Journal of Artificial Intelligence Research x (2019) x-xx Submitted 01/19; published x/xx

Agent-Based Simulation of Offender Mobility: Modeling Activity Nodes from Large-Scale Human Activity Data

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

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

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

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 prevention strategies, and crime prediction purposes, despite facing challenges due to the complex dynamics of crime and the lack of detailed information about offender mobility. This paper presents an agent-based model to explore offender mobility, focusing on the interplay between the agent's awareness space and activity nodes. 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 journey data as a proxy for human movement between regions of the city. 35 mobility strategies have been tested, combining search distance strategies (e.g. Lévy flight, inspired by insights from human dynamics literature) and destination selection strategies (enriched with Foursquare and taxi data). We analyze and compare the different mobility strategies, and show the benefits of using large-scale human activity data to simulate offender mobility. Our strategy provides a basis for comparing offender mobility in crime simulations by inferring 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 intrusive in everyday life.

One of the main research interests within criminology is understanding when crime will occur. The most influential theory that addresses 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 where crime will occur (before it does). 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? In the current paper we address this question by using an explicit agent-based model and by generating a synthetic population of offender agents to 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 in which the offender frequently resides. The awareness space of an offender can be determined, for example by his home, work space, recreation areas, etc., including the routes towards them. So, what occurs to us is 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 those patterns in situ. Due to the complex spatially and temporally distributed nature of these processes, an often-used approach is to employ the simulation technology of Agent-Based Modeling (ABM). Indeed, previous authors and researchers have attempted to simulate crime patterns using ABM. Unfortunately, many of these simulations were based on 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). Often these simulations contained uninformed offender mobility strategies. As an alternative, this paper proposes an ABM technology that describes offender mobility based on more complete large-scale human mobility data.

As a case study, our model is applied to the surface road network of New York City (NYC), where a number of offender agent mobility strategies are compared to each other. Parting 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 human activity proxies valid for the general population. For example we've inferred home addresses from census data (land use information and population density), venues 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 journey data as a proxy for travel patterns, and historic crime location data as a proxy for attractive areas of the city. 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 criminal behavior 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 are 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. In general, simulating crime patterns contributes to the understanding of crime in a spatial environment. First generation crime simulations have mainly been 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 road-network and land-use data in combination with robberies 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 road 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 if crime patterns can be reproduced (Peng & Kurland, 2014). Then too, some have looked at road networks and household information (census and building data) in combination with burglary data to gauge the utility of ABM to prediction 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 above detailed 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, leading to 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.

Given all of the above we ask: Is it possible that by explicitly modeling the movement of offenders, their direction choices, and distances traveled, 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 arrive at a better predictive result? We therefore argue that more realistic and generalizable offender (spatial-temporal) mobility would improve crime simulations.

Thus, in this paper, we consider the importance of studying the basic simulation rules governing offender mobility by building a simulation model to compare a large number of offender-agent mobility strategies using large-scale mobility data for NYC and 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, Frank, Iwanski, Dabbaghian, & Brantingham, 2014), 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. On one hand, some of the models rely on police records for recorded home addresses as starting points to derive their trajectories (Malleson, See, Evans, & Heppenstall, 2014). Such a setup is constrained to simulating reported offenders and especially the ones for whom home addresses have been reported by the police. 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-generated user data (Gonzalez, Hidalgo, & Barabasi, 2008; Song, Qu, Blumm, & Barabasi, 2010), as well as by means of LBSN e.g. Foursquare (Noulas, Scellato, Lambiotte, Pontil, & Mascolo, 2012), and even taxi 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, which 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). Not only can information about mobility patterns and rules governing such movement be gained from LBSN, but information from social media about the location attractiveness can be used as a proxy to model the pull of specific locations within urban areas (Resch, Summa, Zeile, & Strube, 2016), and taxi data can give insights into frequent travel volume from one region of a city into others (Liu, Wang, Xiao, & Gao, 2012). 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).

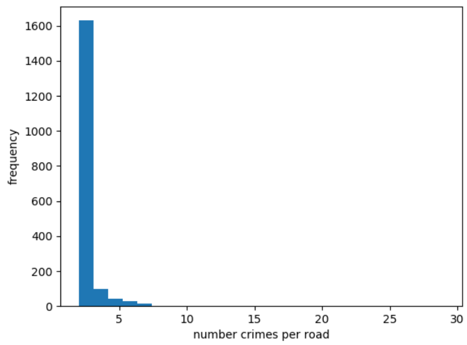

Figure 1: Number of crime locations per road in 1 month.

