. Due to the complex spatially and temporally distributed nature of these processes, a frequently used approach is to employ the simulation technology of Agent-Based Modeling. Indeed, previous authors and researchers have attempted to simulate crime patterns using ABM. Unfortunately, many of these simulations utilized highly incomplete data (e.g., based solely on police records of known offenders) or were not related to real-world data at all (Liu & Eck, 2008). These simulations often contained uninformed offender mobility strategies. As an alternative, this paper proposes an agent-based model (ABM), with no interaction between the agents, that describes offender mobility based on more complete, large-scale human activity data. This simulation is intended as a basis upon which to build more robust ABMs simulating offenders reproducing crime patterns, and will eventually be combined with cognitive models of offender decision making in order to predict when/where crimes will be committed.

As a case study, our model is applied to the surface street network of New York City (NYC), where a number of offender agent mobility strategies are compared to each other. Departing from the notion that crime is a legal definition and does not necessarily define group behavior (Tappan, 1947), the strategies developed here are not only inspired by theories in criminology but also use novel data as human activity proxies that could well represent offenders. For example we've inferred home addresses from census data (land use information and population density); venue location, venue type, and check-in counts from location-based social networks (LBSN) as a proxy for activity nodes and human activity; transitions within the city from taxi trip data as a proxy for travel patterns between city areas; and historic crime location data as a proxy for attractive city areas. We note that no robustness analysis has been conducted for the various datasets inferring the proxies. Consequently, our results must be interpreted with caution and may be regarded as preliminary. The performance of the model is assessed in terms of: (1) the ratio of crimes covered over distance traveled by the agents; and (2) crime locations covered within different areas of the city. Finally, we note that this model could be applied to study social behavior other than that of a criminal by adapting the performance measurement and by including other relevant environmental factors.

This paper is organized as follows. Section 2 describes related work and Section 3 introduces relevant notions for the purpose of this simulation. Section 4 introduces the data included in the simulation. The simulation model is presented in Section 5 and the results are shown in Section 6. In Section 7 we end this paper with a conclusion and outlook.

2. Related Work

Criminology (the study of crime) involves many aspects, whose inter-relationship may be mathematically complex. In the context of related work, we believe that computational social science (CSS) (i.e., using computational approaches to study social phenomena) has begun to present itself as an important explanatory tool for analyzing and predicting crime. We also note that the technologies of CSS, such as simulations, have emerged as tools with the potential to offer explanatory insight across many other complex social issues (Cioffi-Revilla, 2010), but we believe these tools to be particularly relevant for criminology. Across several fields and several decades, the technologies of CSS have consistently demonstrated interdisciplinary explanatory power, especially through the use of agent-based simulation (consider Axelrod, 1986; Crooks & Wise, 2013; Rouly, 2018; Schelling, 1969; Kohler, Kresl, van West, Carr, & Wilshusen, 2000).

In the field of criminology, in particular, scientists are discovering the power of agent-based simulation for various applications involving theory testing (Birks, Townsley, & Stewart, 2014; Brantingham & Tita, 2008; Groff, 2007a; Liu & Eck, 2008), testing of prevention strategies (Bosse & Gerritsen, 2010; Devia & Weber, 2013; Dray, Mazerolle, Perez, & Ritter, 2008; Gunderson & Brown, 2000), and forecasting the development of crime (Gunderson & Brown, 2000; Malleson, Heppenstall, & See, 2010; Peng & Kurland, 2014). Liu and Eck (2008) provide an overview of the basic characteristics of crime simulation models, while Groff, Johnson, and Thornton (2018) discuss practices, potentials, and shortcomings of existing ABM in relation to urban crime. In general, simulating crime patterns contributes to the understanding of crime in a spatial environment. First generation crime simulations were mainly built in synthetic environments without the use of real-world data (e.g., Brantingham & Tita, 2008) to study the underlying mechanisms of crime. However, including real data in a simulation allows an instantiation to support a more realistic environment and allows for a better transfer of the gained information, even though it may complicate the user's comprehension of underlying mechanisms. Indeed, existing simulation models have included street network and land use data, in combination with reported robbery data to test RAT, with basic offender agents moving between a set of static and predefined activity nodes, and deciding whether to offend (Groff, 2007b). Others have considered street and subway networks in combination with burglary data and agents moving between connected nodes at random and/or with heavy-tailed distribution waiting times (inspired by research on human mobility patterns) to test whether crime patterns can be reproduced (Peng & Kurland, 2014). Then, too, some have looked at street networks and household information (census and building data), in combination with reported burglary data, to gauge the utility of ABM for predicting crime. There we see agents modeled in a complex manner using PECS (Physical conditions, Emotional states, Cognitive capabilities, and Social status) (Urban & Schmidt, 2001). These latter simulations consider frameworks that model offender behavior as a series of random home and work locations where the agents build a cognitive map of possible targets within their awareness space (Malleson et al., 2010; Ward, Evans, & Malleson, 2016). One of the common elements characterizing all of the aforementioned simulations is their instantiation of offender behavior. All of these examples concentrate on the cognitive reasons for an offender to commit a crime by including agent-individual characteristics – e.g., wealth measure or target characteristics, and guardianship level of the possible targets, that factor into the offender's decision of whether to offend or not. In contrast, the offender agent mobility characteristics are rather neglected and based on simplified assumptions, with the exception of emergent crime patterns in a 2-D space by means of basic mathematical models (without any data) (Brantingham & Tita, 2008) and the Criminal Movement Model (CriMM) (Reid, Frank, Iwanski, Dabbaghian, & Brantingham, 2013). CriMM simulates travel patterns of known offenders from their residential address towards assigned attractors (major shopping centers). Travel paths are simulated between fixed residential and attractor locations, by means of the shortest path Dijkstra algorithm. The study focuses on analyzing the proximity of crime locations to the simulated traveled paths and identifies the potential of using major shopping centers as crime attractors to simulate travel path destinations.

Given all of the above, we ask: Is it possible that by explicitly modeling the movement of offenders, their direction choices, and distances traveled inferred from real-world open data, and by comparing random walks to more realistic non-random human movement, we might discover that a simple mobility rule could be used together with other behavior rules to reproduce crime patterns that allow for a better predictive result? In contrast to more traditional methods for generating a synthetic population representative of a city (Beckman, Baggerly, & McKay, 1996; Adigaa, Agashea, Arifuzzamana, Barretta, Beckmana, Bisseta, Chena, Chungbaeka, Eubanka, Guptaa, Khana, Kuhlmana, Mortveita, Nordberga, Riversa, Stretza, Swarupa, Wilsona, & Xiea, 2015; Burger, Oz, Crooks, & Kennedy, 2017), we rely on activity and mobility data to build strategies for offenders only, accounting for factors relevant to crime. Moreover, researchers have already shown the potential of using novel types of data in order to account for population at risk (also referred to as ambient population) rather than residential population for the purpose of crime analysis and prediction. Such data sources might include, for example, LBSN data (Kadar & Pletikosa, 2018; Wang, Schoenebeck, Zheng, & Zhao, 2016a), mobile data (Bogomolov, Lepri, Staiano, Oliver, Pianesi, & Pentland, 2014), or Twitter data (Gerber, 2014; Malleson & Andresen, 2014, 2015). We therefore argue that more realistic and generalizable offender (spatial-temporal) mobility behavior, built using novel data sources, would improve crime simulations.

Thus, in this paper, we consider the importance of studying the basic rules governing offender mobility by building a simulation model with a large number of offender-agent mobility strategy variations using large-scale mobility data for NYC and subsequently assess the value of those strategies using historic crime location patterns.

