Università Commerciale "Luigi Bocconi"

BSc in Economic and Social Sciences

Disciplinary Field (code): L-33

StreetCrime:
A Generative Science Approach

Tutor: Paolo Pinotti

Thesis by

Adam Amer

ID number: 3151896

Academic Year: 2022/2023

Abstract

According to data from 2019, Milan emerged as the foremost city in Italy in terms of reported robberies and pickpocketings. Anecdotal evidence of the city being perceived as unsafe has become increasingly prevalent on various social media platforms. This situation raises questions within the realm of social sciences regarding the development of effective crime reduction strategies. One of the most effective approaches involves the ability to predict when and where crimes are likely to occur. However, creating a machine learning model for this purpose is often a formidable challenge, as criminal data tends to be incomplete, difficult to access, or ethically sensitive. In response to these challenges, this thesis adopts an alternative perspective by delving into an old yet recently revitalized methodology within the social sciences: Agent-Based Modeling (ABM). Urban crime presents indeed an ideal environment for the application of ABM due to its reliance on local interactions among agents within a specified spatial context. The thesis commences with an exploration of the foundations of generative social science and the historical evolution of ABM. It then delves into a comprehensive review of criminological theories, empirical literature, and prior ABMs, with the aim of creating a model that aligns with current research on urban crime. Subsequently, the thesis follows on an empirical exploration of real-world crime data. With the addition of environmental data, the model is specified and tested through short-run and long-run iterations. The evaluation process encompasses various dimensions, including external empirical validity, empirical validation, and statistical conclusion validity. Finally, the thesis concludes with an examination of the model's performance and outlines directions for future future research expansion.

Contents

1. Generative Social Science

Without a doubt, Joshua M. Epstein, a professor of Epidemiology at NYU, stands out as the most prominent advocate for generative social science and a pioneer in the field of agent-based modeling. He has referred to his 1999 seminal paper ”Agent-based computational models and generative social science” [1], as a *generativist manifesto* ([2], p.1). This chapter seeks to clarify the notion of applying the ”generative twist” to the domain of social science, drawing inspiration from Epstein's article.

1.1 Society as a Complex Adaptive System

The fundamental tenet of generative science revolves around the idea of society as a *complex adaptive system*. One of the first social scientist to advocate for this perspective was sociologist Walter Buckley. In his 1968 article ”Society as a complex system” [3], Buckley argues that the traditional understanding of society, as an homeostatic framework, had become outdated. Equilibrium systems, by nature, lack internal sources of disturbances and only possess the ability to *self-regulate* to external pressures, maintaining their structure within predetermined boundaries. Buckley affirms that society appears instead intrinsically dynamic, undergoing simultanous internal changes at many levels and possessing a fairly unstable structure. If, for instance, we consider urban crime, it becomes apparent that multiple factors (such as individual decision-making, economic circumstances, and environmental conditions), interact in a complex manner, with feedback loops and cascading consequences that cannot be well represented by static models. When examining an adaptive system rather than an equilibrium system, it becomes evident that the former exhibits "some degree of plasticity vis-a-vis its environment" ([3] p.394). This is characterized by an ongoing and mutual interaction between environmental events and entities involved, whereby both parties exert influence on each other and subsequently adapt their responses. When subject to disturbances, the system *self-adapts*, or changes his structure, while mantaining the network of relations that defines it. These relationship between components exhibit varied degrees of interrelation, including unilateral, multilateral, correlational, and causal connections. According to Buckley, it is precisely from this intricate web of stable links, that an unstable observable structure emerges. Society can thus be conceptualized as a "social neural network", where arriving at a specific equilibrium simply means that the network "has executed a certain computation P*" ([2], p. 16). We can refer again to the example of urban crime, specifically the identification of certain areas as hotspots. We cannot assume that they will remain constant over time, as an equilibrium model would suggest, nor can we predict their response to environmental fluctuations without definining the underlying interactions that generate these hotspots. If we are able to succeed in such endeavour, it is possible that we will not only be capable of accurately predicting the progression of hotspots, but also anticipate the effectiveness of policy interventions for addressing these hotspots. This is precisely the objective that generative scientists put forward.

1.2 The Evolution of Agent-Based Modeling

The generative scientist attempts to find an answer to the following question: which hypothesized microspecification gives rise to an observed macroscopic regularity ([2], p.21)?