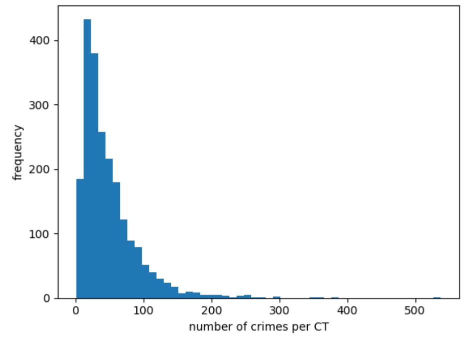

Figure 2: Number of crime locations per CT in 12 months.

4. Data

The simulation described in the next section includes sufficient geographic data to simulate a virtual environment projected onto the NYC area using projected coordinates (per North American Datum of 1983 – NAD83) and allowing measurement in feet. In particular, the simulation builds a road network for NYC including 117,320 road segments collected from the NYC open data portal. The network provides the structure of the road and public transportation system (including ferry lines), upon which the agents may find their way. Additionally, we projected data to census tracts (CT), a statistical unit subdividing counties, defined by the United States Census Bureau[1],for the New York Region. In NYC there are 2,168 CTs with a population of 3,000 to 4,000 and an average of 90 acres of land area, our dataset contains 2,162 CTs excluding CTs containing only water and shorelines. From NYC census data, we have extracted population density information for each CT and have combined it with zoning information on NYC buildings to identify residential areas. Furthermore, crime data has been obtained from the NYC open data portal, and includes anonymized felony crimes at road segment level (projected to the middle or the ends of the segment), which we projected to the road network of the simulation and to the CT for different purposes. Figure 1 shows the counts of crime locations per road in the NYC road network for the month of June 2016 (used for model performance assessment only), with 17 roads having more than 10 crime locations mapped (i.e. the x axis reaches 30 crime locations per road), and Figure 2 shows the counts of crime locations per CT in NYC for June 2015 to May 2016, with 4 CT counting more than 300 crimes in them. Note that crime locations on roads traversing several CTs are counted twice, once for each CT.

The crime data includes information such as type of crime, date, time, and location. Our final dataset contains the following types of crime: burglary, grand larceny, grand larceny of motor vehicle, robbery, and felony assault. Rape and murder incidents have not been used for simulation purposes due to low frequency (1,209 and 357 incidents in 12 months, respectively). Crime data for 12 months (June 2014 to May 2015) has been instantiated in the model on the CT level for one of the simulated scenario variations, which resulted

1. `http://www.census.gov/`

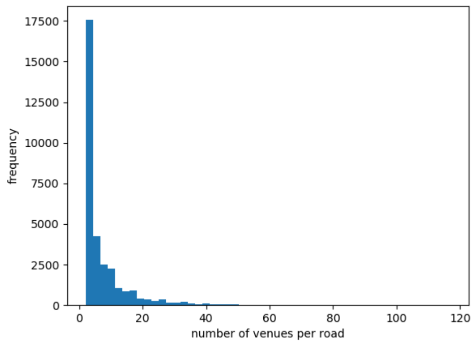

Figure 3: Number of venues per road.

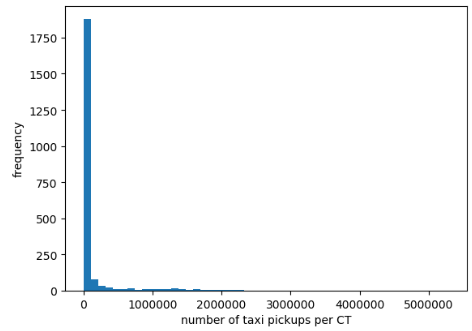

Figure 4: Number of taxi pickups per CT in 12 months.

in 104,532 crime locations mapped: 15,897 burglaries, 43,301 grand larcenies, 7,523 grand larcenies of motor vehicles, 16,413 robberies, and 19,832 felony assaults. And crime data for 1 month (June 2015) has been used to obtain an up-to-date overview of crime patterns for model evaluation on road level, that resulted in 8,503 crime locations mapped: 1,287 burglaries, 3,555 grand larcenies, 580 grand larcenies of motor vehicles, 1,303 robberies, and 1,778 felony assaults.