3. Criminal Offender Mobility

In RAT, *routine activities* are described as everyday activities that tend to happen at the same locations, such as home, work, and shopping areas. Offenders are thought to engage in routine activities, while research has shown that they are more prone to commit crimes close to the areas connecting the different activity nodes (Reid et al., 2013), i.e., within the offender's awareness space. Consequently, including offender agents' home locations and some set of activity nodes in a crime simulation is common practice. In such simulations crime is mainly represented on segments of the street network, the natural domain of police activities (Weisburd, Bushway, Lum, & Yang, 2004; Herrmann, 2013; Davies & Bishop, 2013; Kim, 2016; Rosser, Davies, Bowers, Johnson, & Cheng, 2016). On one hand, some of the models rely on police records to instantiate home addresses as starting points for the travel trajectories (Malleson, See, Evans, & Heppenstall, 2014). Such a setup is constrained to simulating offenders known to the police and especially those for whom home addresses have been reported. On the other hand, little effort has been devoted to defining appropriate activity nodes and reproducing realistic human (e.g., offender) spatial-temporal mobility patterns in simulations. In the era of social media and crowdsourced/location-based user data (Crooks & Wise, 2013), patterns of human activity can be inferred from openly available data. Human mobility patterns have been intensively studied by means of GPS (Global Positioning System) generated user data (Gonzalez, Hidalgo, & Barabasi, 2008; Song, Qu, Blumm, & Barabasi, 2010), as well as by means of LBSN such as Foursquare (Noulas, Scellato, Lambiotte, Pontil, & Mascolo, 2012; Hecht & Stephens, 2015), and even taxi trip data (Tang, Liu, Wang, & Wang, 2015). Such research has confirmed the high regularity of individual human movement and determined basic rules governing it, e.g., suggesting individual human travel distances should be modeled by means of Lévy flight. This is the name given to an actor's set of seemingly random spatial movements where those actual incremental displacements are better represented by a heavy-tailed probability distribution (Mandelbrot, 1982), the probability of short displacements is high while the probability of long displacements is low, hence modeling very well human mobility. Existing research has shown, that not only can we gain information about human mobility patterns from LBSN, but that this information is especially reliable within urban areas as opposed to rural areas (Noulas, Scellato, Mascolo, & Pontil, 2011; Noulas et al., 2012; Cranshaw, Schwartz, Hong, & Sadeh, 2012; Nguyen & Szymanski, 2012; Hecht & Stephens, 2015). Rules governing human mobility behavior extracted from LBSN coincide with those studied for the general population extracted from other methods such as GPS traces from mobile phones (Cheng, Caverlee, Lee, & Sui, 2011). The advantage of LBSN data over cell phone data, is the possibility to derive information regarding the context of the activities. For instance, in Foursquare, users check into locations which are broadly categorized by type of activity (Noulas, Mascolo, & Frias-Martinez, 2013; Cranshaw et al., 2012). Criminology has identified a link between types of activities available in an area of the city and the types of people visiting this area, emphasizing the importance of attractive locations – facilities such as shopping centers, malls, schools, restaurants, and bars– in generating and attracting crime (Brantingham & Brantingham, 1995). Certain city areas can offer a greater number of criminal opportunities simply due to a greater popularity among visitors, i.e., number of possible victims (crime generators), while other areas attract offenders due to the availability of easy targets related to the type of activities in the area, e.g., drunk visitors in bar areas (crime attractors). Thus, information about the location type and popularity can be used as a proxy to model the attractiveness for victims and offenders of specific locations within urban areas (Reid et al., 2013). Far from being perfect, LBSN data present some biases in terms of user characteristics and use cases. Existing research has raised concerns about the representativeness of check-ins due to bias in the user groups and the possibility of false check-ins (Zhang, Zhou, Zhao, Wang, Su, Metzger, Zheng, & Zhao, 2013; Wang et al., 2016a). However, the research community relies on such data due to the ease of accessibility and lack of more representative equivalent alternatives. Previous research has used LBSN check-ins and venues as a proxy for human activity, furnishing insights into the activities unfolding in different neighborhoods (Noulas et al., 2013). Researchers have already identified potential and risks of using crowd-sourced data for crime analysis (Malleson & Andresen, 2014, 2015) and have successfully used LBSN for crime prediction models (Bogomolov et al., 2014; Al Boni & Gerber, 2016; Wang, Kifer, Graif, & Li, 2016b; Yang, Heaney, Tonon, Wang, & Cudré-Mauroux, 2017; Kadar & Pletikosa, 2018). Other data sources such as taxi trip data, can provide insights into frequent travel volume from one region of the city into others. Existing research has used taxi trip data to study human mobility patterns (Liu, Kang, Gao, Xiao, & Tian, 2012a; Liu, Wang, Xiao, & Gao, 2012b), reveal city structure (Liu, Gong, Gong, & Liu, 2015), mine attractiveness of city areas (Yue, Zhuang, Li, & Mao, 2009), and to infer crime rates at neighborhood level (Wang et al., 2016a), among others. Travel patterns extracted from taxi trip data relate to the type and intensity of activities occurring in different city areas, while areas of the city drawing higher numbers of visitors act as crime attractors and generators, resulting in a higher number of crimes (Kinney, Brantingham, Wuschke, Kirk, & Brantingham, 2008). Even though taxi trip data only represents a group of drivers within the network, the high penetration of the service in NYC and the fact that it is commonly used for social and recreational trips makes it a good proxy for this study (New York State Department of Transportation, 2018). Taxi trip data has successfully been used to inform crime prediction models (Wang et al., 2016b; Kadar & Pletikosa, 2018) and can act as a proxy for transitions between city areas. Moreover, activity nodes, and city centers as a special case, have been identified as attracting offenders as well as the general population (Frank, Dabbaghian, Reid, Singh, Cinnamon, & Brantingham, 2011).

4. Data Outline

NYC is home to over 8.5 million inhabitants and is the most densely populated major city in the United States, with around 27,000 people per square mile. In 2015, the NYC government launched the NYC Open Data platform to share the data produced and used by the city's government. The platform, which contains datasets concerning the inhabitants and environment of the city, encourages the citizens and researchers to use the data for value creation. The combination of a large urban population, an open data platform, and widespread digitization of services (e.g., LBSN such as Foursquare) creates enormous amounts of data that can be used to simulate the use of space and activities happening in NYC. Indeed, NYC is the most popular city on Foursquare, an LBSN application in which users check into places as part of their daily activities. Foursquare boasts almost 300,000 venues and a total of 132 million check-ins in NYC (as of May 2016)[1]. In this paper we simulate a simple geographic virtual environment of NYC using openly accessible data. The data is projected onto the NYC area (projected coordinate per North American Datum of 1983), allowing measurement in feet. In this section we offer a detailed description of the data which will build the simulation. We have gathered data to simulate a crime pattern for

1. `https://www.4sqstat.com/`

June 2015. For the simulation the data is projected onto two different levels of granularity, depending on the use case:

- street segments of the street network (see Figure 1), collected from the NYC Open Data Portal in 2016. The NYC street network contains 117,320 street segments [lat/long line coordinates] and provides the structure of the street and public transportation system (including ferry lines). The dataset has been cleaned, removing 4,138 isolated street segments in total.

- census tracts (CT), a statistical unit subdividing counties [lat/long polygon coordinates] defined by the United States Census Bureau[2] for the New York region (see Figure 2). In NYC there are 2,168 CTs with populations ranging from 3,000 to 4,000 and an average land area of 90 acres. The dataset has been cleaned to remove 6 CTs containing only water and shorelines.

Figure 1: Street network in NYC.

Figure 2: Census tracts in NYC.

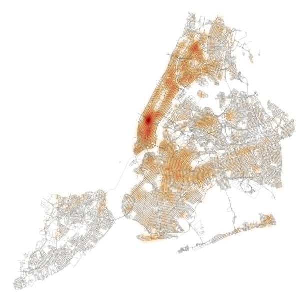

Figure 3: Crime pattern heat map, June 2015.

Census data: From the NYC American Community Survey Public Use Microdata Sample (ACS 2011-2015 5-year PUMS)[3], we extracted census tract population total for the period of 2011-2015 and calculated population density for each CT. The ACS PUMS data gives information about the populations within the different census tracts. ACS 2016 PUMS data was not yet available when data for this study was collected; however, the overall changes in population from one year to another are less then 0.5%. Land use information from the PLUTO[4] database (2016) is used to identify residential areas filtering for one- and two-family buildings, multi family buildings, and mixed residential buildings. The PLUTO data is collected on tax lot (a feature class part of the Department of Finance's Digital Tax Map) level, which identifies each parcel of the city. We combine this information to build a map with residential street segments and resident population density, i.e., each residential street segment is weighted by the population density of the CT it is in.

Crime data: The NYPD (New York Police Department) complaint data[5] contains felony crimes reported to the police. The dataset includes information such as type of

2. `http://www.census.gov/`
3. `https://www.census.gov/programs-surveys/acs/data/pums.html`
4. `https://www1.nyc.gov/site/planning/data-maps/open-data/dwn-pluto-mappluto.page`
5. `https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Current-Year-To-Date-/5uac-w243/data`

crime, date, time, and location [lat/long point coordinates at the middle or any end of a street segment] for each crime. In terms of type of crime, our dataset includes burglary, grand larceny, grand larceny of motor vehicle, robbery, and felony assault. Instances of rape and murder have been excluded from the dataset due to low incidence rates (e.g., 1,209 and 357 incidents in 2015, respectively). Crime data is used for three different purposes: (1) crime data from June 2014 to May 2015, aggregating crime counts per type of crime over CT's, is used within one strategy of the simulation to instantiate the attractiveness level of a given CT for offenders within the last 12 months. This is useful because long-term past crime data (e.g., at least 12 months) is a good indicator for future crime (Groff & La Vigne, 2002) and is therefore often used for training crime prediction models. (2) Crime data from June 2015, aggregating crime counts per type of crime over CT's, is used for model performance assessment (i.e., testing). (3) Crime data from June 2015 on segments of the street network is used for model performance assessment (i.e., testing). Figure 3 shows the crime pattern in a heat map for all crimes in June 2015 and Table 1 details the number of crimes per crime type and dataset. Note that while aggregating crime locations at CT level, crimes on street segments traversing several CTs are counted twice, once in each CT.

Crime type	**Count**
June 2014-May 2015	
Total crimes	102,966
Burglaries	15,897
Grand larcenies	43,301
Grand larcenies of motor vehicles	7,523
Robberies	16,413
Felony assaults	19,832
June 2015	
Total crimes	8,503
Burglaries	1,287
Grand larcenies	3,555
Grand larcenies of motor vehicles	580
Robberies	1,303
Felony assaults	1,778

Table 1: Crime data counts.

LBSN data: Foursquare is an LBSN allowing online interaction between users based on their physical location. Users share their real-time location and check into venues they visit. Foursquare data was collected from the Foursquare API (Application Programming Interface)[6] in June 2016, as in Kadar, Iria, and Pletikosa Cvijikj (2016) as well as Kadar, Rosés Brüngger, and Pletikosa Cvijikj (2017). Figure 4 shows a heat map of the Foursquare venue locations. The dataset contains information such as venue name, location [lat/long point coordinates], check-in counts (accumulated over time), associated categories. The categorization of venues includes arts and entertainment, college and university, events, food, nightlife, shops and services, traveling and transportation, etc. This dataset is used to simu-

6. `http://www.foursquare.com/`

late popularity of locations and to provide context of the activities at the specific locations. It is composed of 236,294 venues in the proximity of every incident from the crime dataset and includes over 119 million check-ins (from the creation of the venue in the platform until the point of data collection date in June 2016) associated with venue categories (i.e., types). The venues have been mapped to the streets of the NYC street network[7]. Table 2 shows the count of different venue types and cumulative check-ins associated with each venue type.