This approach has a long history, tracing its origins back to Adam Smith's notion of the "Invisible Hand" ([4], p.1). In Adam Smith's work, individuals by locally maximizing their utility, contribute to the achievement of an optimal allocation of resources within the market, almost like there was an "invisible hand" administering the economy. Smith was witnessing a phenomenon known as "emergence," wherein complex global patterns arise from simple local rules. Nevertheless, in the absence of technological tools, an in-depth exploration of these systems was unfeasible. The reason is that equations associated with these systems are often non-linear, rendering an analytical solution non-viable. With the advent of computers, numerical solutions of non-linear systems became a possibility. But rather of solely utilizing the computer as a high-performance calculator, Von Neumann hypothesized that it could serve as a groundbreaking tool for scientific analysis ([4], p.4). He argued that, by using the computer, one could employ the scientific method to study emergence through generated data. The steps would entail the construction of a computational model, followed by the execution of experiments. Subsequently, one would compare the hypothesis with the obtained data, and improve the theory by an iterative repetition of these stages. Von Neumann was theorizing the generative approach, which became a reality with the digital revolution. Von Neumann's subsequent work on *cellular automata*, which can be considered forerunners to ABM, illustrated the potential of computers in generating emergent phenomena with efficacy. Soon after, researchers began employing this methodology to study human societies, as exemplified by significant contributions such as Schelling's model of segregation [5]) and Epstein & Axtell's Sugarscape [6]. The fundamental principle underlying ABM involves the examination of a universal pattern that arises from the "decentralized local interactions of heterogeneous autonomous agents" ([2], p.26). The initial stage in the ABM involves the delineation of various groups of individuals within the model, by explicitly defining their characteristics, fixed or stochastic, and defining the interactions between individuals, mostly using conditional statements. While individuals may engage in rational decision-making, they are traditionally different from the standard *homo economicus* due to their heterogeneity and access limited to local information. Given the significant dependence of agent-based models (ABMs) on stated parameters, ABM are usually iterated multiple times. In the following stage, asensitivity analysis is performed to examine the impact of various parameters on the desired outcome. As previously mentioned, not only this methodology encompasses a more accurate understanding of society, but also entails a more restricted delineation of the essence of social science. The utilization of black-box machine learning algorithms, whose internal mechanisms are unknown to researchers, has gained significant popularity in social science research for the purpose of enhancing predicting accuracy. The generativist scientist dismisses this technique as lacking scientific rigor, as it fails to offer a relational explanation of the phenomenon under investigation, despite achieving factual accuracy. Using Epstein words, "if you didn't grow it, you didn't explain it" ([2], p.51). However, caution must be exercised when describing ABM as a method that is entirely dependable. It is possible to attain the appropriate regularity by a microspecification that might be considered implausible or even impossible. Epstein presents the case of the Barlney's Fern Fractal, ([2], p.30), an iterative algorithm relying on local rules to construct a representation of Barlney's Fern. While the end result is visually convincing, the growth pattern is completely unbiological. Consequently, it is expected of the researcher to conduct comprehensive theorethical research in order to address this concern. Despite the hurdles, ABM remains a powerful methodology that deserves further investigation alongside conventional research approaches.

2. Exploring Literature on Urban Crime

The purpose of this chapter is to provide an overview of existing models and theories that try to study street crime dynamics, as this will aid for the development of the model going forward. The chapter will be divided into two section: the first one will present non-ABM approaches, while the second one will be focused on exploring papers that utilize ABM.

2.1 Environmental Explanations of Crime

Historically, criminological research has mostly concentrated on examining the underlying causes that drive individuals to engage in criminal behavior [7]. But since the publication of Cohen and Felson's seminal paper in 1979 on *Routine Activity Theory (RAT)* [8], the field of criminology has witnessed a significant shift towards investigating environmental factors as primary determinants of criminal behavior. Environmental criminology puts emphasis on observable macroscopic patterns to develop hypotheses pertaining to situational opportunities. According to the original *RAT*, criminal events occur at the convergence in space and time of three key elements: a motivated offender, an attractive target, and a lack of capable guardianship. The theory emerged from the observation of a singificant correlation between fluctuations in home activities and variations in crime rates. Cohen & Felson postulated that, by increasing the amount of time potential victims spent within their residences, they reduced potential opportunities for criminals to commit offenses. The theory has stood the test of time and serves as the foundational framework for subsequent theories of environmental criminology. The reason for the success can be attributed to the wealth of empirical confirmations that came afterwards. Focusing on the victimization risk, Biderman et al. [9] discovered that individuals most at risk of victimization were more likely to be engaging in lifestyles characterized by increased presence in public spaces (particularly during nighttime), extended periods of separation from the household and proximity to high-offending groups. Additional confirmation was supplied by the examination of the victimization data from Seattle of 1990 [10]. After the publication of RAT, a new pattern was discovered: the clustring of crimes in hotspot in urban centers. Using data from reported crimes in the city of Minneapolis, Sherman et al. revealed that a significant majority of reported crimes (exceeding 50%), were concentrated within a disproportionately small number of places, accounting for a mere 3% of the total area [11]. These hotspots exhibited a notable level of temporal stability over extended periods of time. To incorporate the aforementioned emprical regularity in a new theorethical framework, Brantingham and Brantingham developed *Crime Pattern Theory*[12]. It combines *RAT* with a rational choice perspective in the context of criminal decision-making. The authors suggest that criminals engage in a hierarchical decision-making process as they navigate their everyday activities, selecting specific geographical locations on factors such as effort, anticipated risk, and potential rewards. They posits that hotspots emerge in locations with high levels of activity which offer abundant opportunities for criminals. Studies have demonstrated there is a correlation between high levels of criminal activity and busy nodes such as transit stations or places of interest. Additionally, research has also found that areas with weaker guardianship and proximity to the residences of criminals intensify this effect ([7], p.135). An alternative rationale for the existence of hotspots is provided by *Place Management Theory* [13], which also takes into account the association between crime and factors such as poverty and race. According to Eck & Madensen, guardianship in crime prevention extends beyond police officers, include also place managers. Given that minorities and low socioeconomic groups possess limited political influence, property owners feel a lack of accountability for correct place management. Acting ineffectivly or indifferently, they neglect buildings, permit the accumulation of code violations and contribute to the emergence of conditions that facilitate criminal activities. According to the researchers, the theory suggests that the correlation between poverty and race is spurious.