To instantiate attractive locations, Foursquare data was collected from the Foursquare API (Application Programming Interface)[2], as in Kadar, Iria, and Pletikosa Cvijikj (2016), Kadar, Rosés Brüngger, and Pletikosa Cvijikj (2017), including information about venues in the area of NYC: venue name, location, check-in counts (accumulated over time), associated categories, etc. The set is composed of 273,149 venues in the proximity of every incident from the crime data set with over 122 million check-ins (from creation of the venue in the platform until data collection in June 2016) associated with venue categories ranging from arts and entertainment, college and university, events, food, nightlife, shops and services, etc. The venues have been mapped to the roads of the NYC road network. Figure 3 shows the distribution of venue counts per road. 54 roads contain over 60 venues with a maximum of 120 venues per road (i.e. outliers in Figure 3). In the simulation model, Foursquare venues are used as proxies for activity nodes and the check-ins are used to quantify the attractiveness of the activity nodes. Serving as a proxy for human dynamics, taxi data, including information about travel starting and ending points, are projected on the CTs and give insights into the connectivity and popularity of transition between CTs. This latter dataset is composed of over 248 million taxi trips within 12 months (July 2014–June 2015) obtained from the official website NYC [3]. See Figure 4 for taxi journey pickup frequency per CT.

2. `http://www.foursquare.com/`
3. `http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml`

5. The Simulation

Inspired by previous ABMs simulating crime, in this paper we study offender mobility by assessing the performance of different agent mobility strategies in several scenarios emulating a large-scale urban environment. In the simulation, offender agents travel the road network of an urban area, which includes geo-located information about historic crime locations. These agents move from one spatial destination to another and memorize the historic crime locations as they pass throughout the simulation (i.e. as a proxy for measuring mobility performance). The agents represent criminal offenders and travel from a starting location to a number of activity nodes before returning to the starting location. The goal of the simulation is that agents pass as many new crime locations as possible along the path. Each simulation step (epoch) represents 1 day of the month and the model runs for 30 days, consistent with one-month crime data. The performance of all agents is evaluated after the total period of 30 steps.

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 environment, including a road network, spatial destinations representing activity nodes, and geo-spatial reported 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 the next section.

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

5.1 Basic Functionality Formalization

The variables in Section 5.1 are used in the following section to introduce the model features in detail.

The simulation model instantiates agents traveling from a starting position s to a destination position x, before returning to position s at the end of the epoch (step). Agents are created and newly positioned at each step, whereas starting and ending at a location s can be assumed. Over one model run (30 steps), the agents collect information about the historic crime locations c they pass by, including details about the type of crime. As the agents embody anonymous offenders, s is inferred from residential areas weighted by the population density of each area. Agents are placed on the closest road within 80 feet from a residential building. The residential building is chosen by weighting each building according to the population density of the CT where it is situated.

Within the same step, the agents search for a destination x in area a to travel to, while the value of a and the possibilities of x depend on the offenders' strategies. Strategies for choosing a and x are combined into different simulation scenarios, detailed in section 5.2. The number of trips *xtrip* an agent performs between several x, within the same epoch, before returning to s, is drawn from U(0,2×atrip), where *atrip* is the statistical average number of trips performed by the NYC population (3.8 trips per day) (New York State Department

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

Table 1: Variables in the simulation model.

of Transportation, 2012), thus the number of x each agent visits per step varies. The model is run for the different scenarios and their performance is assessed using the results of tc (crime locations traveled) and td (distances traveled) by the agent. The performance is measured using several metrics: (1) comparing an adaption of the Predictive Accuracy Index (PAI) (Chainey, Tompson, & Uhlig, 2008) over the scenarios. 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. For assessing the performance of this model, PAI has been adapted as follows:

$$adapted\ PAI = \frac{\frac{\Sigma\ tc}{\Sigma\ c}}{\frac{\Sigma\ td}{\Sigma\ d}} \tag{1}$$

The adapted PAI shows the relationship between percentage of distinct crime locations passed by the agents and the percentage of distinct distances traveled (i.e. length of new roads within the road network). The higher the resulting index, the better the performance of the model, i.e. traveling between more crime locations per distance. The index is computed counting each different crime locations passed by any agent only once.