Venue type	**Venue count**	**Check-ins**
Food	47,590	37,906,768
Outdoors & Recreation	18,011	16,397,944
Shop & Service	62,627	16,008,377
Professional & Other Places	64,055	14,118,615
Nightlife Spot	11,140	12,382,224
Travel & Transport	13,911	11,939,597
Arts & Entertainment	11,794	8,011,780
College & University	7,082	3,094,084
Event	84	20,386
Total	236,294	119,879,775

Table 2: Foursquare venue types.

Taxi trip data: The NYC Open Data Platform offers very large sets of taxi trip data. We combine Yellow Taxi Trip Data[8] and Green Taxi Trip Data[9] into one dataset for a one-year time period (July 2014 to June 2015). Both yellow and green taxi services pick up passengers hailing from the street and cover all of NYC, while serving different city areas (i.e., pick-up locations in different areas). Yellow taxis are concentrated around Manhattan as wel as the JFK International Airport and LaGuardia Airport; green taxis offer their services above 110th Street in Manhattan and in the outer boroughs of NYC. The dataset includes pick-up and drop-off dates/times/locations [lat/long point coordinates], trip distances, fares, rate types, payment types, and driver-reported passenger counts. For this study we use pick-up and drop-off locations and project them on CTs, creating a new dataset pairing CTs with each other and weighting by total number of pick-ups and drop-offs. Thus, the dataset reveals information about the connectivity and popularity of transitions from one CT to any other CT in the city. The dataset is composed of over 248 million taxi trips within 12 months (July 2014 to June 2015). Figure 5 shows the counts of taxi trip pick-ups and Figure 6 shows the counts of taxi trip drop-offs at CT level.

Table 3 summarizes the data described in this section and shows how it is used in the simulation model described in the next section.

5. The Simulation Model

Inspired by previous ABMs simulating crime, we built a simulation model to explore several scenarios of offender mobility strategies. The model aims to generate RAT-based patterns of individual offender movement which cumulatively match the spatial distribution of historic

7. 8,254 venues where not within 80 feet of a street segment, and therefore not included in this study
8. `https://data.cityofnewyork.us/Transportation/2014-Yellow-Taxi-Trip-Data/gn7m-em8n`
9. `https://data.cityofnewyork.us/Transportation/2014-Green-Taxi-Trip-Data/2np7-5jsg`

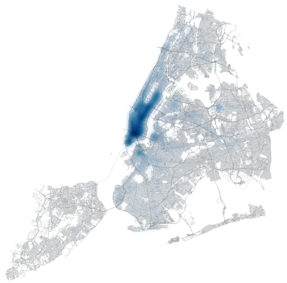

Figure 4: Foursquare venues heat map.

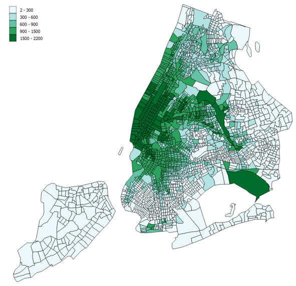

Figure 5: Taxi trip pickup locations count on CT level.

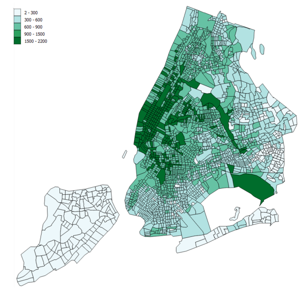

Figure 6: Taxi trip drop-off locations count on CT level.

Layer	**Representation**	**Use**
Street network	street network	travel paths
Census data	residential areas	home locations, starting and ending position
Foursquare	activity nodes and activity	activity node locations and popularity, destinations
Taxi trip data	transition trends between CTs	destination area (i.e., CT) popularity
Crime data	attractiveness of CT	destination area (i.e., CT) popularity

Table 3: Use of data layers within the simulation.

crime locations. Due to the fact that this model focuses only on mobility strategies, the offender agents engage in movement throughout the street network, traveling past locations with historic crime points but without engaging in any criminal action or interaction within this simulation. The number of historic crime locations passed by the agents is used for model performance assessment only. In order to simulate offender mobility, the following aspects are relevant: (1) the optimal number of agents influencing the spatial coverage area; (2) the characteristics of the simulation model environment, including a street network, spatial destinations representing activity nodes, and geo-spatial historic crime data; (3) the agents, starting positions affecting the future possibilities due to path dependency; and (4) the movement preferences and strategies of the agents. These points are formalized in Section 5.1.

Using Mesa, an agent-based modeling framework in Python (Masad & Kazil, 2015), a simplified version of NYC is instantiated in this model, providing the structure of the street network, zoning features for residential areas with population densities, venues with their popularity from location-based social networks, aggregated taxi trip data, and aggregated crime data including crime locations per type of crime (burglary, robbery, grand larceny, larceny of motor vehicles, and felony assault)[10].

5.1 Basic Functionality Formalization

The variables in Table 4 are used in the following subsection to introduce the model in detail. The simulation emulates a large-scale urban environment with offender agents traveling the street network. Each simulation step (epoch) represents one day of the month (24 hours)

10. The python and sql code for this simulation is available on GitHub: `https://github.com/rraquel/ABM-crime-mobility-NYC`

and the model runs for 30 days, consistent with one-month crime data emulated in the environment. The performance of all agents within each scenario is evaluated after the total period of 30 days.

Model var.	Explanation	Agent var.	Explanation
n	number of agents instantiated	s	starting position
rd	street in NYC street network	x	travel destination
c	historic crime locations	tc	traveled crime location
d	distance (length) of NYC street network	td	traveled distance
v	activity node (i.e., Foursquare venue)	r	radius distance
$atrip$	average number of travel trips in a day	a	search area
		$xtrip$	number of trips in a day

Table 4: Variables in the simulation model.

In detail, the instantiated agents are created and newly positioned at each step (i.e., day) to reduce path dependency bias. They start and end each step at a location s, representing the home location. As the agents instantiate anonymous offenders, they are placed on the closest street to a residential building. s is inferred from residential areas weighted by the population density of each area. The population density of each CT influences the probability of each residential building to be chosen as an s; higher density corresponds to higher probability.

Within each step, the offender agents travel from a starting position s to several destination positions x. Each agent draws the number of trips to destinations $atrip$ for the current day (i.e., step) from $U(0, 2 \times atrip)$, while $atrip$ is the statistical average number of trips undertaken by the NYC population (3.8 trips per day) (New York State Department of Transportation, 2012). For each travel trip, an agent chooses a destination area a and a destination x within a. The choice of a and x depends on the offenders' mobility strategy, detailed in Section 5.2. A scenario is simulated for each possible offender mobility strategy. The agents travel using Dijkstra's shortest path algorithm[11], taking into account street segment length. Over one model run (30 steps), each agent collects information about the historic crime locations c encountered along the traveled paths tc and the distance traveled td. The performance of the scenarios is assessed using the cumulative tc and td over all agents within a simulation (see Section 5.3 for more details). The agents' behavior is independent from other agents and starts at a new location every day, thus there is no difference between running the simulation multiple times or running it for a high number of agents within the simulation. Additionally, the adequate (Swarup, 2019) number of runs or agents within each scenario is unknown. Drawing upon the insights from our previous preliminary work (Rosés, Kadar, Gerritsen, & Rouly, 2018), we decide to fix the running time (30 steps) and instantiate 1,000 agents within each simulation, a largely sufficient number of agents to cover all historic crime locations (multiple times). This results in 30,000 individual agents simulated for each scenario. To assess the performance of the model, and to find the adequate number of agents within the simulation, we evaluate the results for varying numbers of agents, meaning that we only account for the desired number of agents, only taking into

11. `https://networkx.github.io/documentation/stable/reference/algorithms/shortest_paths.html`

account the paths of the first desired number of agents for each simulated day within a model run.

5.2 Mobility Scenario Strategies

We build 35 mobility scenarios by combining 5 options for area-selection strategies and 7 options for destination-selection strategies. For each travel trip, agents first choose an area to travel to, followed by a concrete location within the same area. The area can either be a radial distance or a CT, whereas the concrete destination location is always on a street segment (i.e., Foursquare venues are mapped to street segments). In line with the knowledge described in Section 3, the strategies are detailed as follows:

Area selection:

1. **Static**: The static distance allows agents to move only within a specific distance from their current location, resulting in a radial area a. The static distance is set to a radius of 40,000 feet with a 5% boundary, inspired by the average trip length for NYC's population (New York State Department of Transportation, 2012), and represents a baseline strategy.

2. **Uniform**: The uniformly distributed distance builds upon the static distance, uniformly drawing distances from a distribution with an average trip length for NYC's population: $R \sim U(0, 2r)$ so that $E[R] = r$ with a 5% boundary; this results in a radial area a. Consequently, each traveled distance varies, representing an improved baseline.

3. **Power**: The Lévy flight distance draws distances from a power law distribution using Lévy flight. Lévy flight distribution is introduced to mimic realistic distance choices. The Lévy flight formula is transformed to allow for the drawing of distances (r) from the probability distribution within NYC, with β =0.6, determined to be the optimal value for NYC (Brockmann, Hufnagel, & Geisel, 2006), and an extra boundary of 5%, resulting in a radial area a:
$$P(r) \sim r^{-(1+\beta)} \rightarrow r \sim \frac{1}{P(r)} \times \mathrm{e}^{\frac{1}{1+\beta}} \tag{1}$$

4. **Taxi**: The taxi distance provides agents with a list of destination areas corresponding to census tracts weighted by the frequency of trips between the CT of origin and any other census tract in NYC. Census tracts with higher transition frequencies are weighted higher. The aggregated taxi trip data is included as a proxy for more realistic travel preferences.

