Agent-Based Simulation of Offender Mobility: Integrating Activity Nodes from Location-Based Social Networks

Raquel Rosés
ETH Zurich
Zurich, Switzerland
rroses@ethz.ch

Cristina Kadar
ETH Zurich
Zurich, Switzerland
ckadar@ethz.ch

Charlotte Gerritsen
NSCR
Amsterdam, Netherlands
cgerritsen@nscr.nl

Chris Rouly
ETH Zurich
Zurich, Switzerland
chris.rouly@gess.ethz.ch

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 and location-based social networks data to represent activity nodes as proxy for human activity. 18 mobility strategies have been tested, combining search distance strategies (e.g. Lévy flight, inspired by insights in human dynamics literature) and destination selection strategies (enriched with Foursquare data). We analyze and compare the different mobility strategies, and show the impact of using activity nodes extracted from social networks to simulate offender mobility. This agent-based model provides a basis for comparing offender mobility in crime simulations by inferring offender mobility in urban areas from real world data.

CCS CONCEPTS

• **Applied computing → Law, social and behavioral sciences**;

KEYWORDS

Agent-based simulation; crime; LBSN; human mobility patterns

ACM Reference Format:
Raquel Rosés, Cristina Kadar, Charlotte Gerritsen, and Chris Rouly. 2018. Agent-Based Simulation of Offender Mobility: Integrating Activity Nodes from Location-Based Social Networks. In *Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), Stockholm, Sweden, July 10–15, 2018,* IFAAMAS, 9 pages.

1 INTRODUCTION

Criminology is a multidisciplinary research field that aims to explain, predict and prevent criminal behavior. Although criminals represent only a minority of the overall population, people come in contact with criminal behavior (either by being criminal or by being a victim), anytime anyplace. Crime is 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) [12]. 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 answer the question 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 [7]. 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 generate a synthetic population of offender agents navigating the urban environment.

Previous studies [8] have shown us that higher crime concentration rates are found within the 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 appears to us is that the area lying 'between' frequently visited activity nodes should be the field of operation of 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 Agent-Based Modelling (ABM). Indeed, previous authors and researchers have attempted to simulate crime patterns using ABM. Unfortunately, many of these models were often 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 [23], and combine uninformed offender mobility strategies. As an alternative, this paper proposes an ABM technology that describes offender mobility based on more complete data, inferred from structural data (e.g. road network), census data, and activity nodes from Location-Based Social Networks (LBSN). Note that we focus on mobility patterns and not on the agent' decision whether to offend. As a case study, our model

Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), M. Dastani, G. Sukthankar, E. André, S. Koenig (eds.), July 10–15, 2018, Stockholm, Sweden.

is applied to the surface transportation network of New York City, 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 [33], the strategies developed here are not only inspired by theories in criminology but use human activity proxies gained from location-based social networks, we infer home addresses from census data and activity nodes from location-based social networks. The performance of the model is assessed in terms of number of crimes covered and distance traveled by the agents. We note that this model could be applied to study social behavior other than crime 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. The paper concludes with a discussion in Section 7.

2 RELATED WORK

Criminology, the study of crime, involves many aspects, the interrelationship of which may be mathematically complex. In the context of related work, we believe that Computational Social Science has begun to present itself as an important explanatory tool for analyzing and predicting crime. We also note that the technologies of Computational Social Science have emerged as tools with the potential to offer explanatory insight across many other complex social issues [11]. But, we think this is particularly true for Criminology. Across several fields and several decades, the technologies of Computational Social Science have consistently demonstrated interdisciplinary explanatory power through the use of agent-based simulation (consider [3, 13, 34, 35]). Especially in the field of Criminology, scientists are discovering the power of agent-based simulation for various applications involving theory testing [4, 6, 18, 23], testing of prevention strategies [5, 14, 15, 20] and forecasting the development of crime [20, 24, 28], [23] provides an overview for basic characteristics of crime simulation models). In general, simulating crime patterns contributes to the understanding of crime in a spatial environment. Crime simulations have mainly been built in virtual (simulated) environments without the use of real world data (e.g. [6]) to study the underlying mechanisms of crime. However, including environmental 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 street robberies to test RAT with basic offender agent moving between a set of static and predefined activity nodes, and deciding whether to offend [19]. Others have considered street and subway network in combination with burglary data and agents moving between connected nodes at random and/or with heavy-tailed distribution waiting times (inspired by research in human mobility patterns), to test if crime patterns can be reproduced [28]. Then too, some have looked at street networks, household information (census and building data) in combination with burglary data to test the use of ABM for crime prediction with agents modeled in a complex manner using PECS (Physical conditions, Emotional states, Cognitive capabilities and Social status) [36]. These latter simulations consider frameworks that model offender behavior and move between randomly assigned home and work locations to build a cognitive map of passible targets within their awareness space [24, 37]. 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 crime by including agent individual characteristics, e.g. wealth measure and target characteristics, and guardianship level of the possible targets, leading to the offender's decision whether to offend or not. In contrast, the offender agent mobility characteristic are rather neglected and based on simplified assumptions, with the exception of [6] who have studied simple offender movement strategies leading to emergent crime patterns in a 2-D space by means of basic mathematical and ABM models. Given all of the foregoing we ask: Is it possible that by explicitly modeling movement offender direction choices, distances traveled, and by comparing random walks to Lévy flight probability distribution might we discover that simple mobility rules could be used together with other behavior rules to reproduce crime patterns that arrive at a better predictive result? In this sense, we argue that more realistic and generalizable offender (spatial-temporal) mobility would improve crime simulations. Lévy flight, being 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 [26].