Besides assessing the scenario's performance at the road level, (2) we also assess the performance of the most successful scenarios on the CT-level, comparing the coverage area (crime locations traveled) over different CTs within one scenario, giving us information on whether the agents cover crime locations equally across various CTs of the city.

Additionally, the optimal number of agents is determined by comparing the performance of the simulations 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 in White, Rassweiler, Samhouri, Stier, and White (2014), where they advised against it for social simulations.

5.2 Mobility Scenario Strategies

Thirty-five agent mobility scenarios were built by varying agent mobility strategies and by applying the knowledge described in section 3. In particular, the Lévy flight distribution was built to mimic realistic distance choices for each agent's movements, alternatively aggregated taxi journey data was included as a proxy for travel patterns, and aggregated historic crime data was included as a proxy for the general attractiveness of areas with previous crime, while venues from location-based social networks (including information on the popularity of each venue) were derived as a proxy for activity nodes and human activity, accounting for the attractiveness of specific destinations within the general population. The scenarios combined 5 options for area-selection strategies and 7 options for destination-selection strategies, resulting in thirty-five strategy combinations as described in the following.

Area selection:

1. **Static**: The static distance allows agents to move only in one specific distance, set to a radius of 40,000 feet with a 5% boundary, resulting in area a, the average trip length for NYC's population (New York State Department of Transportation, 2012).

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, resulting in area a.

3. **Power**: The Lévy flight distance draws distances from a power law distribution using Lévy flight. The Lévy flight formula is transformed to allow drawing distances 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 area a:
$$P(r) \sim r^{-(1+\beta)} \rightarrow r \sim \frac{1}{P(r)} \times \mathrm{e}^{\frac{1}{1+\beta}} \tag{2}$$

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 census tract at the starting position s and any other census tract in NYC. Census tracts with higher transition frequencies are weighted higher.

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 s are weighted higher.

Destination selection within area *a*:

1. **Random roads**: The first option is the most basic one, offering any random road as a destination.

2. **Random venues**: The second destination choice is any random activity node (Foursquare venue).

3. **Random venues-center**: The third destination choice accounts for the attractiveness of the city center, allowing a choice of any activity node, and weighting roads in the direction 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 a choice of activity nodes weighted by their popularity, determined using check-in counts from Foursquare as in equation 3. The higher the number of check-ins, the higher the weight of the venue.

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

5. **Popular venues**: The fifth strategy offers a choice of activity nodes weighted by their popularity (determined using check-in counts from Foursquare) and by their proximity to the center of the city, using the center score (described in item 3 of this section).

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

6. **Popular venues-center**: The sixth strategy offers a choice of random activity nodes weighted by the popularity of the venue category, determined by total check-in count per venue category in all of NYC. The higher the overall number of check-ins for a category, the higher the weight of the venues within this category.

$$P[x] = \frac{\Sigma \textit{check-ins category}}{\Sigma\ \textit{check-ins total}} \tag{5}$$

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

$$P[x] = \frac{\textit{check-ins}}{\Sigma\ \textit{check-ins within } r} \times \frac{\Sigma \textit{check-ins category}}{\Sigma\ \textit{check-ins total}} \tag{6}$$

6. Simulation Results

For the purpose of assessing the performance of various offender mobility strategies described in the previous section, we ran multiple simulations across several different scenarios. 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 on the CT level.

6.1 Scenario performance for all types of crimes

In the first step, we explored the performance of the scenarios on road level for all types of crimes. Each of the thirty-five scenarios was 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 Figure 5 for destination strategies combined with static area strategy. See Figure 6 for destination strategies combined with uniformly distributed area strategy. See Figure 7 for destination strategies combined with power-law distributed area strategy. See Figure 8 for destination strategies combined with taxi area strategy. And see Figure 9 for destination strategies combined with crime area strategy.

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

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 previously seen graphs (see Table 2). 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, and showed 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 overall performing scenarios. Conversely, power-law, uniform random, and static area selection strategies, each combined with random roads, performed worst with average adapted PAI values between 1.15 and 1.18.