5. **Crime**: The crime distance provides agents with a list of destination areas corresponding to CTs weighted by crime location counts (all crimes combined) and by their distance to the starting position s. CTs with higher historic crime location counts and closer to the CT of origin are weighted higher. The aggregated historic crime data is included as a proxy for the general attractiveness of areas in line with previous crime locations. As stated earlier, historic crime is a good indicator for future crime.

Destination selection within area *a*:

1. **Random streets**: The first option is the most basic one, offering any random street of the street network as a destination.

2. **Random venues**: The second option offers a choice of any random activity node (i.e., Foursquare venue) as a destination.

3. **Random venues-center**: The third destination option accounts for the attractiveness of the city center using a center score, allowing a choice of any activity node, and weighting activity nodes in proximity of the center of NYC higher. The center score assigns values from 10 to 100 to the venues, decreasing in value with increasing distance from the city center.

4. **Random venues-type**: The fourth option offers any activity node weighing nodes with more popular activity types higher. The popularity of the activity type at the node is determined by the total check-ins count per venue category in all of NYC.

5. **Popular venues**: The fifth strategy offers a choice of activity nodes weighted by popularity (determined using check-in counts from Foursquare). The higher the number of check-ins at the venue, the higher the weight of the activity node.

$$P[x] = \frac{\textit{check-ins}}{\Sigma\ \textit{check-ins within } r} \tag{2}$$

6. **Popular venues-center**: The sixth strategy offers a choice of activity nodes weighted by popularity (determined using check-in counts from Foursquare) and by proximity to the center of the city, using the center score (described in item 3 of this section). The higher the number of check-ins and the closer the venue is to the city center, the higher the weight of the activity node.

$$P[x] = \frac{\textit{check-ins}}{\Sigma\ \textit{check-ins within } r} \times \textit{center\_score} \tag{3}$$

7. **Popular venues-type**: The seventh strategy offers a choice of activity nodes weighted by popularity (determined by check-in counts from Foursquare) and by popularity of the activity type (total check-ins count per venue category in all of NYC). The higher the number of check-ins at the venue and the higher the number of check-ins for the activity type in all of NYC, the higher the weight of the activity node.

$$P[x] = \frac{\textit{check-ins}}{\Sigma\ \textit{check-ins within } r} \times \frac{\Sigma \textit{check-ins category}}{\Sigma\ \textit{check-ins total}} \tag{4}$$

5.3 Performance Assessment

The cumulative number of crimes and distance traveled over the agents of each simulated scenario is used to assess the performance of each scenario. We develop two metrics at different levels of granularity to assess the performance of the simulated scenarios. First, we develop a metric for assessing the performance of the scenarios at street segment level.

We adopt the Predictive Accuracy Index (PAI) (Chainey, Tompson, & Uhlig, 2008) and adapt it for use within the context of our simulations. PAI is a standard measure applied in criminology to evaluate performance of crime prediction models, overcoming the challenges posed by sparseness of point processes for performance measurement. The original PAI, was specifically developed to assess the performance of models predicting crime hotspots (i.e., areas of a map with high crime intensity). The equation considers hit rate of crimes against prediction area with respect to the total area. See Equation (5), where *HitRate* is the percentage of predicted crimes within the prediction area and *AreaPercentage* is the prediction area in relation to the whole study area. The value of PAI is 1, if the model predicts all crimes in the whole study area.

$$PAI = \frac{HitRate}{AreaPercentage} \tag{5}$$

Inspired by PAI, in order to assess the performance of the simulations in this study, we evaluate the relationship between two ratios: (1) distinct crime locations traveled (each crime only counted once) over the total crime locations as *TraveledCrimesRatio* and; (2) the distinct distances traveled (i.e., length of streets within the street network, only counted once) by the agents over the travel space (i.e., total length of the street segments) as *TraveledDistanceRatio*. We have therefore adapted the PAI index as in Equation (6). The resulting index for adapted PAI shows better model performance if the resulting value is higher, meaning that the simulation has covered a higher number of crime locations per distance.

$$adapted\ PAI = \frac{TraveledCrimesRatio}{TraveledDistanceRatio} = \frac{\frac{\Sigma\ tc}{\Sigma\ c}}{\frac{\Sigma\ td}{\Sigma\ d}} \tag{6}$$

Using the adapted PAI measure, we choose a limited number of successful scenarios and conduct further analysis to assess their performance at CT level. We compare the coverage area (crime locations along the paths traveled by the agents) of different CTs within one scenario, obtaining information about whether the agents cover crime locations equally across various CTs of the city.

Furthermore, the optimal number of agents within the simulation is determined by comparing the performance of the simulations when run with different numbers of agents, ranging from 5 to 1,000. Note that no significance test was conducted for comparing the performance of different scenarios following the recommendations of White, Rassweiler, Samhouri, Stier, and White (2014), who advised against it for social simulations.

6. Simulation Results

For the purpose of assessing the performance of various offender mobility strategies described in the previous section, we ran multiple simulations, one for each different scenario with the maximum number of agents (i.e., 1,000 agents). In the following subsections: (1) we highlight the most interesting results over all simulated scenarios for all types of crimes and choose the two best performing strategies; (2) we engage in a deeper analysis of the scenario performance for different types of crimes; and (3) we assess the spatial performance of the best strategy at the CT level.

6.1 Scenario Performance for All Types of Crimes

We explore the performance of the scenarios at street segment level for all types of crimes. Each of the 35 scenarios is evaluated in terms of adapted PAI for a varying number of simulated agents n (5, 25, 50, 75, 100, 125, 150,..., 1,000). To ease readability of the overall result, we've grouped the adapted PAI results into five graphs, one for each area strategy in combination with the various destination strategies (see Table 5).

Area strategy	**Destination strategy**	**Figure**
Static area strategy	all destination strategies	Figure 7
Uniformly distributed area strategy		Figure 8
Power-law distributed area strategy		Figure 9
Taxi area strategy		Figure 10
Crime area strategy		Figure 11

Table 5: Variables in the simulation model.

A preliminary visual inspection of the resulting graphs reveals the consistent underperformance of the most basic destination strategy (offering a choice between random streets) compared to more elaborate destination strategies across all five figures. The remaining destination strategies perform rather similarly and can be split into the following broad categories: (1) proxies for activity nodes (random venues, random venues-center, random venues-type) and (2) proxies for human activity at these nodes (popular venues, popular venues-center, popular venues-type), with the latter showing overall slightly higher adapted PAI values throughout Figures 7-11.

For a thorough investigation of the overall performance of each scenario, we applied a holistic measure. We considered the area under the curve (AUC) for each result line in the graphs depicted in Table 5 and show the resulting values in Table 6. This allowed us to compare the average performance of each scenario in terms of adapted PAI over a varying number of agents. Overall, combining static area strategy with popular venues-center performs best, showing an AUC value corresponding to an average adapted PAI of 1.35. The scenarios combining static area with popular venues-type (1.34 average adapted PAI) and static area with popular venues (1.33 average adapted PAI) followed as second and third best performing scenarios overall. Conversely, power-law, uniform random, and static area selection strategies, each combined with random streets, perform worst, with average adapted PAI values between 1.15 and 1.18.

Defining the most basic strategy (static area combined with random streets destination) as the baseline, we compare the relative AUC improvement of each scenario while grouping destination selection strategies by area selection strategies. In Table 6, we conclude that a static area selection strategy combined with a popular venues-center strategy performs best, yielding a 14.31% improvement over the baseline. This is followed by crime area strategy combined with popular venues-type, which exhibits an improvement of 12.63%. Using a power-law strategy combined with popular venues-center results in only a 9.75% improvement. Furthermore, when empirical taxi trip data is combined with a simple popular venues strategy, a 9.46% improvement is seen. Finally, a uniform random area strategy combined with a popular venues-center strategy delivers an improvement of only a 9.03%. However, a pattern emerges as the special cases for popular venues perform best within each

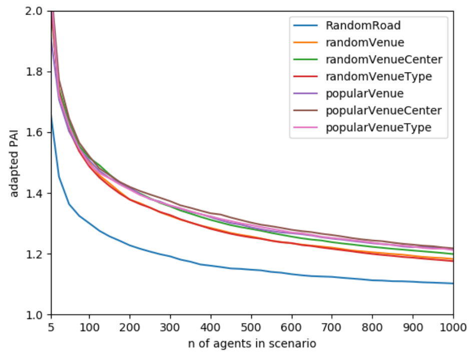

Figure 7: Adapted PAI (all crime types) for static distance and number of agents.

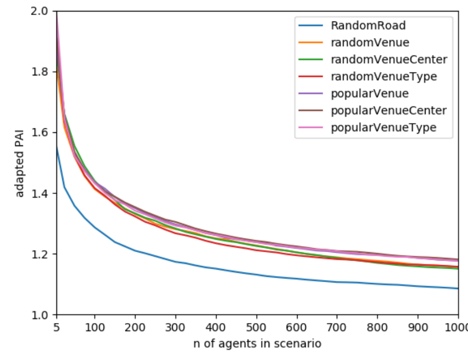

Figure 8: Adapted PAI (all crime types) for uniform distance.

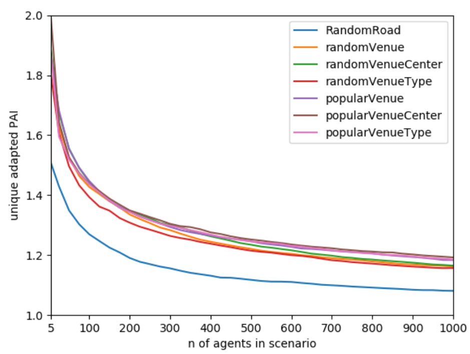

Figure 9: Adapted PAI (all crime types) for power distance.