Thus, in this paper, we consider the importance of studying the basic simulation rules governing offender mobility, and build a simulation model to compare a large number of offender agent mobility strategies on New York City's (NYC) transportation network with historic crime data of the city.

3 CRIMINAL OFFENDER MOBILITY

In RAT, *routine activities* are described as everyday activities which 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 the areas connecting the different activity nodes [29],i.e. awareness space. Consequently, including offender agents' home location 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 [25], while such a setup is limited to simulating reported offenders and especially the ones for which home addresses have been reported by the police. On the other hand, little effort has been put into defining appropriate activity nodes and reproduced realistic human (e.g. offender) spatial-temporal mobility patterns in simulations. In the era of social media and crowdsourced/location-based user data [13], patterns in human activity can be inferred from openly available data. Human mobility patterns have been intensively studied by means of GPS generated user data [17, 32] as well as by means of LBSN, e.g. Foursquare [27]. Such research has confirmed the high regularity of individual human movement

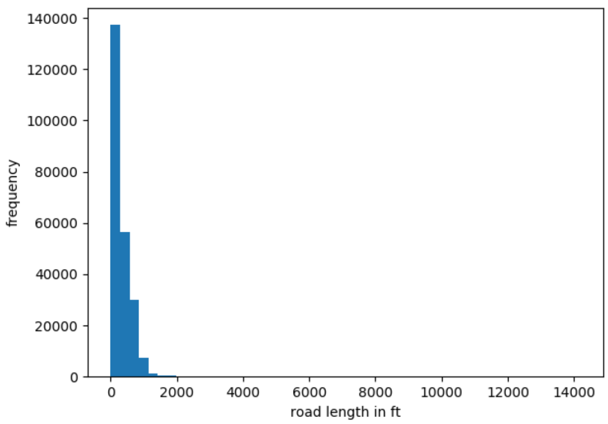

Figure 1: Road length histogram.

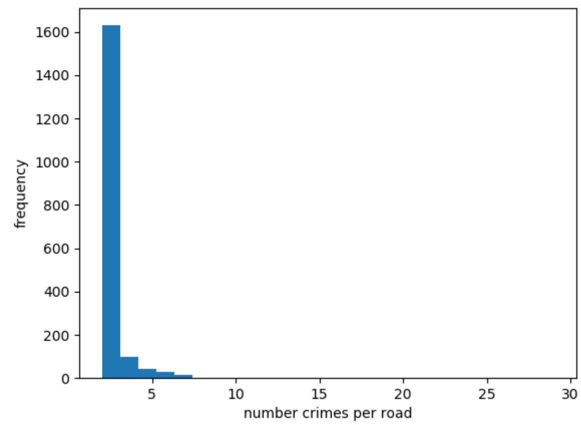

Figure 2: Number of crimes per road histogram.

and determined basic rules governing it, e.g. suggesting individual human travel distances should be modeled by means of Lévy flight. 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 certain locations within urban areas [30]. Moreover, activity nodes and city centers as a special case, have been identified to attract offenders as well as the general population [16].

4 DATA

The simulation described in the next section, includes sufficient geographic data to simulate a virtual environment projected onto New York City area using rectified coordinates (per NAD83) and allowing measurement in feet. In particular, the simulation builds a road network for NYC including 117,321 street segments collected from 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. The length of the roads approximates a Poisson distribution (see Figure 1), with a number of outliers: 440 roads with a length of over 2000 feet (ft), hence the x axis of the graph reaches 14000 ft. From NYC census data , we have extracted population density information for each census tract and have combined it with zoning information on NYC buildings to identify residential areas. Furthermore, crime data has also been obtained from the NYC open data portal, and includes anonymized felony crimes at street segment level (projected to the middle or the ends of the segment), which we projected to the road network of the simulation. Figure 2 shows the counts of crime per road in the NYC road network, with 17 roads having more than 10 crimes mapped (i.e. the x axis reaches 30 crimes per road). The crime data includes information such as type of crime, date, time, etc., and the following types of crime: burglary, grand larceny, grand larceny of motor vehicle, robbery, and felony assault. Rape and murder incidents in this dataset have not been used for simulation purpose due to low frequency (114 and 23 incidents, respectively).