Defining the most basic strategy (static area combined with random roads destination) as the baseline, we compare the relative AUC improvement of each scenario while grouping destination selection strategies by area selection strategies. In Table 2, we conclude that when static area selection strategy is combined with a popular venues-center strategy it performs best, gaining a 14.31% improvement over the baseline. Then, followed by crime area strategy combined with popular venues-type, a 12.63% improvement is seen. Using a power-law strategy combined with popular venues-center results in only 9.75% improvement. Further, when empirical taxi data is combined with a simple popular venues strategy, a 9.46% improvement is seen. Finally, a uniform random area strategy combined with popular venues-center strategy delivers only a 9.03% improvement. However, a pattern emerges as the special cases for popular venues perform best within each of the area strategies. Additionally, the difference in performance for popular venues, popular venues-center and popular venues-type is very smal, 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 crimes spots covered within each simulated scenario (see Figure 10). We defined efficiency to mean achieving the highest adapted PAI value while covering a reasonable amount of the crime locations within a simulation needing the least number of agents (percentage of crime locations traveled). For a crime locations coverage of 80% and 90%, we determined the adapted PAI value and number of agents (see Table 3). 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

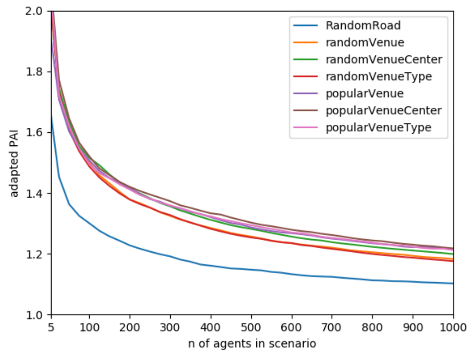

Figure 5: Adapted PAI (all crime types) for static distance and number of agents.

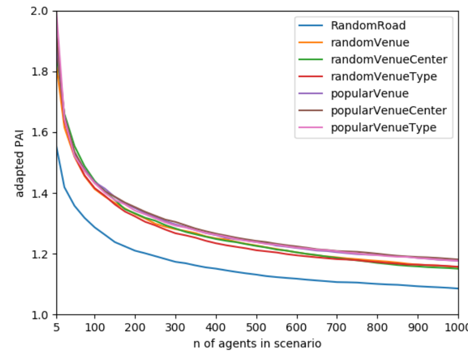

Figure 6: Adapted PAI (all crime types) for uniform distance.

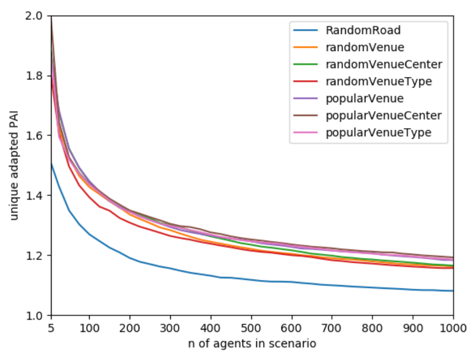

Figure 7: Adapted PAI (all crime types) for power distance.

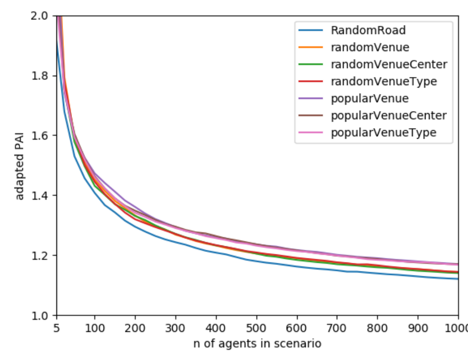

Figure 8: Adapted PAI (all crime types) for taxi area.

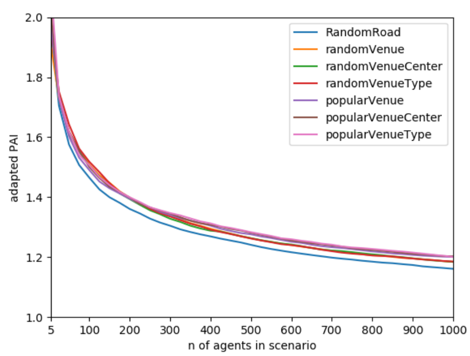

Figure 9: Adapted PAI (all crime types) for crime area.

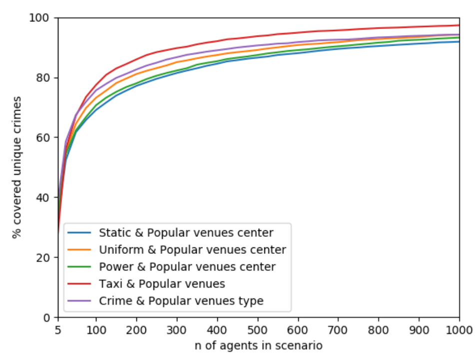

Figure 10: Crime locations coverage for the 5 best performing scenarios.