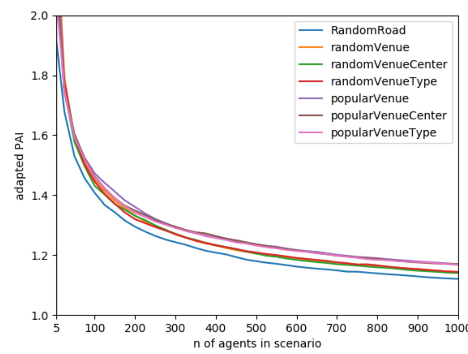

Figure 10: Adapted PAI (all crime types) for taxi area.

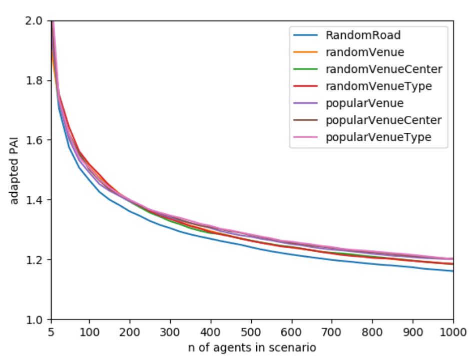

Figure 11: Adapted PAI (all crime types) for crime area.

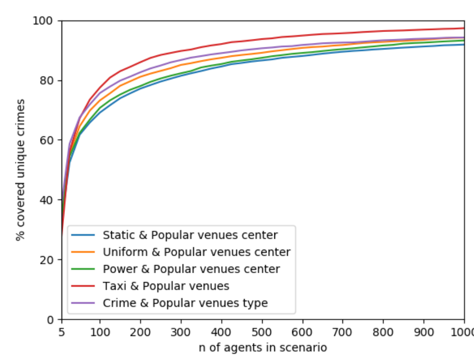

Figure 12: Crime locations coverage for the 5 best performing scenarios.

Distance	**Destination**	**AUC**		
		Overall	**Avg. adapted PAI**	**Improvement**
Static	Random streets	1,172.38	1.18	0.00%
	Random venues	1,299.69	1.31	10.86%
	Random venues-center	1,322.00	1.33	12.76%
	Random venues-type	1,296.07	1.30	10.55%
	Popular venues	1,325.57	1.33	13.07%
	Popular venues-center	**1,340.10**	1.35	**14.31%**
	Popular venues-type	1,328.54	1.34	13.32%
Uniform	Random streets	1,156.94	1.16	-1.32%
	Random venues	1,256.75	1.26	7.20%
	Random venues-center	1,259.40	1.27	7.42%
	Random venues-type	1,251.11	1.26	6.72%
	Popular venues	1,274.25	1.28	8.69%
	Popular venues-center	1,278.21	1.28	**9.03%**
	Popular venues-type	1,273.78	1.28	8.65%
Power	Random streets	1,144.43	1.15	-2.38%
	Random venues	1,259.58	1.27	7.44%
	Random venues-center	1,271.79	1.28	8.48%
	Random venues-type	1,246.03	1.25	6.28%
	Popular venues	1,280.04	1.29	9.18%
	Popular venues-center	1,286.73	1.29	**9.75%**
	Popular venues-type	1,278.11	1.28	9.02%
Taxi	Random streets	1,225.00	1.23	4.49%
	Random venues	1,260.04	1.27	7.48%
	Random venues-center	1,254.06	1.26	6.97%
	Random venues-type	1,257.54	1.26	7.26%
	Popular venues	1,283.25	1.29	**9.46%**
	Popular venues-center	1,279.06	1.29	9.10%
	Popular venues-type	1,275.94	1.28	8.83%
Crime	Random streets	1,278.31	1.28	9.04%
	Random venues	1,305.52	1.31	11.36%
	Random venues-center	1,304.03	1.31	11.23%
	Random venues-type	1,306.97	1.31	11.48%
	Popular venues	1,311.59	1.32	11.87%
	Popular venues-center	1,317.35	1.32	12.37%
	Popular venues-type	1,320.49	1.32	**12.63%**

Table 6: Overall performance comparison for all scenarios, improvement over static combined with random streets.

of the area strategies. Additionally, the difference in performance among popular venues, popular venues-center, and popular venues-type is very small, within each scenario grouped by area strategies.

From the previous analysis, we observed the best performing scenarios for each area selection strategy and used this information to analyze the efficiency of those scenarios by looking into the percentage of crime locations covered within each simulated scenario (see Figure 12). We define efficiency as achieving the highest adapted PAI value while covering a reasonable proportion of the crime locations within a simulation and requiring the lowest number of agents (percentage of crime locations along the agents traveled paths). For covering 80% and 90% of crime locations, we determined the adapted PAI value and number of agents (see Table 7). To cover 80% of total crime locations, the adapted PAI values vary between 1.33 and 1.44, while the highest adapted PAI value is achieved by taxi area combined with popular venues, with only 125 agents within the simulated scenario. In turn, to cover 90% of the total crime locations, the values for adapted PAI vary between 1.25 and 1.29; the highest value is achieved by the scenario combining crime with popular venues-type for 475 agents, noting that taxi combined with popular venues achieves a very similar adapted PAI value (1.28) for only 325 simulated agents. These results give us an idea about the number of agents needed to simulate each scenario, depending on the desired coverage of crime locations.

Distance	**Destination**	**80 % crime locations coverage**		**90 % crime locations coverage**	
		n	**Adapted PAI**	**n**	**Adapted PAI**
Static	Popular venues-center	275	1.38	775	1.25
Uniform	Popular venues-center	200	1.35	575	1.23
Power	Popular venues-center	250	1.33	675	1.22
Taxi	Popular venues	**125**	**1.44**	325	1.28
Crime	Popular venues-type	175	1.42	**475**	**1.29**

Table 7: Efficiency and coverage of crime locations within the simulation.

6.2 Performance for Single Types of Crimes

In this subsection, we engage in a deeper analysis of the two best performing strategies determined by the analysis conducted so far. In particular, we look into the performance of taxi area combined with popular venues and crime combined with popular venues-type, over adapted PAI by varying the number of agents for different types of crime: burglary, robbery, grand larceny, larceny of motor vehicle, and felony assault. See Figure 13 for taxi combined with popular venues and Figure 14 for crime combined with popular venues-type. A visual inspection of the graphs reveals a clear over-performance of the scenarios for robbery, followed by grand larceny, which performs similarly to all types of crimes combined. Both scenarios under-perform for the remaining crime types (burglary, grand larceny of motor vehicle, and felony assault) as compared to all types of crimes aggregated. We note that both over-performing crime types can be grouped into a larger category referred to as "street crimes". Consistent with the analysis in the previous subsection, we show in Table 8 that there is value in the application of a holistic measure for assessing the overall performance of the different crime type within the scenarios. We calculate AUC and the corresponding average PAI over varying numbers of agents, as well as the percentage of AUC improvement over the baseline (all crimes combined). This results in two baselines, one for each scenario.

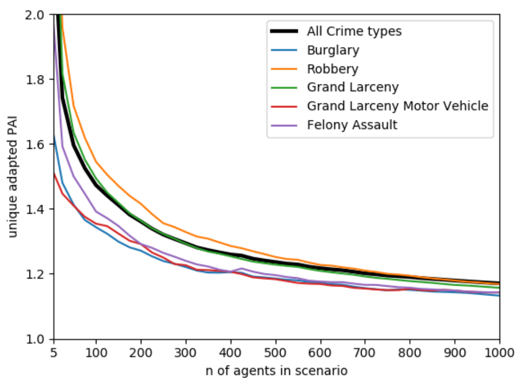

Figure 13: Adapted PAI for different types of crimes in Taxi & Popular venues.

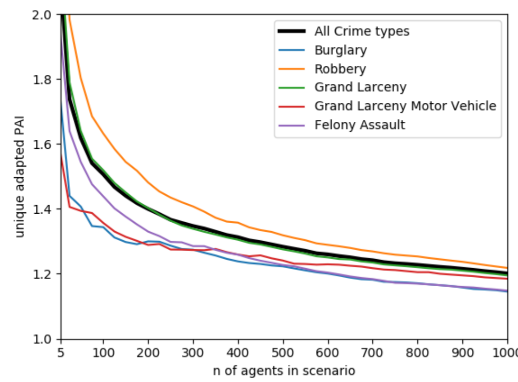

Figure 14: Adapted PAI for different types of crimes in Crime & Popular venues-type.

Scenario	**Crime type**	**AUC**		
		Overall	**Avg. adapted PAI**	**Improvement**
Taxi & Popular venues	All crimes	1,320.49	1.33	0.00%
	Burglary	1,232.26	1.24	-6.68%
	Robbery	**1,378.38**	1.39	**4.38%**
	Grand larceny	1,318.86	1.33	-0.12%
	Grand larceny of motor vehicle	1,251.42	1.26	-5.23%
	Felony assault	1,259.67	1.33	-4.61%
Crime & Popular venues-type	All crimes	1,283.25	1.29	0.00%
	Burglary	1,208.46	1.21	-5.83%
	Robbery	**1,316.73**	1.32	**2.61%**
	Grand larceny	1,284.22	1.29	0.08%
	Grand larceny of motor vehicle	1,210.07	1.22	-5.70%
	Felony assault	1,231.87	1.24	-4.00%

Table 8: Performance comparison for best scenarios and all types of crime.