2.2 Agent-Based Modeling in Urban Crime Research

Since the mid-2000s, there has been a noticeable increase in the prominence of criminological research that use Agent-Based Modeling (ABM) as its chosen technique. Environmental criminology presents a perfect field to develop an ABM as the local interactions between criminals and residents induce emergent regularities. Moreover, due to the frequent unavailability or incompleteness of crime data and the impracticality or ethical concerns surrounding experiments for crime reduction, ABM can potentially function as a substitute for empirical research. A recent paper by Groff et al. [14] presents an overview of the current advancements in ABM in urban crime. The authors claim that a unique challenge in ABM is evaluating the validity. They thus present three types of validity which, according to their perspective, must be upheld: *empirical external validity*, which can be attained by verifying consistent findings across various model definitions and by ensuring that results can be replicated and made transparent; *statistical conclusion validity*, which can be achieved by running the model multiple times, analyzing the distribution of outcomes, and subjecting it to appropriate statistical tests; *empirical validation*, which involves the calibration of model parameters on known empirical data and conducting sensitivity analysis in situations of uncertainty. They also cite the paper of Wang et al. [15] which mentions which regularities should be observed in generated data. Among these factors, there are a high degree of crime clustering, few offenders responsible for a large proporton of crime, victmization concentrated among a small number of victims and a high risk of a crime occuring nearby a crime that just occured. In their review, the authors also advise that a researcher should include a complete specification of the model to ensure perfect reproducibility. The list should include the criminological theory employed, the used software, the number of agents as well as the spatial and temporal characteristics and the decision-making process of the offenders. Finally, they categorize ABM in three categories based on the primary objective of the research paper: policy-focused models, theoretical models, and forecast models. The objective of this thesis is to develop a model suitable for forecasting crimes and investating crime reduction policies. As such, an examination of the two categories with two exemplary papers was conducted. Regarding the forecast model, the recent paper of Zhu & Wang (2020) [16] was reviewed. In their specification, agents are characterized by a prescribed set of activities, consisting of both fixed (work and rest) and flexible components. Additionally, certain agents possess an inherent trait related to criminal motive. Upon completion of the model development, a sensitivity analysis was conducted. Subsequently, the researchers proceeded to gather results and validate the model using real data. This validation process involved the utilization of three metrics: the hit rate (H), which quantifies the ratio of reported crimes within predicted hotspots to the total number of crimes; the forecast precision index (FPI), which adjusts the hit rate based on the study area; and the forecast accuracy index (FAI), which assesses the difference between reported and simulated crimes occurring within predicted hotspots, adjusted by the number of crimes. The model with the highest level of accuracy demonstrates an hit rate of 63% and a forecast accuracy index (FAI) approaching 1 thus demonstrating the possibility to achieve a fairly accurate outcome with ABM. Regarding policy, the paper of Weisburd et al. was examined [17]. The primary objective of their research inquiry was to design and execute an experimental study aimed at isolating the impact of hotspot policing on crime rates. The researchers constructed a fictitious borough in the United States consisting of grid cells that possessed varying levels of riskiness scores. This simulated borough was populated by 40,000 individuals, of which 9.25% were involved in criminal activities. Additionally, the borough was equipped with a police force of 36 officers. The study provided evidence that the implementation of high-intensity hotspot policing, characterized by a greater allocation of officers exclusively dedicated to hotspots, resulted in a 10% decrease in the average number of robberies. In contrast, a low-intensity hotspot policing, which involved most officers patrolling hotspots only 50% of the time, led to a 2% reduction in crime.

3. Collecting Empirical Data

3.1 Geographic, Demographic & Economic Data

In order to create the spatial matrix that enabled agent movement, it was necessary to acquire Geographic Information System (GIS) data of the city of Milan. Specifically, data partaining to the road and public transportation systems, as well as the architectural structures and administrative boundaries. Although the city of Milan has freely accessible and up-to-date GIS data[18], the OpenStreetMap API was used to strengthen the model's adaptability to other urban areas. The API is provided free of charge and enables the retrieval of current geographical data to a specified location. The `osmnx` package [19] provides a convenient implementation of the API for street networks in Python, allowing users to obtain a multi-directed graph, simplified according to a specific tolerance. In the current model, a threshold of 20m was selected as a higher resolution seemed unnecessary for the scope of this thesis. The `osmnx` package also provides access to speed limit data for individual roads, enabling the estimation of trip times between nodes. To capture the effect of traffic congestion, trip times were calculated on speeds 30% under the limit. The downloaded and simplified road network for this thesis has 9321 nodes and 60859 edges. However, the retrieval of public transport routes is not yet implemented by `osmnx`. By utilizing OverPass Turbo [20], a web-based platform designed for the OpenStreetMap API, data on subway, bus, and tram routes was successfully acquired. Subsequently, the obtained graph was merged with the existing road network, connecting the stations with the nearest road node. Projected speeds for public transports were 65 km/h for the subway, and 35 km/h for trams. While `osmnx` also allows to download data on buildings ("features") from OpenStreetMap, shapefiles from the Milan GIS database were used, having already pre-processed the included 33.744 buildings. Based on the functional classification provided in the database, each building was classified into several categories, namely residential, occupational, daytime recreational, or nighttime recreational spaces. OpenStreetMap, however, frequently lacks information on official or neighborhood boundaries. A file containing the 88 neighborhoods of Milan, referred to as "NIL - Nuclei d'Identità Locale", was obtained from the Milan database [21], along with the corresponding 2022 demographic data [22]. Although income data is not directly available for every NIL, the Italian Ministry of Finances offers binned income data by postal code (CAP) [23]. A skewed normal income distribution was constructed for every NIL by taking into account data for every intersecting CAP. At the end of the thesis, an appendix with the computed income distributions is available A.1.