Distance	Destination	AUC		
		Overall	Avg. adapted PAI	Improvement
Static	Random roads	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 roads	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 roads	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 roads	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 roads	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 2: Overall performance comparison for all scenarios, improvement over static combined with random roads.

Distance	Destination	80 % crime locations coverage		90 % crime locations coverage area	
		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 3: Efficiency and coverage of crime locations over the simulation.

option, with only 125 agents within the simulated scenario. In turn, to cover 90% of 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 for each scenario that depends on our desired crime locations coverage 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 number of agents for different types of crime: burglary, robbery, grand larceny, larceny of motor vehicle, and felony assault. See Figure 11 for taxi combined with popular venues and Figure 12 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 similar to all types of crimes combined. Both scenarios under-perform for the remaining crime types of burglary, grand larceny of motor vehicle, and felony assault compared to all types of crimes aggregated. We note that both over-performing crime types can be grouped into a larger crime category referred to as street crimes. Consistent with the analysis in the previous subsection, we show in Table 4 that the application of a holistic measure for assessing the overall performance of the different crime type within the scenarios has value. 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), resulting in two baselines, one for each scenario.

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-performing when compared to the baseline. The results over both scenarios are highly consistent.

Again, we analyze the efficiency of the best performing crime types within each scenario, covering 80% and 90% of total crime locations within the simulated scenarios. The results of this are shown in Table 4. 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 ran 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

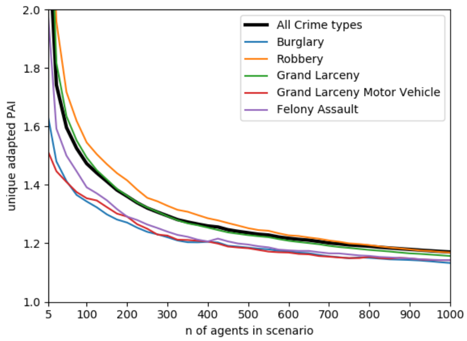

Figure 11: Adapted PAI for different types of crimes in taxi & Popular venues.

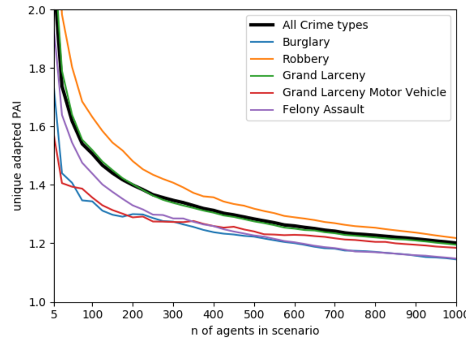

Figure 12: Adapted PAI for different types of crimes in Crime & Popular venues-type.

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

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 area**	
		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 5: Efficiency and coverage of crime locations per type of crime for the best simulated scenarios.

6.3 Best scenario performance on CT

In this section, we present the results of our investigation of the spatial distribution of crime locations coverage on the 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. And then, we compare the real number of robberies in each CT from the original crime dataset to the robberies covered by the agents within the mentioned 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 the road level onto CTs, resulting in 1,303 robberies spread over 781 CTs, with a maximum of 9 robberies in a CT (see Figure 13). 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 14). The number of untraveled robberies per CT varies between 0 and 3. A visual comparison of the Figures 13 and 14 reveals little difference between actual robberies and robberies traveled within the simulated scenario. Our simulated scenario using a taxi area strategy covered 1,175 robberies, leaving 128 (9.82%) robberies in 52 (6.66%) CTs untraveled (see Figure 15). 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.