The highest AUC value is achieved by robbery within the taxi combined with popular venues scenario, corresponding to an average adapted PAI of 1.39. This is followed by grand larceny within the same scenario (1.33 average adapted PAI) and by robbery in crime combined with popular venues-type (1.32 average adapted PAI). In terms of improvement over the baseline, for the scenario combining taxi with popular venues, robbery shows the highest improvement (4.38%), followed by grand larceny (-0.12%). Both slightly under-perform compared to all types of crimes combined. Likewise, for the scenario that combines crime areas with popular venues-type, robbery shows the highest improvement (2.61%), followed by grand larceny (0.08%), which slightly over-performs when compared to the baseline. The results for both scenarios are highly consistent. Thus, both strategy combinations perform best for robbery.

Again, we analyze the efficiency of the best performing crime types within each scenario, covering 80% and 90% of total crime locations within the simulation. The results of this are shown in Table 9. For 80% coverage of crime locations, the adapted PAI values vary between 1.45 and 1.63, and for 90% coverage, the adapted PAI values vary between 1.28 and 1.41. By comparing the adapted PAI values for all crime types combined (see previous section), simulations run to account only for robbery and grand larceny revealed themselves to be more efficient in terms of adapted PAI. The highest adapted PAI value was achieved by robbery within the scenario combining *crime with popular venues-type* for 80% and for 90% coverage, with respective adapted PAI values of 1.63 and 1.41 for 100 and 300 agents within the simulation. Both scenarios perform slightly better for robbery than for grand larceny. This strongly indicates the usefulness of simulating specific scenarios for street crimes rather than for other types of criminal behaviors.

Scenario	**Crime type**	**80 % crime locations coverage**		**90 % crime locations coverage**	
		n	**Adapted PAI**	**n**	**Adapted PAI**
Taxi & Popular venues	Robbery	**100**	**1.55**	**225**	**1.38**
	Grand larceny	125	1.45	325	1.28
Crime & Popular venues-type	Robbery	**100**	**1.63**	**300**	**1.41**
	Grand larceny	150	1.45	500	1.28

Table 9: Efficiency and coverage of crime locations per type of crime for the best simulated scenarios.

6.3 Best Scenario Performance at CT Level

In this section, we present the results of our investigation of the spatial distribution of crime location coverage at CT level. We look at the two best performing scenarios, that of crime area combined with popular venues-type destination strategy and taxi area combined with popular venues destination strategy, both for robberies only. We then compare the real number of robberies in each CT from the original crime dataset to the robberies covered by the agents within the aforementioned simulated scenarios and assess whether there is a pattern of CTs in which the scenario under-performs.

For this part of the experiment we mapped the robberies (at street segment level) onto CTs, resulting in 1,303 robberies spread over 781 CTs, with a maximum of 9 robberies in a CT (see Figure 15). In contrast, our simulated scenario using crime areas covered 1,178 of those robberies, leaving 125 (9.59%) robberies in 53 (6.79%) CTs untraveled (see Figure 16). The number of untraveled robberies per CT varies between 0 and 3. A visual comparison of Figures 15 and 16 reveals little difference between actual robberies and robberies traveled within the simulated scenario. Our simulated scenario using the taxi area strategy covered 1,175 robberies, leaving 128 (9.82%) robberies in 52 (6.66%) CTs untraveled (see Figure 17). For this scenario, the maximum number of undiscovered robberies in a CT is also 3. In our opinion, the differences between real and traveled robberies do not seem to be clustered in specific regions of the city, even though not all robbery locations are traveled by the agents within each simulated scenario. This suggests a good performance balance across the simulated scenario strategies in space.

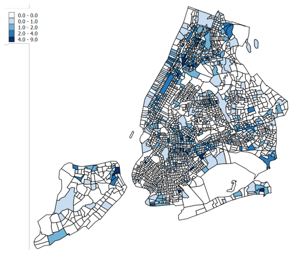

Figure 15: Robbery locations per CT from the original dataset, for 1 month (June 2015).

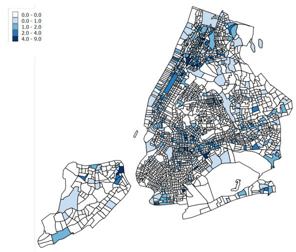

Figure 16: Traveled robbery locations for crime area strategy combined with popular venues-type per CT.

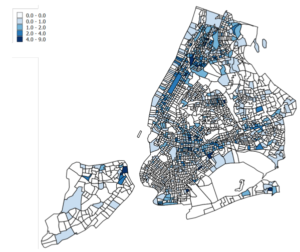

Figure 17: Traveled robbery locations for taxi area strategy combined with popular venues per CT.

7. Conclusion

The goal of the simulation was to formulate rules of offender movement behavior which cumulatively match the spatial distribution of historic crime locations. In order to achieve this, and taking into account that literature in criminology suggests that criminals are prone to offend between frequently visited activity nodes (i.e., their awareness space), we proposed and tested 35 offender mobility scenarios with various mobility strategies. Our model instantiated structural and large-scale mobility data for NYC: (1) the NYC street network with an abstract notion of residential areas and NYC population density; (2) a set of NYC crime locations (June 2015) mapped to the street segments and CTs for model evaluation; (3) venues and check-ins from LBSN (i.e., Foursquare) as proxies for activity nodes and human activity; (4) aggregated taxi trip data mapped to CTs as a proxy for travel patterns; and (5) NYC crime location data for the previous year (June 2014 to May 2015) mapped to CTs as a proxy for attractive crime areas. Moreover, by explicitly creating a simulation experiment with behavioral heuristics driving the mobility of the agent offenders, we gave ourselves a solid quantitative, spatial basis for evaluating our work in terms of a comparison between experimental results and known, empirical data.

7.1 Discussion and Implications

To determine the most useful strategies for simulating offender mobility, we analyzed and compared the simulated scenarios for various numbers of agents in terms of adapted PAI (a measure based on calculating a ratio between historic crime locations passed and the distance traveled by the agents). In fact, our results showed that all of the scenarios produced very high adapted PAI values (i.e., high performance) when simulating a low number of agents. We understand this to be a consequence of a reduced accumulated travel distance. As an artifact-of-simulation, this was a result of the agents not needing to cover (travel by) a minimum percentage of historic crime locations within the simulated environment. Knowing this in advance was, in part, why we ran the simulations with various numbers of agents and assessed them for a minimal crime coverage percentage. We achieved improved results with more plausible numbers of simulated criminal perpetrators.

Our overall analysis based on average adapted PAI over varying numbers of agents within the simulation, revealed a consistent over-performance of destination selection strategies inferred with proxies for human activity (derived from LBSN). Indeed, using information on activity nodes, including the popularity of those nodes, proves beneficial to the simulation of offender mobility. In terms of agent area selection strategies, it appears that using static distance selection (agents always traveling between 38,000 and 42,000 feet) performs best, but only when assessing the average PAI over a varying number of agents. This result was somewhat surprising because according to the mobility literature presented in Section 3, a strategy that applied a Lévy flight trajectory selection (one that mimics individual human movement) should have produced a better result, at least compared to agents traveling static distances. Nonetheless, our work confirmed a hypothesis that finely tuned input parameters — in this case adapted explicitly to NYC and in accordance with the average trip length of the NYC population — leads to plausible output results, which are comparable to more elaborate parameters inferring data from large-scale human mobility sources.

However, a more relevant measure with regard to the overall adapted PAI performance is to consider which simulation scenarios performed best for covering a minimum percentage of historic crime locations within the simulation, i.e., 90%. The highest adapted PAI value was achieved using a proxy for attractive crime areas (from historic crime data for the previous year) combined with a human activity proxy, simulating 475 agents. The next best performance was achieved by using a travel patterns proxy (from taxi trip data) combined with a human activity proxy, simulating 325 agents. Consequently, the scenarios including rich real data (LBSN in combination with taxi trip data or historic crime data) performed best compared to various strategies using only average travel distance within NYC. This was again consistent with our hypothesis that an empirically grounded and explicit ABM using large-scale mobility data would prove a powerful complex system diagnostic tool.

We engaged in a deeper analysis of the results, focusing on exploring the two best performing scenarios in terms of simulated crime types (i.e., evaluating only agents passing specific types of crime locations). In terms of average PAI over a varying number of agents, both scenarios (proxies for travel patterns and attractive crime areas combined with a human activity proxy) performed best for robbery, followed by grand larceny (performing similarly to all crime types combined). This result still holds when assessing the scenarios for a 90% coverage of crimes, and both scenarios perform best for robbery. The highest performance was achieved by an attractive crime areas proxy combined with a human activity proxy for 300 agents, with an adapted PAI reaching a value of 1.41. The next best performance was the scenario combining a travel patterns proxy with a human activity proxy in a simulation using 225 agents, which achieved an adapted PAI value of 1.38. Hence, we conclude that those scenarios that included large-scale human activity data proved most useful for simulating offender mobility in robbery simulations. Moreover, adding taxi trip data as a proxy for travel patterns resulted in simulation outputs comparable to using historic crime data as a proxy for attractive crime areas.

Consequently, our scenarios, especially those including real data, are most useful for simulating offender mobility for specific street crimes as opposed to other crime types or all crime types combined. On the one hand, these results are in line with our previous research, which showed that accounting for human activity (e.g., Foursquare venues and check-ins) and travel patterns (taxi trip data) improved predictive accuracy, especially for models predicting robbery and grand larceny (Kadar et al., 2017; Kadar & Pletikosa, 2018). On the other hand, the range of adapted PAI values achieved for our models was within the lower but acceptable range compared to PAI values in the works of others, e.g., between 1.2 and 3.37 for burglary prediction models (Adepeju, Rosser, & Cheng, 2016). Note that the values for adapted PAI achieved in this simulation are not directly comparable to the original PAI applied in crime prediction models. In this simulation we counted historic crime locations seen by agents without accounting for crime committing capabilities. The original PAI only counts occurrences of crime.