3.2 Crime Data

As previously mentioned, a significant challenge in developing crime models stems from the inherent problems associated with obtaining reliable criminal data. However, since one of the objectives of this model is to forecast actual crime data, it is necessary to conduct some form of model validation. To properly calibrate model parameters, data on the 2019 annual reported crimes was acquired from the ISTAT website [24]. Nevertheless, it was necessary to acquire more granular data in order to validate the accuracy of the forecasts. Andrea Caldarini, a security expert, supported the project by granting permission to use the dataset of *MilanoCrimeMonitoring* [25]. In this platform, he has meticulously gathered crime-related information from multiple sources, including newspapers and social networks, spanning the period from January 2022 to June 2023. Additionally, cooperation was also offered by MineCrime, a privately-owned organization specialized in the aggregation of criminal information from diverse journalistic outlets. They provided crime data for every initial quarter from the year 2020 to 2023 [26]. The obtained dataset is bound to be incomplete: crimes are, by nature, underreported and only sensational crimes that occur in areas of interests (e.g. nightlife or popular turist destinations) receive media coverage. However, these datasets serve as the most extensive and accessible sources for this thesis, with the potential for further expansion in the future. A prelminary analysis of this dataset provides insights that could potentially contribute to the development of the model. The initial step involves calculating the correlation between the average income of neighborhoods and the proportion of crimes committed in these neighborhoods. When examining the data from both Milano Crime Monitoring and MineCrime, it is evident that there exists a strong positive correlation, around 0.4. A subsequent examination of the distribution of crime among various neighborhoods indicates that certain areas exhibit a notably high prevalence of criminal activity. These areas include "Duomo," "Stazione Centrale - Ponte Seveso," and "Buenos Aires - Porta Venezia." The "Duomo" area, for instance, accounts for 7.06% and 8.45% of the total reported offences in the two datasets. These findings are aligned with Crime Pattern Theory[12], as they support the idea that crime tends to occur more frequently in densely populated and affluent areas as they provide better opportunities for illicit activities. However, it also indicates that the crime data is not suitable for inferring the residential locations of offenders, as it seems highly improbable that the bulk of criminals reside in more affluent areas. Therefore, when assigning places of residence to `Criminals`, it was assumed that areas with lower income levels will have a higher concentration of individuals with criminal tendencies. This assumption, other than being in accordance with Place Management Theory[13], also relates well to empirical findings. Poverty and crime are notoriously correlated. A noteworthy occurrence is the neighborhood of "Taliedo - Morsenchio - Q.re Forlanini", which exhibits a substantial prevalence of criminal activities according to the MineCrime dataset (11%); however it conversely ranks among the least crime-ridden areas in the Milano Crime Monitoring dataset (2%). The potential explanation could be attributed to the different data collection methodologies employed for the two datasets. Milano Crime Monitoring dataset relies more on social media, while MineCrime solely relies on newspaper sources. Forlanini is a known criminal neighborhood. As such, it might be that crimes are reported less frequently on social media than on newspapers, since newspapers try to offer a more comprehensive depiction of the city while posts on social media focus mostly on central areas. The last significant fact is that the mean of the variance across the years for neighborhoods is 2.78%. This implies that, although there may be certain variations, areas with high crime rates tend to exhibit a consistent pattern across time. It should also be noted that part of this analysis is dependent on the inclusion of years during which the Covid epidemic occurred, which had a substantial impact on crime patterns. For example, the true value of the variance during "conventional" years, might be different than the calculated figure. Lastly, to assess the clustering of crimes, a DBSCAN algorithm was used with a threshold of 500m to compute the silhouette score. The silhouette score provides a measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation). The score ranges from -1 to +1, with a perfect value of 1 indicating that clusters are completely separated between each other. The observed value is -0.09, which seems surprising since it suggests that data does not exhibit a high degree of clustering, thus contradicting a common pattern of crime data. The value might be a clue that the dataset is incomplete.