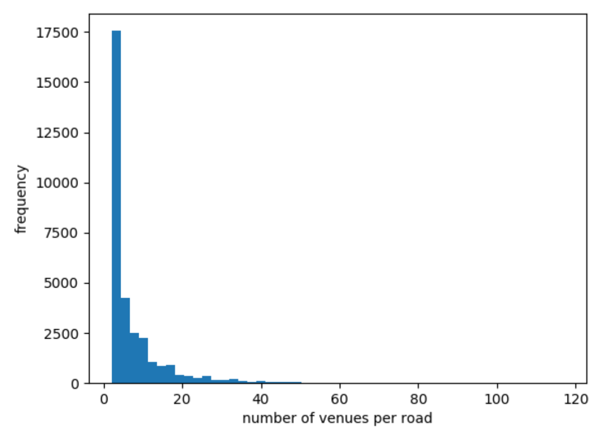

Figure 3: Number of venues per road histogram.

Crime data for 1 month (June 2015, arbitrary decision) has been instantiated in the model, to obtain an up-to-date overview of crime patterns for a time period that resulted in 8,494 crime incidents mapped: 1,287 burglaries, 3,555 grand larcenies, 580 grand larcenies of motor vehicle, 1,301 robberies, and 1,778 felony assaults. To instantiate attractive locations, Foursquare data was collected from the Foursquare API (as in [21, 22]), 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 to the venues and categories ranging from Arts and Entertainment, College and University, Events, Food, Nightlife Spot to Shop and Service, 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 (attractive locations) and the check-ins are used as proxies for attractiveness of the activity nodes.

5 SIMULATION MODEL

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 crimes, moving from one spatial destination to another, while memorizing the historic crimes 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 crimes 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 more detail, 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 agent. These points are formalized in the next section.

Using Mesa an agent-based modeling framework in Python, a simplified version of New York City is instantiated in this model, providing the structure of the road and public transportation network (including ferry lines), zoning features for residential areas and population density in these areas, as well as venues from location-based social networks including popularity of each venue as proxies for activity nodes (from Foursquare), and crime locations per type of crime (burglary, robbery, grand larceny, larceny of motor vehicle, and felony assault) for one month (June in 2015).

5.1 Basic Functionality Formalization

The variables in Table 1 are used in the following section to introduce the model feature 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). For each step, n agents are instantiated, starting and ending at a location s can be assumed. Agents are created and newly positioned at each step. Over one model run (30 steps) the agents collect information about the historic crimes c they pass by, including details about the type of crime. As the agents embody unknown offenders, s is inferred from residential areas weighted by 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 census tract where it is situated. Within the same step, the agents search for a destination x in distance r (10 %) to travel to, while the value of r and the possibilities of x depend on the offenders'

Table 1: Variables in simulation model.

Level	Name	Explanation
`Model`	n	number of agents instantiated
	rd	road in NYC road network
	c	historic crime
	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 crimes
	td	traveled distance
	r	search distance
	$xtrip$	number trips in a day

strategy. Strategies for choosing r 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 NYC population (3.8 trips per day) [1], 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 (crimes traveled) and tc (distances traveled)by the agent. In detail, the performance is measured using several metrics: (1) comparing the coverage area, ratio of distinct (i.e. new) tc by all agents within the simulation (without multiple counts), over c (total histori crimes), for all crimes and per crime type; (2) comparing an adaption of the Predictive Accuracy Index (PAI) [10] 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, the PAI index has been adapted as follows:

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

Adapted PAI shows the relationship between percentage of distinct crimes 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. The index is computed counting each different crime passed by any agent only once. Additionally, the optimal number of agents is determined by comparing the performance of the simulations with different number of agents: 25, 50, 75, 100, 125 and 150. Note, that no significance test will be conducted for comparing the performance of different scenarios following the recommendations in [38], advising against it for social simulations.

5.2 Mobility Scenario Strategies

Fifteen agent mobility scenarios are built by varying agent mobility strategies and by applying the knowledge exposed in section

3. In particular, the Lévy flight distribution is built to mimic realistic distance choices for the agent's movement, while venues from location-based social networks (including information on popularity of each venue) are derived as a proxy for activity nodes accounting for their attractiveness for the general population. The scenarios combine 3 options for search distance selection and 5 options for destination selection. **Distance selection**:

(1) The static distance allows agents to move only in one specific distance, set to r=40,000 feet, the average trip length for NYC's population [1].
(2) The uniformly distributed distance builds upon the static distance, uniformly drawing distances from a distribution with average trip length for NYC's population: $R \sim U(0, 2r)$ so that E[R]= r.
(3) 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 the boundaries of NYC, with β =0.6, determined to be the optimal value for NYC [9]:

$$P(r) \sim r^{-(1+\beta)} \to r \sim \frac{1}{P(r)} \times e^{\frac{1}{1+\beta}} \tag{2}$$

Destination selection within selected distance:

(1) The first option is the most basic one, offering any random road as a destination within distance *r*.
(2) The second destination choice is any random activity node (Foursquare venue) within the distance *r*.
(3) The third destination choice accounts for the attractiveness of the city center, allowing a choice of any road within distance *r* but 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) The fourth option offers a choice of activity nodes weighted by their popularity, determined using check-in counts from Foursquare as follows:

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

(5) The fifth strategy offers a choice of activity nodes weighted by their popularity and by their proximity to the center of the city, using the center score.