Finally, we evaluated the spatial coverage of historic crime locations at CT level for our best performing scenario in order, to gain insight as to whether or not there were recognizable spatial patterns of crime locations not covered by the agents within the simulation. We did not recognize any spatial patterns and therefore conclude that the simulation for these scenarios was balanced and covered crime locations equally throughout the CTs.

The results presented by this paper provide extensive insights into the construction of more accurate rules governing offender mobility in crime simulations and suggests that integrating more realistic offender mobility strategies, informed with novel large-scale human mobility data, can improve such simulations.

7.2 Limitations and Future Work

Simulating criminal behavior can improve our understanding of the mechanisms underlying crime and contribute to: (1) more informed testing of crime prevention strategies and (2) more accurate crime predictions. Developing informed rules governing the spatial movement strategies of mobile agents is crucial for crime simulations. Building on our previous work (Rosés et al., 2018) and the work of many others, this paper extends the state of the art by proposing and testing numerous offender mobility scenarios.

We urge caution regarding the limitations inherent to the datasets used for inferring different types of proxies. The robustness and validity of the proxies for human activity and mobility have not been tested. We have relied on existing literature and studies to choose the datasets used to build the proxies. Thus, we cannot be sure about the accuracy with which our proxies reproduce reality. This implies that our results are preliminary. Moreover, we acknowledge the bias in the datasets (see Section 3). First, Foursquare data has geographical and social biases (e.g., user age). Second, taxi trip data is also biased towards specific areas of the city, such as Manhattan. Considering that we have aggregated data from those datasets, these issues are mitigated. We also acknowledge the inherent bias in historic crime data, as it only contains crimes reported to the police, leaving unreported crimes unaccounted for. Moreover, these biases may manifest themselves geographically if the data is skewed towards certain city areas. This could have impacted the results in a positive way if each data set is biased towards the city areas with more crime (e.g., Manhattan) or in a negative way if each dataset is biased towards different areas of the city. We encourage further research to verify the presented results using different data sources.

Further limitations include the fact that our study was only conducted for NYC and may not be valid for other cities, especially those cities with basic structural differences. Future work could compare the performance of mobility strategies across different cities. Furthermore, in order to understand the impact of improving offender mobility rules in yet more general crime simulations, our crime simulation should be extended to include agents having the capability to decide whether or not to commit a new crime (as in Peng & Kurland, 2014). This additional capability can be implemented with or without the mobility behavior described in this paper. In general, the combination of heuristic mobility strategies, as we have shown in this work, with the capability of the agents to decide whether or not to offend (commit a new crime along their travel paths) would provide further insights into the utility of crime simulation.

In terms of model evaluation, we suggest that future work might involve the use of a machine learning technique perhaps one like that of Mnih, Kavukcuoglu, Silver, Rusu, Veness, Bellemare, Graves, Riedmiller, Fidjeland, Ostrovski, Petersen, Beattie, Sadik, Antonoglou, King, Kumaran, Wierstra, Legg, and Hassabis (2015) or Mnih, Puigdomènech Badia, Mirza, Graves, Harley, Lillicrap, Silver, and Kavukcuoglu (2016). It might be possible to construct an engine similar to theirs and to use it to assess the emergence (or non-emergence) of patterns in the data, especially when assessing how the simulation covers crime patterns over various areas of the city.

In addition to highlighting the importance of offender mobility within crime simulation, this work also highlights the impact of explicit ABM techniques that incorporate: (1) environmental data into crime simulations; (2) LBSN data; (3) and taxi trip data. These can all improve crime simulations by plausibly accounting for human activity. We argue for the importance of including newly available, rich data sources to improve crime simulations, especially for increasing the transferability of simulated results to the real-world. In summary, we believe that scientific research like ours, and like the many other works we have cited in this paper, has the potential to contribute to the success of law enforcement organizations and individual police officers around the world as they test crime prevention strategies *in silico*.

Acknowledgments

This paper extends a paper presented at AAMAS'18: Rosés, R., Kadar, C., Gerritsen, C., & Rouly, C. (2018). Agent-based simulation of offender mobility: integrating activity nodes from location-based social networks. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS)* (pp. 804-812). We would like to thank the anonymous reviewers of the conference version of this paper as well as the current version for their helpful feedback.

Abbreviations

AAMAS, Autonomous Agents and Multi-Agent Systems;
ABM, Agent-Based Model/Models/Modeling;
AUC, Area Under the Curve;
CSS, Computational Social Science;
CT, Census Tract;
GPS, Global Positioning System;
LBSN, Location-Based Social Networks;
NAD, North American Datum;
NYC, New York City;
PAI, Predictive Accuracy Index;
PECS, Physical, Emotional, Cognitive, and Social;
RAT, Routine Activity Theory

References

Adepeju, M., Rosser, G., & Cheng, T. (2016). Novel evaluation metrics for sparse spatio-temporal point process hotspot predictions - a crime case study. *International Journal of Geographical Information Science*, *30*(11), 2133–2154.

Adigaa, A., Agashea, A., Arifuzzamana, S., Barretta, C. L., Beckmana, R., Bisseta, K., Chena, J., Chungbaeka, Y., Eubanka, S., Guptaa, E., Khana, M., Kuhlmana, C. J., Mortveita, H. S., Nordberga, E., Riversa, C., Stretza, P., Swarupa, S., Wilsona, A., & Xiea, D. (2015). Generating a synthetic population of the United States. Technical report 15-009, Network Dynamics and Simulation Science Laboratory.

Al Boni, M., & Gerber, M. S. (2016). Predicting crime with routine activity patterns inferred from social media. In *Proceedings of the 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*.

Axelrod, R. (1986). An evolutionary approach to norms. *The American Political Science Review*, *80*(4), 1075–1111.

Beckman, R. J., Baggerly, K. A., & McKay, M. D. (1996). Creating synthetic baseline populations. *Transportation Research Part A: Policy and Practice*, *30*(6), 415–429.

Birks, D. J., Townsley, M., & Stewart, A. (2014). Emergent regularities of interpersonal victimization: An agent-based investigation. *Journal of Research in Crime and Delinquency*, *51*(1), 119–140.

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 *Proceedings of the 16th ACM International Conference on Multimodal Interaction (ICMI)*, pp. 427–434.

Bosse, T., & Gerritsen, C. (2010). A model-based reasoning approach to prevent crime. In Kacprzyk, J., Magnani, L., Carnielli, W., & Pizzi, C. (Eds.), *Model-Based Reasoning in Science and Technology*, Vol. 314 of *Studies in computational intelligence*, pp. 159–177. Springer, Berlin, Heidelberg.

Brantingham, P. L., & Brantingham, P. J. (1993). Environment, routine, and situation: Toward a pattern theory of crime. In Clarke, R. V., & Felson, M. (Eds.), *Routine Activity and Rational Choice*, Advances in Criminological Theory, pp. 259–294. Transaction Publisher, New Brunswick.

Brantingham, P. L., & Brantingham, P. J. (1995). Criminality of place. *European Journal on Criminal Policy and Research*, *3*(3), 5–26.

Brantingham, P. J., & Tita, G. (2008). Offender mobility and crime pattern formation from first principles. In Liu, L., & Eck, J. (Eds.), *Artificial crime analysis systems*. Information Science Reference, Hershey, N.Y. and London.

Brockmann, D., Hufnagel, L., & Geisel, T. (2006). The scaling laws of human travel. *Nature*, *439*(7075), 462–465.

Burger, A., Oz, T., Crooks, A., & Kennedy, W. G. (2017). Generation of realistic megacity populations and social networks for agent-based modeling. In *Proceedings of the International Conference of The Computational Social Science Society of the Americas (CSSSA)*, pp. 1–7, New York, New York, USA. ACM Press.

Chainey, S., Tompson, L., & Uhlig, S. (2008). The utility of hotspot mapping for predicting spatial patterns of crime. *Security Journal*, *21*(1-2), 4–28.

Cheng, Z., Caverlee, J., Lee, K., & Sui, D. Z. (2011). Exploring millions of footprints in location sharing services. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM)*, pp. 81–88.

Cioffi-Revilla, C. (2010). A methodology for complex social simulations. *Journal of Artificial Societies and Social Simulation*, *13*(1).

Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review*, *44*(4), 588–608.

Cranshaw, J., Schwartz, R., Hong, J. I., & Sadeh, N. (2012). The livehoods project: Understanding collective activity patterns of a city from social media. In *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media (ICWSM)*, pp. 58–65.

Crooks, A. T., & Wise, S. (2013). Gis and agent-based models for humanitarian assistance. *Computers, Environment and Urban Systems*, *41*, 100–111.

Davies, T. P., & Bishop, S. R. (2013). Modelling patterns of burglary on street networks. *Crime Science*, *2*(1), 10.

Devia, N., & Weber, R. (2013). Generating crime data using agent-based simulation. *Computers, Environment and Urban Systems*, *42*, 26–41.

Dray, A., Mazerolle, L., Perez, P., & Ritter, A. (2008). Drug law enforcement in an agent-based model: Simulating the disruption to street-level drug markets. In Liu, L., & Eck, J. (Eds.), *Artificial crime analysis systems*. Information Science Reference, Hershey, N.Y. and London.

Frank, R., Dabbaghian, V., Reid, A., Singh, S., Cinnamon, J., & Brantingham, P. (2011). Power of criminal attractors: Modeling the pull of activity nodes. *Journal of Artificial Societies and Social Simulation*, *14*(1).

Gerber, M. S. (2014). Predicting crime using Twitter and kernel density estimation. *Decision Support Systems*, *61*, 115–125.

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, *453*(7196), 779–782.

Groff, E. R. (2007a). Simulation for theory testing and experimentation: An example using routine activity theory and street robbery. *Journal of Quantitative Criminology*, *23*(2), 75–103.