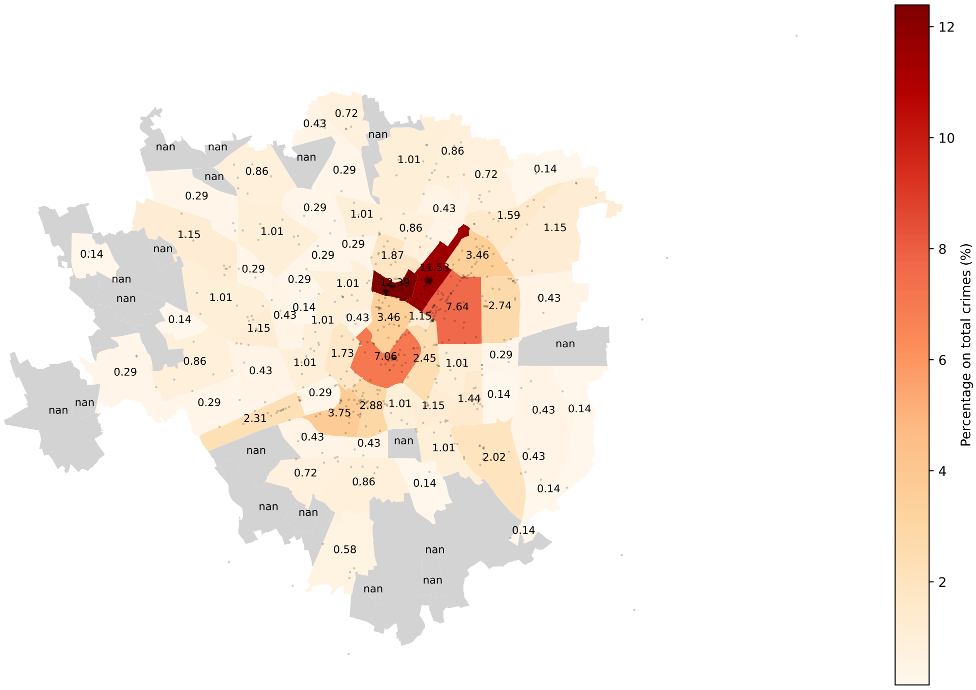

Figure 3.1: Crime Distribution according to Milano Crime Monitoring (all crimes)

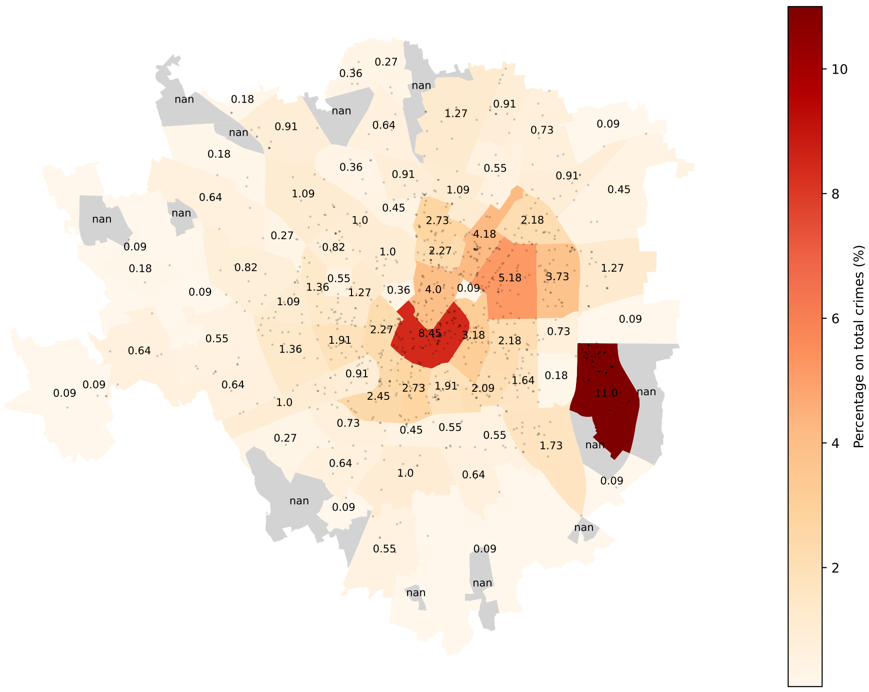

Figure 3.2: Crime Distribution according to MineCrime (robbery and theft)

4. The Model

This chapter will present a `StreetCrime` model. The model is inspired by: the criminological theory of Crime Pattern Theory [12], the ABM implementation of Zhu & Wang [16] and the recommendations of Groff et al. in their review[14]. The goals of the model presented in this chapter are the following: firstly, to create a model of crime that can reliably detect hotspots of street crime in the city of Milan; secondly is to provide a general open-source framework for criminal ABM which can be extended to different cities, asssumptions or situations with ease. This objective represents the cornerstone of generative research. For this reason, the dataset and the code is freely accessible on Github[27] .

4.1 Agents Specification

The model is developed using the `mesa-geo` package [28]. In `mesa-geo` the base class of all agents is the `GeoAgent` class. A `GeoAgent` is created with a `unique_id` and a `geometry(POINT)` attribute which defines his location on the map. At every step of the, `GeoAgents` execute their `step` method in random order. The subsequent subsections provide a comprehensive overview of each class.

4.1.1 InformedMover

The `InformedMover` serves as the foundational class for all agents within the model. The InformedMover participates in urban activities by navigating through the city utilizing either the public transportation system or the road network. The latter is selected only if they possess a personal vehicle and the destination is further than a distance threshold (set