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

6 SIMULATION RESULTS

The simulation model described in the previous section was run with the above presented scenarios to assess the performance of offender agent mobility strategies. In this section, we present the most interesting results over all simulated scenarios.

6.1 Performance for all types of crimes

In the first step we explored performance of the scenarios for all types of crimes, see Figure 4 for destination strategies combined with static distance, Figure 5 for destination strategies combined with uniformly distributed distances, and Figure 6 for destination strategies combined with Lévy flight distance. The figures show the

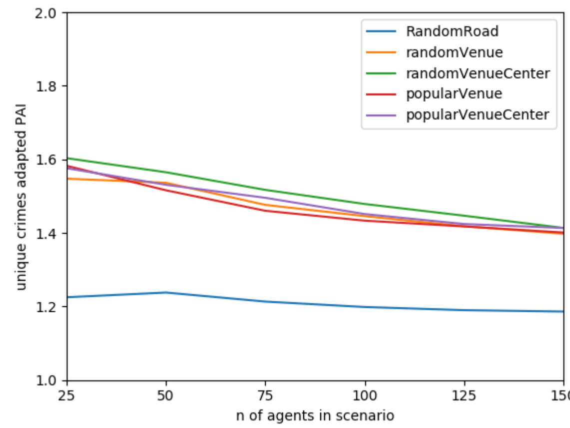

Figure 4: Adapted PAI (all crime types) for static search distance and number of agents.

performance of the adapted PAI measurement per simulation for different number of agents in the scenarios *n* (25, 50, 75, 100, 125, 150). Across the three figures, the most basic destination strategy offering a choice between random roads, underperforms compared to more elaborate strategies. The remaining destination strategies can be seen as special cases of activity nodes, and as such perform similarly. Random venues weighted to the center of the city perform slightly better combined with static distance, popular venues weighted to the center of the city perform slightly better combined with uniformly distributed distances. In Figure 6, popular venues weighted to the center of the city, popular venues and random venues weighted to the center of the city perform best for different agent count in the simulation, while popular venues prevail more often. Thus the variance of destination strategies is slightly higher for scenarios with Lévy flight distances compared to other scenarios. A closer look at the best performing strategies (see Figure 7) shows the highest performing destination strategy for each search distance, from which we conclude, that the static search distance performs slightly better than the uniformly distributed and the Lévy flight distance, although the difference in performance is relatively small. Throughout the simulation scenarios the performance decays while the number of agents per simulation is increased, due to the fact that an increasing number of agents also increases the traveled distance, which normalizes the performance index. Thus, we obtain a notion about the impact of increasing the number of agents in each scenario, but do not gain information about the optimal number of agents for this simulations. For more details, in Figure 8 we show the percentage of crimes covered by all agents in the best performing scenarios. The scenarios behave consistent with the previous figures and cover between 41.04% - 74.09% of distinct crimes. The growth of the trend line for the different scenarios starts to stabilize around 100 agents and static distance combined with random venues weighted to the center of the city obtains the highest percentages of covered crimes. This meaning that the number of new crimes traveled by the agents within the scenario starts growing slower from 100 agents on. Thus, we assume the optimal

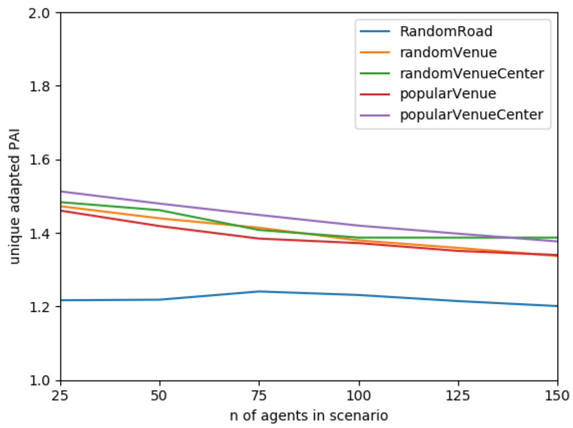

Figure 5: Adapted PAI (all crime types) for uniformly distributed search distance.

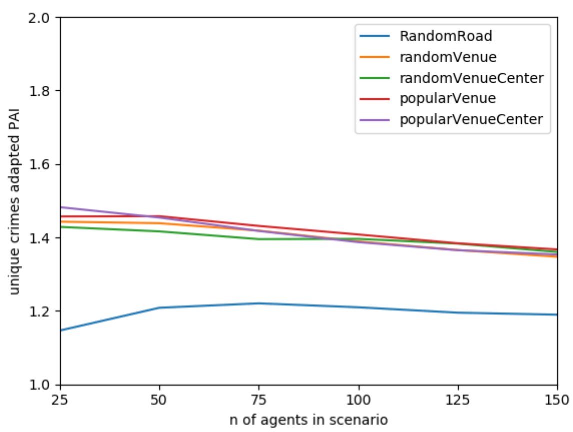

Figure 6: Adapted PAI (all crime types) for Lévy flight distance.