Groff, E. R. (2007b). 'situating' simulation to model human spatio-temporal interactions: An example using crime events. *Transactions in GIS*, *11*(4), 507–530.

Groff, E. R., Johnson, S. D., & Thornton, A. (2018). State of the art in agent-based modeling of urban crime: An overview. *Journal of Quantitative Criminology*, *76*(1), 21.

Groff, E. R., & La Vigne, N. G. (2002). Forecasting the future of predictive crime mapping. *Crime Prevention Studies*, *13*, 29–57.

Gunderson, L., & Brown, D. (2000). Using a multi-agent model to predict both physical and cyber criminal activity. In *Proceedings of the 2000 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pp. 2338–2343, Piscataway, N.J.

Hecht, B., & Stephens, M. (2015). A tale of cities: Urban biases in volunteered geographic information. In *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media (ICWSM)*, pp. 197–205.

Herrmann, C. R. (2013). Street-level spatiotemporal crime analysis: Examples from Bronx County, NY (2006–2010). In Leitner, M. (Ed.), *Crime Modeling and Mapping Using Geospatial Technologies*, pp. 73–104. Springer Netherlands, Dordrecht.

Kadar, C., Iria, J., & Pletikosa Cvijikj, I. (2016). Exploring Foursquare-derived features for crime prediction in New York City. In *Proceedings of the 5th SIGKDD International Workshop on Urban Computing (UrbComp)*.

Kadar, C., & Pletikosa, I. (2018). Mining large-scale human mobility data for long-term crime prediction. *EPJ Data Science*, *7*(26), 1–27.

Kadar, C., Rosés Brüngger, R., & Pletikosa Cvijikj, I. (2017). Measuring ambient population from location-based social networks to describe urban crime. In Ciampaglia, G. L., Mashhadi, A., & Yasseri, T. (Eds.), *Social Informatics*, Lecture Notes in Computer Science, pp. 521–535. Springer International Publishing, Cham.

Kim, Y.-A. (2016). 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?. *Journal of Quantitative Criminology*, *27*(3), 247.

Kinney, J. B., Brantingham, P. L., Wuschke, K., Kirk, M. G., & Brantingham, P. J. (2008). Crime attractors, generators and detractors: Land use and urban crime opportunities. *Built Environment*, *34*(1), 62–74.

Kohler, T. A., Kresl, J., van West, C., Carr, E., & Wilshusen, R. H. (2000). Be there then: a modeling approach to settlement determinants and spatial efficiency among late ancestral pueblo populations of the Mesa Verde region, U.S. southwest. In Kohler, T. A., & Gumerman, G. J. (Eds.), *Dynamics in human and primate societies*, Santa

Fe Institute studies on the sciences of complexity, pp. 145–178. Oxford University Press, New York.

Liu, L., & Eck, J. (Eds.). (2008). *Artificial crime analysis systems: Using computer simulations and geographic information systems*. Information Science Reference, Hershey, N.Y. and London.

Liu, X., Gong, L., Gong, Y., & Liu, Y. (2015). Revealing travel patterns and city structure with taxi trip data. *Journal of Transport Geography*, *43*, 78–90.

Liu, Y., Kang, C., Gao, S., Xiao, Y., & Tian, Y. (2012a). Understanding intra-urban trip patterns from taxi trajectory data. *Journal of Geographical Systems*, *14*(4), 463–483.

Liu, Y., Wang, F., Xiao, Y., & Gao, S. (2012b). Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai. *Landscape and Urban Planning*, *106*(1), 73–87.

Malleson, N., Heppenstall, A., & See, L. (2010). Crime reduction through simulation: An agent-based model of burglary. *Computers, Environment and Urban Systems*, *34*(3), 236–250.

Malleson, N., See, L., Evans, A., & Heppenstall, A. (2014). Optimising an agent-based model to explore the behaviour of simulated burglars. In Dabbaghian, V., & Mago, V. K. (Eds.), *Theories and Simulations of Complex Social Systems*, Vol. 52 of *Intelligent Systems Reference Library*, pp. 179–204. Springer, Berlin, Heidelberg.

Malleson, N., & Andresen, M. A. (2014). The impact of using social media data in crime rate calculations: Shifting hot spots and changing spatial patterns. *Cartography and Geographic Information Science*, *42*(2), 112–121.

Malleson, N., & Andresen, M. A. (2015). Spatio-temporal crime hotspots and the ambient population. *Crime Science*, *4*(1), 258.

Mandelbrot, B. B. (1982). *The Fractal Geometry of Nature*. W. H. Freeman, New York.

Masad, D., & Kazil, J. (2015). Mesa: An agent-based modeling framework. In *Proceedings of the 14th Python in Science Conference (SciPy)*.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529–533.

Mnih, V., Puigdomènech Badia, A., Mirza, M., Graves, A., Harley, T., Lillicrap, T. P., Silver, D., & Kavukcuoglu, K. (2016). Asynchronous methods for deep reinforcement learning. In *Proceedings of The 33rd International Conference on Machine Learning, PMLR*, Vol. 48, pp. 1928–1937.

New York State Department of Transportation (2012). A Transportation Profile Of New York State. Technical report.

New York State Department of Transportation (2018). New York City Mobility Report. Technical report.

Nguyen, T., & Szymanski, B. K. (2012). Using location-based social networks to validate human mobility and relationships models. In *Proceedings of the 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 1215–1221.

Noulas, A., Mascolo, C., & Frias-Martinez, E. (2013). Exploiting Foursquare and cellular data to infer user activity in urban environments. In *Proceedings of the 14th IEEE International Conference on Mobile Data Management (MDM)*, pp. 167–176. IEEE.

Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., & Mascolo, C. (2012). A tale of many cities: universal patterns in human urban mobility. *PloS one*, *7*(5).

Noulas, A., Scellato, S., Mascolo, C., & Pontil, M. (2011). An empirical study of geographic user activity patterns in Foursquare. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM)*.

Peng, C., & Kurland, J. (2014). The agent-based spatial simulation to the burglary in Beijing. In *Proceedings of the 14th International Conference on Computational Science and its Applications (ICCSA)*, Vol. 8582 of *Lecture Notes in Computer Science*, pp. 31–43, Cham. Springer International Publishing.

Reid, A. A., Frank, R., Iwanski, N., Dabbaghian, V., & Brantingham, P. (2013). Uncovering the spatial patterning of crimes: A criminal movement model (crimm). *Journal of Research in Crime and Delinquency*, *51*(2), 230–255.

Rosés, R., Kadar, C., Gerritsen, C., & Rouly, C. (2018). Agent-based simulation of offender mobility: Integrating activity nodes from location-based social networks. In *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*.

Rosser, G., Davies, T., Bowers, K. J., Johnson, S. D., & Cheng, T. (2016). Predictive crime mapping: Arbitrary grids or street networks?. *Journal of Quantitative Criminology*.

Rouly, O. C. (2018). A computer simulation to investigate the association between gene-based gifting and pair-bonding in early hominins. *Journal of Human Evolution*, *116C*, 43–56.

Schelling, T. C. (1969). Models of segregation. *The American Economic Review*, *59*(2), 488–493.

Song, C., Qu, Z., Blumm, N., & Barabasi, A.-L. (2010). Limits of predictability in human mobility. *Science*, *327*(5968), 1018–1021.

Swarup, S. (2019). Adequacy: What makes a simulation good enough?. In *Proceedings of the Spring Simulation Conference (SpringSim)*, pp. 1–12.

Tang, J., Liu, F., Wang, Y., & Wang, H. (2015). Uncovering urban human mobility from large scale taxi GPS data. *Physica A: Statistical Mechanics and its Applications*, *438*, 140–153.

Tappan, P. W. (1947). Who is the criminal?. *American Sociological Review*, *12*(1), 96–102.

Urban, C., & Schmidt, B. (2001). PECS - Agent-based modeling of human behavior. Technical report FS-01-02, AAI.

Wang, G., Schoenebeck, S. Y., Zheng, H., & Zhao, B. Y. (2016a). "Will check-in for badges": Understanding bias and misbehavior on location-based social networks. In *Proceedings of the 10th International AAAI Conference on Web and Social Media (ICWSM)*.

Wang, H., Kifer, D., Graif, C., & Li, Z. (2016b). Crime rate inference with big data. In *The 22nd ACM SIGKDD International Conference*, pp. 635–644.

Ward, J. A., Evans, A. J., & Malleson, N. (2016). Dynamic calibration of agent-based models using data assimilation. *Royal Society open science*, *3*(4), 150703.

Weisburd, D. A., Bushway, S., Lum, C., & Yang, S.-M. (2004). Trajectories of crime at places: A longitudinal study of street segments in the city of Seattle. *Criminology*, *42*(2), 283–322.

White, J. W., Rassweiler, A., Samhouri, J. F., Stier, A. C., & White, C. (2014). Ecologists should not use statistical significance tests to interpret simulation model results. *Oikos*, *123*(4), 385–388.

Yang, D., Heaney, T., Tonon, A., Wang, L., & Cudré-Mauroux, P. (2017). CrimeTelescope: Crime hotspot prediction based on urban and social media data fusion. *World Wide Web*, *4*(3), 114.

Yue, Y., Zhuang, Y., Li, Q., & Mao, Q. (2009). Mining time-dependent attractive areas and movement patterns from taxi trajectory data. In *Proceedings of the 17th International Conference on Geoinformatics (CPGIS)*, pp. 1–6.

Zhang, Z., Zhou, L., Zhao, X., Wang, G., Su, Y., Metzger, M., Zheng, H., & Zhao, B. Y. (2013). On the validity of geosocial mobility traces. In *Proceedings of the 12th ACM Workshop on Hot Topics in Networks (HotNets)*, Vol. 11, pp. 1–7.