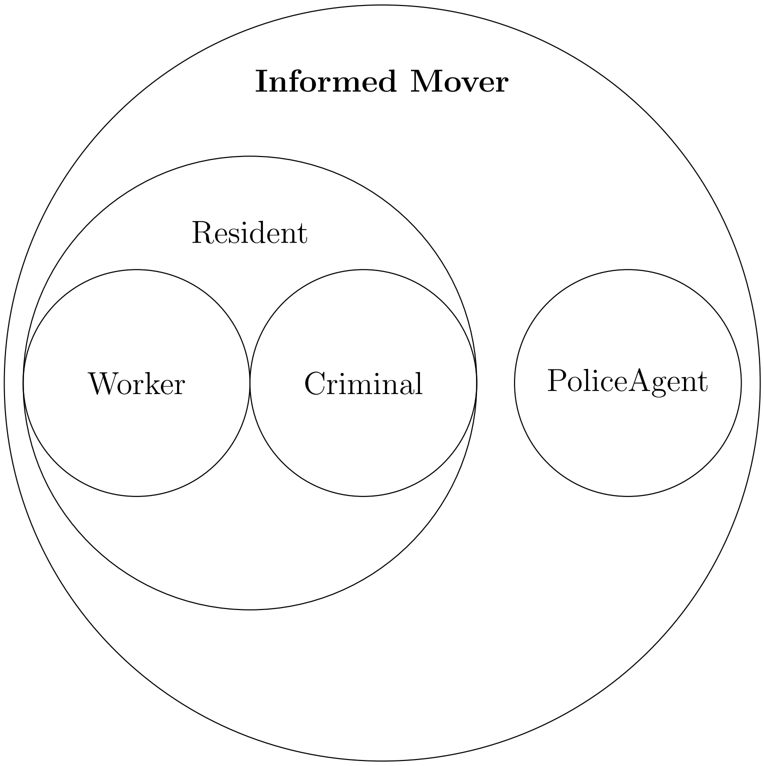

Figure 4.1: A diagram representing class dependencies

at 5km). In order to determine the next activity, the individual employs a probabilistic decision-making process based on his available information. Upon selecting an activity, the individual proceeds to calculate the shortest path between their current location and the desired destination using Dijkstra's algorithm. They then navigate around the map through the network nodes, by adjusting their position according to the length of the model step and their current speed. When arrived at the activity, a countdown is generated according to a specified normal distribution. Everyday at midnight, the `InformedMover` updates his information on the city by retrieving data from the previous day, according to his `p_information` attribute. When the value of `p_information`, is set to 1, the individual possesses perfect information. Conversely, with `p_information = 0`, the individual lacks any knowledge about the city. If the type of information is a list (e.g. the crimes occured in the previous day), the agent obtains a random sample of `p_information`. If the type of information is instead a value (e.g. the number of workers that passed through a certain neighborhood the past day), the agent obtains the value v disturbed by $\mathcal{N}(0, \sigma^2)$ where $\sigma^2 = \frac{(1-\text{p information})\cdot v}{2}$, such that there's a 95% probability that the disturbed value will be between $v \pm [1 - \text{p information}] * v$.

4.1.2 Resident

The `Resident` class extends the functionality of the `InformedMover` class with the inclusion of a home address, a resting schedule and a yearly income. The initiation and termination times of rest periods are determined on a daily basis by following a predetermined normal distribution and by respecting the working schedule for `Workers`. Once the designated period for rest arrives, the `Resident` returns to his home and remains there until the conclusion of the resting period. His income is extracted from the estimated income distribution of their neighborhood of residence. If an individual's income exceeds a specified threshold, there is a higher probability that they will possess an automobile.

4.1.3 PoliceAgent

The `PoliceAgent` has the role of a silent guardian. He travels around the city of Milan, preventing criminal activity simply by his sheer presence. Unless specified otherwise, the `PoliceAgent` makes a strategic decision on the selection of neighborhoods to visit. The agent engages in hotspot policing, by making a choice with recent and historical crime statistics, of which they possess perfect knowledge.

4.1.4 Worker

The `Worker` class encompasses the majority of the agents within the model. A `Worker` is assigned a specific work location and working hours. Outside of this period, if he does not have to rest, he chooses an activity to go to, making the selection on various criteria, such as the proximity, the recent and the historical crimes in the location's neighborhood. When on the journey, the `Worker` can be a target for `Criminals`, although `Pickpockets` would not pose a threat if he travels by car. The `Worker` possesses also a `crime_attractiveness` attribute, contigent on his income, which affects the probability of being targeted by a criminal. Finally, he has a `self_defense` attribute, initiated randomly, which influences the likelihood of resisting a robbery.

4.1.5 Criminal

The `Criminal` class is shared by both `Pickpockets` and `Robbers`. The `Criminal` exhibits a `crime_motivation` attribute that depends on his income. In accordance with Crime Pattern Theory [12], the `Criminal` engages in criminal behavior while carring out his daily routine. Nevertheless, the Criminal adopts a rational decision-making approach, by strategically selecting affluent and under-policed locations. If he is a `Pickpocket`, he will also tend to target highly frequented areas, as these locations offer greater opportunities for committing crimes. If the individual is instead a `Robber`, he will strategically select less populated locations in order to reduce the presence of potential witnesses, hence reducing the likelihood of encountering any deterrents to his criminal activity. While travelling, the Criminal will actively seek for potential crime opportunities, by surveying the area within a specified range of awareness (set at 150m). In the event that he encounters a `PoliceAgent`, he will not commit any crime but will mentally annotate his location for future reference, so that overtime he will be able to develop a map of the areas subjected to frequent patrols and avoid them. If instead there are no `PoliceAgents` present but there is at least one `Worker`, the `Criminal` will select the victim with the highest level of `crime_attractiveness`. If the `crime_attractiveness` of the victim is lower than the `Criminal` motivation, the `Criminal` will not commit any crime. If, on the other hand, he encounters a suitable target, he will attempt to conduct a criminal act, the success of which is contingent upon the following conditions. If the criminal is a `Pickpocket`:

$$\textit{crime motivation} + \textit{crowd effect} \cdot \text{n of witnesses} + \mathcal{N}(0, 0.5) \geq \mathcal{U}(0, 1)$$

If the criminal is instead a `Robber`:

$$\begin{cases} \textit{crime motivation} - \textit{crowd effect} \cdot \textit{n of witnesses} + \mathcal{N}(0, 0.5) \geq \mathcal{U}(0, 1) \\ \textit{crime motivation} \geq \textit{victim's self defence} \end{cases}$$

4.2 Evaluation of the Model

In this section the collected results will be analyzed. In the first section a sensitivity analysis will be performed by exploring the effect that varying parameters have on generated data, to ensure a form of *empirical external validity* and *statistical conclusion validity*, as defined by Groff et al. in [14]. In the second section *empirical validity* will be investigated. With the use of some performance metrics, different model specifications will be compared to determine which is the best performing one and how closely it resembles reality. The final section will explore the evolution of crime patterns in the model with two long-run iterations and a comprehensive analysis taking into account all the generated data.