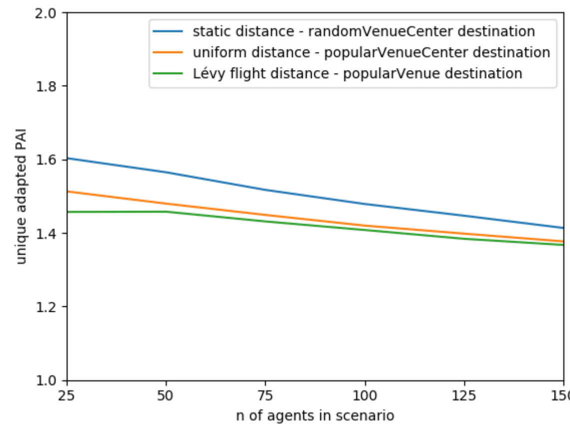

Figure 7: Adapted PAI (all crime types) for best performing search distance strategies.

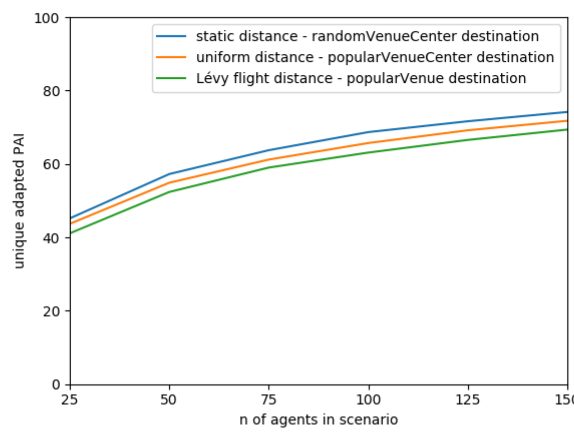

Figure 8: Percentage of crimes covered for best performing scenarios.

number of agents for these scenarios to be approximated by 100 agents. Furthermore, we look into the roads traveled by 100 agents for the best performing strategy and notice a nearly even spread of the agents over the road network, with higher transit across main roads. The Map in Figure 9 shows the roads traveled by 100 agents after 30 epochs, with gradual size increase for traveled frequency.

6.2 Performance for single types of crimes

We now proceed to a deeper analysis of the overall best performing scenario by looking into the performance for different crime types. In this example scenario, grand larcenies and robberies achieve the highest adapted PAI values. These crimes can be grouped into street crimes, which perform even higher than all crime types combined. In contrast, grand larceny of motor vehicle, burglary and felony assault underperform compared to all crime types combined (see Figure 10). In terms of percentage covered by the different crime types (see Table 2), the highest rate of crimes passed by at least one agent is 77.97% and 76.88%, for robbery and grand larceny with 150 agents respectively. In contrast, the lowest rate is achieved by felony assault with 33.24% for 25 agents.

7 DISCUSSION

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. While informed rules governing spatial movement strategies of mobile agents is crucial for crime simulations. Building on previous work [6, 31], this paper extends the state of the art by proposing and testing fifteen offender mobility scenarios combining search distance and destination selection strategies. The goal

Table 2: Percentage of crimes passed for crime types.

Number of agents	All crimes	Burglary	Robbery	Larceny	Larceny Motor	Assault
25	45.11%	39.78%	47.89%	49.23%	41.03%	37.06%
50	57.19%	51.05%	60.17%	61.10%	51.55%	49.38%
75	63.68%	57.50%	67.23%	66.92%	58.28%	55.96%
100	68.61%	62.55%	72.22%	71.87%	63.79%	60.29%
125	71.56%	65.73%	75.44%	74.57%	66.90%	63.16%
150	74.09%	68.07%	77.97%	76.88%	68.62%	66.25%

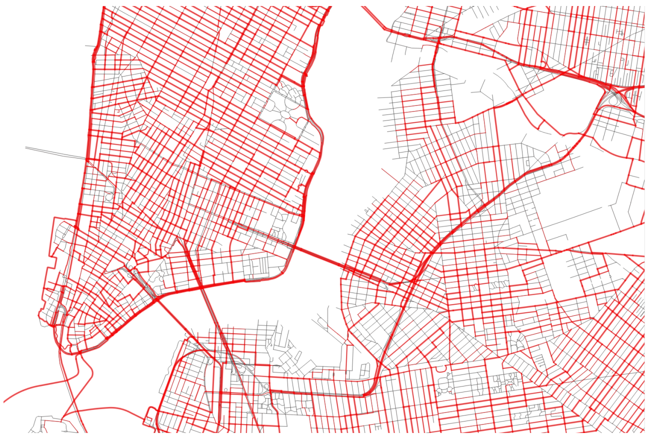

Figure 9: Map of NYC for best performing scenario with traveled roads marked in red.