7. Conclusion

The goal of the simulation was to find strategies governing offender mobility from starting positions to daily activities as inspired by RAT. 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 thirty-five offender mobility scenarios with specific criminal movement patterns. Our model instantiated structural and large-scale mobility data for NYC: (1) the NYC road network with an abstract notion of residential areas and NYC population density; (2) a set of NYC crime locations (June 2015) mapped to the roads and CT for model evaluation; (3) venues and check-ins from LBSN (i.e. Foursquare) as proxies for activity nodes and human activity;(4) aggregated taxi journey data mapped to CTs as a proxy for travel patterns; and (5) NYC crime location data for the previous year (June 2014–May 2015) mapped to CT 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

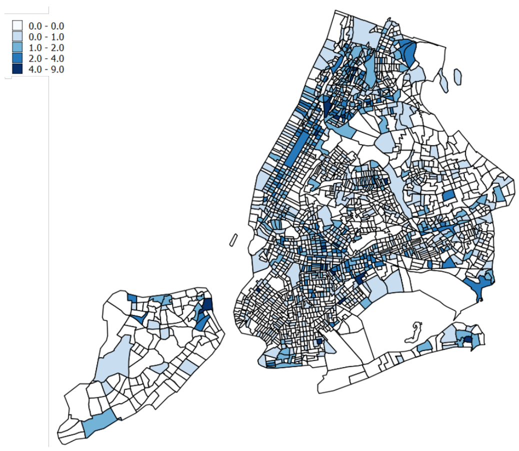

Figure 13: Robbery locations per CT from the original data set, for 1 month (June 2015).

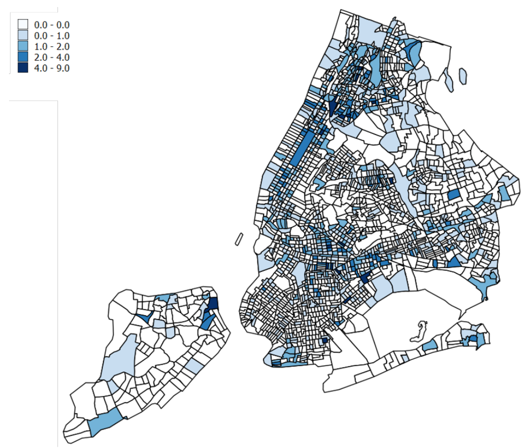

Figure 14: Traveled robbery locations for crime area strategy combined with popular venues-type per CT.

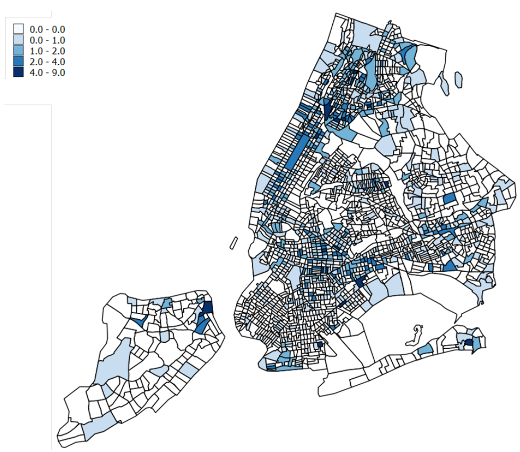

Figure 15: Traveled robbery locations for taxi area strategy combined with popular venues per CT.

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, brings a benefit to the simulation of offender mobility. In terms of agent area selection strategies, it appears 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 elaborated parameters inferring data from large-scale human mobility sources.

However, a more relevant measure to observer 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 highest performance was achieved by using a travel patterns proxy (from taxi journey data) combined with a human activity proxy, simulating 325 agents. Consequently, the scenarios including rich real data (LBSN in combination with taxi data or historic crime data) performed best compared to various strategies only using average travel distance within NYC. This was again consistent with our hypothesis that an empirically-grounded and explicit ABM using large-scale mobility data was 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 a specific type of crime locations). In terms of average PAI over 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 similar 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 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 the ones including real data, are most useful for simulating offender mobility for specific street crimes opposed to other crime types, or all crime types combined. On 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 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 rate 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 crime occurrences.

Finally, we evaluated the spatial coverage of historic crime locations at the CT level for our best performing scenarios, 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 conclude 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 (Brantingham & Tita, 2008; Rosés Brüngger, Bader, Kadar, & Pletikosa Cvijikj, 2017; Rosés, Kadar, Gerritsen, & Rouly, 2018) and the work of many others, this paper extends the state of the art by proposing and testing numerous offender mobility scenarios.