4.2.1 Sensitivity Analysis

The objective of this subsection is to determine whether the model is heavily reliant on the parameters and whether the effect of the parameters is expected from model definition, from criminological theories and from empirical data. In table 4.1 various parameters are presented, with the confidence intervals of the coefficients resulting from a multiple regression with logarithmic transformations $log(Y) = log(X)\beta$.

Model Parameter	$\frac{\Delta\%\ \textbf{Crime Rate}}{\Delta x\%}$	$\frac{\Delta\%\ \textbf{Successful Crimes}}{\Delta x\%}$	$\frac{\Delta\%\ \textbf{Robberies}}{\Delta x\%}$	$\frac{\Delta\%\ \textbf{Pickpocketing}}{\Delta x\%}$
P. of PoliceAgent	$[-111.16, 50.57]$	$[-89.29, -8.64]^{*}$	$[-36.1, 173.41]$	$[-94.99, 37.96]$
P. of Pickpocket	$[-9.65, 188.32]$	$[-112.52, -13.8]^{*}$	$[-87.75, 168.51]$	$[-119.2, 43.55]$
P. of Robber	$[-28.55, 120.67]$	$[-84.43, -10.02]^{*}$	$[30.16, 223.31]^{*}$	$[-154.75, -32.08]^{*}$
P. of Worker	$[-1238.77, 1325.3]$	$[-1379.75, -101.14]^{*}$	$[-718.35, 2600.64]$	$[-1773.61, 334.29]$
P. information of Worker	$[-134.15, 259.98]$	$[-272.58, -76.05]^{*}$	$[-283.55, 278.09]$	$[-214.44, 109.57]$
Opportunity awareness	$[78.72, 552.15]^{*}$	$[10.99, 247.07]^{*}$	$[-680.74, 62.57]$	$[-158.5, 230.7]$

Table 4.1: Sensitivity Analysis (95% CIs)

The table provides valuable insights into the model's behavior. Firstly, it is essential to note that only a handful of selected parameters display statistical significance. Additionally, a high degree of sensitivity to changes in specifications can be observed. A mere 1% alteration in these parameters results in a substantial percentage change in various model outcomes. Three potential explanations can help clarify the results. Firstly, the role of the limited sample size may play a role. Currently, only 30 runs, mostly spanning three days, have been completed. Given the stochastic nature of the model, this sample size may not suffice to yield statistically significant results. Secondly, the model's non-linearity may induce significant biasedness when completing a multilinear regression. The last form of uncertainty may be given from the high collinearity of agents proportions, since there exist a perfect linear reletionship between them. If we now turn to the direction of the effect of the significant parameters, results that seem in line with the assumptions of Crime Pattern Theory [12] can be observed. For example, the introduction of additional police agents, leads to a significant reduction in the proportion of successful crimes. However conclusions about the observed effect must be drawn with care, as the current sample may not provide criminals with enough time to adapt their expectations. Another confirmation is given by the coefficient for the proportion of criminals which seems to have no effect on the crime rate but reduces significantly the proportion of successful crimes. CPT might suggest that while the presence of more criminals determines a higher possibility of being involved in a crime for a single worker, criminal opportunities are reduced since there are less workers to target, hence having a mixed effect on the crime rate. Furthermore, a higher proportion of criminals means that the risk of being detected by a guardian increases substantially. Further investigation in subsequent sections will offer a more comprehensive understanding of the model's dynamics.

4.2.2 Empirical Validation

The performance of the model can be now evaluated against real-world data, by running multiple iterations with different parameters. The used model parameters are collected in table 4.2, while the results are in table 4.3. Only the 10 best runs are presented

Run	% Police	% Pickpockets	% Robbers	Crowd Effect	% Worker information	Criminal Decision Rule
1	10.0	10.0	5.0	1.0	50.0	Logarithmic weighted distance
2	2.0	4.0	8.0	1.0	50.0	Linearly weighted distance
3	2.0	4.0	5.0	1.0	50.0	Linearly weighted distance
4	2.0	4.0	2.0	1.0	50.0	Linearly weighted distance
5	2.0	14.0	2.0	1.0	50.0	Linearly weighted distance
6	2.0	2.0	2.0	1.0	50.0	Linearly weighted distance
7	10.0	10.0	5.0	1.0	40.0	Logarithmic weighted distance
8	2.0	8.0	2.0	1.0	50.0	Linearly weighted distance
9	3.0	4.0	2.0	1.0	50.0	Linearly weighted distance
10	2.0	4.0	14.0	1.0	50.0	Linearly weighted distance