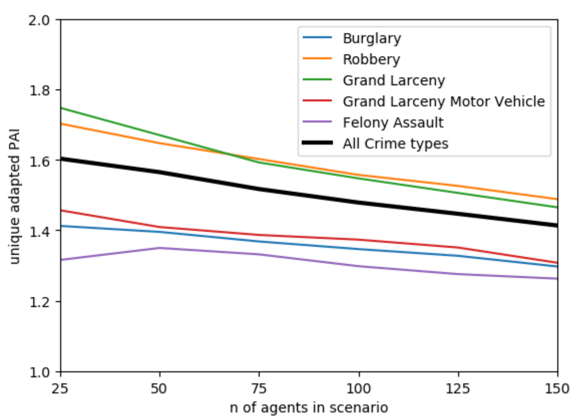

Figure 10: Adapted PAI for static distance and random venues weighted to the center of the city.

of the simulation was to find strategies governing offender mobility from starting positions to daily activities (inspired by RAT) creating the offenders' awareness space (area between the traveled points). According to literature in Criminology, offenders are known to be prone to offend within their awareness space. For a more realistic representation of the environment, the model instantiates the NYC road network, population density and residential areas within the road network, crimes mapped to the roads and venues from Foursquare (including user check-ins per venue) as proxies for activity nodes. To measure the performance of the different scenarios we count historic crimes traveled by each offender within the scenario (each crime only counted once) and build a performance measure based on PAI (i.e. referred to as adapted PAI), accounting for the percentage of historic crimes passed and the distance traveled by the agents. The first results compare the performance of search distances strategies, suggesting, that the static search distance performs slightly better than the uniformly distributed and Lévy flight search distance. The similarity between the search distance strategies confirms, the correct fine tuning of the parameters for the uniform and Lévy search distance, which have been adapted to NYC and are in accordance with the average trip length of NYC population (i.e. the static distance in the scenarios). Next, the results for comparing the destination selection strategies show that random roads as destination for offender mobility highly underperform compared to destinations representing activity nodes. The different activity nodes (random venues, random venues weighted to the city center, popular venues, and popular venues weighted to the city center) can be seen as special cases accounting differently for activity nodes (i.e. venues) and show little variations in performance when compared to each other. We conclude, that the improvement in performance comes from including activity nodes in the simulation and the variations make little difference, indeed the overall best performing strategy combines static search distance with random venue destinations weighted to the center of the city. Furthermore, the range of adapted PAI values is within the lower but acceptable rate compared to PAI index values in other work, (e.g. between 1.2 and 3.37 for burglary prediction models) [2]. 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 have counted crimes seen by agents without accounting for a crime committing action, the original PAI counts crime occurrences.

The optimal number of agents for this type of simulation can be approximated by looking into the percentage of distinct crimes covered by the scenarios with different number of agents within the scenario. For the best performing scenario, we achieve 68.61% of distinct crimes covered with 100 agents, which we argue approximates the optimal number of agents for this type of simulations. We note that further simulations with higher number of agents could give further insights into the optimal number of agents for these simulations, while at this point the running times where prohibitive (10 hours for a scenario with 150 agents). Furthermore, we look into the performance of the best scenario for different type of crimes, and show it performs best for street crimes (grand larceny and robbery), better than for all types of crimes combined. In contrast, the performance for grand larceny for motor vehicles, burglary and assault underperform compared to all crime types combined. These results are in line with research completed using Foursquare venues to improve crime prediction models, showing that accounting for human activity (e.g. Foursquare venues and check-ins) improves predictive accuracy, this is so, especially for models predicting grand larceny and robberies [22]. Thus, we confirm the consistency of our results obtained using activity nodes.

The results achieved thorough this paper provide preliminary insights into constructing more accurate rules governed offender mobility for crime simulations. The integration of more realistic offender mobility strategies informed with novel environmental data can improve crime simulations, while this study was only conducted for NYC and may not be generalizable to other cities, especially not for those with basic structural differences. Future work, could compare the performance of mobility strategies across different cities and include a larger variety of activity nodes (e.g. human activity derived from taxi flow data), as well as more complex destination choice mechanisms (e.g. choice of different destination types at different times).

To understand the impact of improving offender mobility rules on crime simulations, a crime simulation including agent decision making whether to offend, such as [28] could be implemented with and without the mobility behavior described in this paper. The combination of mobility strategies with decisions whether to offend along the travel path would provide further insights into the utility of mobility strategies.

In addition to the importance of offender mobility within crime simulation, this work highlights the impact of including environmental data into crime simulations, and explores how LBSN data can improve crime simulations by accounting for human activity. We argue about the importance of including newly available rich data sources to improve crime simulations especially for increasing the transferability simulated results to the real world e.g. implications for police officers testing prevention strategies in-silico.

REFERENCES

[1] 2012. *A Transportation Profile Of New York State.* Technical Report. New York State Department of Transportation, Albany, New York.

[2] Monsuru Adepeju, Gabriel Rosser, and Tao Cheng. 2016. Novel evaluation metrics for sparse spatio-temporal point process hotspot predictions - a crime case study. *International Journal of Geographical Information Science* 30, 11 (2016), 2133–2154. https://doi.org/10.1080/13658816.2016.1159684

[3] Robert Axelrod. 1986. An Evolutionary Approach to Norms. *The American Political Science Review* 80, 4 (1986), 1075–1111.