We caution that our study was only conducted for NYC and may not be valid for other cities, especially those cities with basic structural differences. Additional suggestions for future work could easily compare the performance of mobility strategies across different cities. Moreover, to understand the impact of improving offender mobility rules in yet more general crime simulations, our crime simulation could 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. Overall, we think quantitatively valid mobility results, as we have shown in this work, combined with the use of qualitative behavioral heuristics that drive agent mobility and crime committing decision strategies, would add plausibility. In general, the combination of heuristic mobility strategies along with the capability of the agents to decide whether to offend (commit a new crime) or not along their travel paths would provide further insights into the utility of crime simulation.

Further limitations encompass the bias in the data sets we have used for inferring different types of proxies. First, Foursquare data has geographical and social biases (user age). Second, Taxi journey data is also biased towards specific areas of the city, such as Manhattan and the airports. Considering that we have aggregated data from those datasets, these issues are mitigated. We also acknowledge the inherent bias in crime locations data, as it only contains crimes reported to the police, therefore leaving unreported crimes unaccounted for.

In terms of model evaluation, we suggest that future work might involve a Machine Learning technique using an Artificial Recurrent Neural Network to assess the emergence (or non-emergence) of patterns in the data, especially when assessing how the simulation covers crimes 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 include: (1) environmental data into crime simulations; (2) LBSN data; (3) and taxi journey 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 scientific research like ours, and like the many other works we have cited in this paper, have 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 Autonomous Agents and Multi-Agent Systems (AAMAS) 2018 (Rosés et al., 2018).

Abbreviations

AAMAS, Autonomous Agents and Multi-Agent Systems; *ABM*, Agent-Based Model/Models/Modeling; *AUC*, Area Under the Curve; *CSS*, Computational social science; *CSS*, Computational social science; *CT*, Census Tract; *GPS*, Global Positioning System; *LBSN*, Location-Based Social Networks; *NAD*, North American Datum; *NYC*, New York City; *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.

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

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.

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, 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.

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.

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.

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

David Masad, J. K. (Ed.). (2015). *Mesa: An Agent-Based Modeling Framework*.

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).

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.

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

Kadar, C., Iria, J., & Pletikosa Cvijikj, I. (2016). Exploring foursquare-derived features for crime prediction in new york city..

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.

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 in 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, Y., Wang, F., Xiao, Y., & Gao, S. (2012). Urban land uses and traffic 'source-sink areas': Evidence from gps-enabled taxi data in shanghai. *Landscape and Urban Planning*, *106*, 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.

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

New York State Department of Transportation (2012). A transportation profile of New York state..

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), e37027.

Peng, C., & Kurland, J. (2014). The agent-based spatial simulation to the burglary in beijing. In Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J. M., Kobsa, A., Mattern, F., Mitchell, J. C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Terzopoulos, D., Tygar, D., Weikum, G., Murgante, B., Misra, S., Rocha, A. M. A. C., Torre, C., Rocha, J. G., Falcão, M. I., Taniar, D., Apduhan, B. O., & Gervasi, O. (Eds.), *Computational Science and Its Applications – ICCSA 2014*, Vol. 8582 of *Lecture Notes in Computer Science*, pp. 31–43. Springer International Publishing, Cham.

Reid, A. A., Frank, R., Iwanski, N., Dabbaghian, V., & Brantingham, P. (2014). Uncovering the spatial patterning of crimes: A criminal movement model (crimm). *Journal of Research in Crime and Delinquency*, *51*(2), 230–255.

Resch, B., Summa, A., Zeile, P., & Strube, M. (2016). Citizen-centric urban planning through extracting emotion information from twitter in an interdisciplinary space-time-linguistics algorithm. *Urban Planning*, *1*(2), 114.

Rosés, R., Kadar, C., Gerritsen, C., & Rouly, C. (2018). Agent-based simulation of offender mobility: Integrating activity nodes from location-based social networks. In Dastani, M., Sukthankar, G., André, E., & Koenig, S. (Eds.), *17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018)*.

Rosés Brüngger, R., Bader, R., Kadar, C., & Pletikosa Cvijikj, I. (2017). Towards simulating criminal offender movement based on insights from human dynamics and location-based social networks. In Ciampaglia, G. L., Mashhadi, A., & Yasseri, T. (Eds.), *Social Informatics*, Vol. 2 of *Lecture Notes in Computer Science*, pp. 458–465. Springer International Publishing, Cham.

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 (New York, N.Y.)*, *327*(5968), 1018–1021.

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. *In Emotional and Intelligent- The Tangled Knot of Social Cognition, AAAI Fall Symposoim Series*.

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.

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.