[4] Daniel J. Birks, M. Townsley, and A. Stewart. 2014. Emergent Regularities of Interpersonal Victimization: An Agent-Based Investigation. *Journal of Research in Crime and Delinquency* 51, 1 (2014), 119–140. https://doi.org/10.1177/0022427813487353

[5] Tibor Bosse and Charlotte Gerritsen. 2010. A Model-Based Reasoning Approach to Prevent Crime. In *Model-Based Reasoning in Science and Technology*, Janusz Kacprzyk, Lorenzo Magnani, Walter Carnielli, and Claudio Pizzi (Eds.). Studies in computational intelligence, Vol. 314. Springer Berlin Heidelberg, Berlin, Heidelberg, 159–177. https://doi.org/10.1007/978-3-642-15223-8{_}8

[6] Paul J. Brantingham and George Tita. 2008. Offender Mobility and Crime Pattern Formation from First Principles. In *Artificial crime analysis systems*, Lin Liu and John Eck (Eds.). Information Science Reference, Hershey, N.Y. and London.

[7] Patricia L. Brantingham and Paul J. Brantingham. 1993. Environment, Routine, and Situation: Toward a Pattern Theory of Crime. In *Routine Activity and Rational Choice*, Ronald V. Clarke and Marcus Felson (Eds.). Transaction Publisher, New Brunswick, 259–294.

[8] Patricia L. Brantingham and Paul J. Brantingham. 1995. Criminality of place. *European Journal on Criminal Policy and Research* 3, 3 (1995), 5–26. https://doi.org/10.1007/BF02242925

[9] Dirk Brockmann, Levente Hufnagel, and Theo Geisel. 2006. The scaling laws of human travel. *Nature* 439, 7075 (2006), 462–465. https://doi.org/10.1038/nature04292

[10] Spencer Chainey, Lisa Tompson, and Sebastian Uhlig. 2008. The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime. *Security Journal* 21, 1-2 (2008), 4–28. https://doi.org/10.1057/palgrave.sj.8350066

[11] Claudio Cioffi-Revilla. 2010. A Methodology for Complex Social Simulations. *Journal of Artificial Societies and Social Simulation* 13, 1 (2010). https://doi.org/10.18564/jasss.1528

[12] Lawrence E. Cohen and Marcus Felson. 1979. Social Change and Crime Rate Trends: A Routine Activity Approach. *American Sociological Review* 44, 4 (1979), 588–608.

[13] Andrew T. Crooks and Sarah Wise. 2013. GIS and agent-based models for humanitarian assistance. *Computers, Environment and Urban Systems* 41 (2013), 100–111. https://doi.org/10.1016/j.compenvurbsys.2013.05.003

[14] Nelson Devia and Richard Weber. 2013. Generating crime data using agent-based simulation. *Computers, Environment and Urban Systems* 42 (2013), 26–41. https://doi.org/10.1016/j.compenvurbsys.2013.09.001

[15] Anne Dray, Lorraine Mazerolle, Pascal Perez, and Alison Ritter. 2008. Drug Law Enforcement in an Agent-Based Model: Simulating the Disruption to Street-Level Drug Markets. In *Artificial crime analysis systems*, Lin Liu and John Eck (Eds.). Information Science Reference, Hershey, N.Y. and London.

[16] Richard Frank, Vahid Dabbaghian, Andrew Reid, Suraj Singh, Jonathan Cinnamon, and Patricia Brantingham. 2011. Power of Criminal Attractors: Modeling the Pull of Activity Nodes. *Journal of Artificial Societies and Social Simulation* 14, 1 (2011). https://doi.org/10.18564/jasss.1734

[17] Marta C. Gonzalez, Cesar A. Hidalgo, and Albert-Laszlo Barabasi. 2008. Understanding individual human mobility patterns. *Nature* 453, 7196 (2008), 779–782. https://doi.org/10.1038/nature06958

[18] Elizabeth R. Groff. 2007. Simulation for Theory Testing and Experimentation: An Example Using Routine Activity Theory and Street Robbery. *Journal of Quantitative Criminology* 23, 2 (2007), 75–103. https://doi.org/10.1007/s10940-006-9021-z

[19] Elizabeth R. Groff. 2007. 'Situating' Simulation to Model Human Spatio-Temporal Interactions: An Example Using Crime Events. *Transactions in GIS* 11, 4 (2007), 507–530.

[20] Louise F. Gunderson and Donald Brown. 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)*. Piscataway, N.J., 2338–2343. https://doi.org/10.1109/ICSMC.2000.884340

[21] Cristina Kadar, José Iria, and Irena Pletikosa Cvijikj. 2016. Exploring Foursquare-derived features for crime prediction in New York City. (2016).

[22] Cristina Kadar, Raquel Rosés Brüngger, and Irena Pletikosa Cvijikj. 2017. Measuring Ambient Population from Location-Based Social Networks to Describe Urban Crime. In *Social Informatics*, Giovanni Luca Ciampaglia, Afra Mashhadi, and Taha Yasseri (Eds.). Springer International Publishing, Cham, 521–535.

[23] Lin Liu and John Eck (Eds.). 2008. *Artificial crime analysis systems: Using computer simulations and geographic information systems.* Information Science Reference, Hershey, N.Y. and London.

[24] Nicolas Malleson, Alison Heppenstall, and Linda See. 2010. Crime reduction through simulation: An agent-based model of burglary. *Computers, Environment and Urban Systems* 34, 3 (2010), 236–250. https://doi.org/10.1016/j.compenvurbsys.2009.10.005

[25] Nicolas Malleson, Linda See, Andrew Evans, and Alison Heppenstall. 2014. Optimising an Agent-Based Model to Explore the Behaviour of Simulated Burglars. In *Theories and Simulations of Complex Social Systems*, Vahid Dabbaghian and Vijay Kumar Mago (Eds.). Intelligent Systems Reference Library, Vol. 52. Springer, Berlin, Heidelberg, 179–204. https://doi.org/10.1007/978-3-642-39149-1{_}12

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

[27] Anastasios Noulas, Salvatore Scellato, Renaud Lambiotte, Massimiliano Pontil, and Cecilia Mascolo. 2012. A tale of many cities: universal patterns in human urban mobility. *PloS one* 7, 5 (2012), e37027. https://doi.org/10.1371/journal.pone.0037027

[28] Chen Peng and Justin Kurland. 2014. The Agent-Based Spatial Simulation to the Burglary in Beijing. In *Computational Science and Its Applications – ICCSA 2014*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Alfred Kobsa, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Demetri Terzopoulos, Doug Tygar, Gerhard Weikum, Beniamino Murgante, Sanjay Misra, Ana Maria A. C. Rocha, Carmelo Torre, Jorge Gustavo Rocha, Maria Irene Falcão, David Taniar, Bernady O. Apduhan, and Osvaldo Gervasi (Eds.). Lecture Notes in Computer Science, Vol. 8582. Springer International Publishing, Cham, 31–43. https://doi.org/10.1007/978-3-319-09147-1{_}3

[29] Andrew A. Reid, Richard Frank, Natalia Iwanski, Vahid Dabbaghian, and Patricia Brantingham. 2014. Uncovering the Spatial Patterning of Crimes: A Criminal Movement Model (CriMM). *Journal of Research in Crime and Delinquency* 51, 2 (2014), 230–255. https://doi.org/10.1177/0022427813483753

[30] Bernd Resch, Anja Summa, Peter Zeile, and Michael Strube. 2016. Citizen-Centric Urban Planning through Extracting Emotion Information from Twitter in an Interdisciplinary Space-Time-Linguistics Algorithm. *Urban Planning* 1, 2 (2016), 114. https://doi.org/10.17645/up.v1i2.617

[31] Raquel Rosés Brüngger, Robin Bader, Cristina Kadar, and Irena Pletikosa Cvijikj. 2017. Towards Simulating Criminal Offender Movement Based on Insights from Human Dynamics and Location-Based Social Networks. In *Social Informatics*, Giovanni Luca Ciampaglia, Afra Mashhadi, and Taha Yasseri (Eds.). Lecture Notes in Computer Science, Vol. 2. Springer International Publishing, Cham, 458–465. https://link.springer.com/content/pdf/10.1007%2F978-3-319-67256-4_36.pdf

[32] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-Laszlo Barabasi. 2010. Limits of predictability in human mobility. *Science (New York, N.Y.)* 327, 5968 (2010), 1018–1021. https://doi.org/10.1126/science.1177170

[33] Paul W. Tappan. 1947. Who is the Criminal? *American Sociological Review* 12, 1 (1947), 96–102.

[34] Thomas C. Schelling. 1969. Models of Segregation. *The American Economic Review* 59, 2 (1969), 488–493. http://www.jstor.org/stable/1823701

[35] Timothy A. Kohler, James Kresl, Carla van West, Eric Carr, and Richard H. Wilshusen. 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 *Dynamics in human and primate societies*, Timothy A. Kohler and George J. Gumerman (Eds.). Oxford University Press, New York, 145–178.

[36] Christoph Urban and Bernd Schmidt. 2001. PECS - Agent-Based Modeling of Human Behavior. *In Emotional and Intelligent- The Tangled Knot of Social Cognition, AAAI Fall Symposoim Series* (2001).

[37] Jonathan A. Ward, Andrew J. Evans, and Nicolas Malleson. 2016. Dynamic calibration of agent-based models using data assimilation. *Royal Society open science* 3, 4 (2016), 150703. https://doi.org/10.1098/rsos.150703

[38] J. Wilson White, Andrew Rassweiler, Jameal F. Samhouri, Adrian C. Stier, and Crow White. 2014. Ecologists should not use statistical significance tests to interpret simulation model results. *Oikos* 123, 4 (2014), 385–388. https://doi.org/10.1111/j.1600-0706.2013.01073.x