Table 4.2: Model Parameters (3 Days, 700 InformedMovers, steps of 15 minutes)

here. Each run required significant computational resources, with a single run taking approximately 3 hours on a system featuring an Intel i7-6770k processor and 16GB of RAM. To expedite the collection of data, multiprocessing through the `deco` package was implemented for the concurrent execution of four runs, substantially reducing the overall runtime. Inspired by the metrics proposed by Zhu & Wang [16] three key metrics were created in a similar fashion. The first, roughly corresponding to Zhu & Wang hit rate, is the average deviation between the generated neighborhood's proportion of crime and the actual neighborhood's crime proportion. Promisingly, the model demonstrated the capability to predict crime proportions with a margin of accuracy within 1%. However the result might be skewed by the fact that, given the clustring of crimes, only a handful of neighborhoods. As a matter of fact, given the clustering of crimes, it may be the case that the deviation is low for most neighborhoods, but potentially high for the few genuine hotspots. The second metric, inspired by the Forecast Accuracy Index, assesses the average distance of a generated crime from the location of the closest real crime. The key assumption is that every link between generated and actual crimes must be unique as to avoid artificially reducing the deviation. Once again findings are promising, indicating an average deviation of less than 300m. Lastly, the deviation was computed between the actual and the generated crime rate. In this case, the overestimation is significant, by a factor of 2. The discrepancy can be attributed to two factors: the parametrization of the model and the characterization of the criminal population. Empirical data suggests that the expected number of crimes in the city of Milan for a population of 700 inhabitants would be 1.44. To generate a substantial amount of data, the proportion of criminals had to be increased, affecting however the relationship with the real-data proportion. Secondly, criminals are assumed to be hyper-rational and hyper-dedicated individuals. They are devoid of traditional employment and engage continuously in criminal activities, with no breaks and no mechanisms for imprisonment or reduced motivation after a crime.

4.2.3 Long-Run Iterations and Comprehensive Analysis

In this final section, data derived from two extended iterations of the model will be analyzed. The model was run for 14 days for 700 residents. The iterations adhere to the standard specification which includes: hotspot policing with perfect information for police agents, perfect information for criminals and the specified parameters outlined in run 1 of table 4.3. Running longer iterations is central to the evaluation of the model because it allows of the observation of the evolution of agent behaviors through time. Specifically, 3 performance metrics will be analyzed, which refer to the observed regularities of Wang et al. [15]: clustering of crimes, the phenomenon of repeated victimization, and

Run	Deviation from real proportion of crime (%)	Deviation from closest crime (m)	Deviation from crime rate (yearly crimes/100k residents)
1	$[0.67, 1.33]$	$[-140.14, 239.64]$	26649.86
2	$[0.68, 1.37]$	$[65.02, 222.07]$	144840.33
3	$[0.59, 1.27]$	$[96.15, 216.39]$	193507.0
4	$[0.7, 1.39]$	$[93.55, 226.48]$	132673.67
5	$[0.65, 1.32]$	$[142.04, 232.95]$	290840.33
6	$[0.69, 1.35]$	$[36.45, 378.76]$	59673.67
7	$[0.71, 1.38]$	$[143.07, 300.63]$	271069.5
8	$[0.66, 1.33]$	$[167.19, 291.68]$	193507.0
9	$[0.68, 1.37]$	$[113.16, 350.77]$	157007.0
10	$[0.71, 1.41]$	$[165.86, 315.38]$	376007.0

Table 4.3: Comparison between generated and real data (95% CIs)

the concentration of offenses among a minority of offenders. Additionally, we will calculate the daily variation in the neighborhood proportion of crimes, akin to the approach in section 3.2. The results are mixed with respect to the ones obtained from shorter iterations. Firsly, the clustering of crimes was assessed again with the silhouette score. The value of 0.13 suggests that crimes within the same cluster tend to be relatively close, although with some overlap among clusters. Notably, this value is higher than the one observed for real crime data, which was -0.09. At least for what concerns clustering, there seems to be a discrepancy between empirical data and established pattern in the literature. Furthermore, the Gini index was employed to assess repeated victimization and repeated offense. The Gini index quantifies the inequality in the distribution of a specific value within a population, with values ranging from 0 (perfect equality) to 1 (perfect inequality). For offenders, we obtained a Gini index of 0.15, indicating a relatively equal distribution, while 0.14 was obtained for victims. While denoting some variations, the values describe a fairly equal distribution. The source of this small inequality might be given by the unequal income distribution, which directly affects crime attractiveness and crime motivation. Other agent-based model have obtained a similar result with repeated victimization ranging from 0.15 to 0.25 [29]. Tseloni & al. [30] construct instead a Lorenz distribution of victims, specifying that theft and robberies are among the lowest repeated crimes, but they do not present the computed value. The Gini index is indeed likely to be lower for theft and robberies with respect to other crimes because they rely on situational opportunities. Unfortunately, without having access to empirical data, it is hard to make a definitive conclusion. as previously done with the real data, the daily average deviation of neighborhoods from their expected percentages was computed, resulting in a value of 15%. While the value is not directly comparable to the annual real-world value of 2.98%, a daily varation of this kind appears unlikely. This value indicates that in the generated dataset, hotspots are dynamic and heavily dependent on the temporal frame. This assumption seems to be confirmed from the distribution in figure 4.2, which represent crimes that are evenly distributed on the map. Although it would have been logical to also conduct an analysis of the impact of time on the crime statistics, similarly to the one conducted in sensitivty analysis (Table 4.1), there was not enough data to determine the statistical significance of time's effect on these statistics. Consequently, the analysis was excluded. Shifting focus to the comprehensive examination, we can visualize the distribution of all generated crimes on Figure ??. While it may seem methodologically incorrect to combine data from different iterations, I argue that it is precisely by combining all data that we obtain a clearer picture of the functioning of the model. Due to the random sampling of agent residences and the heterogeneity of the iterations, each iteration effectively serves as a random sample of residents. The resultant crime distribution should smooth out the heterogeneity, revealing macroscopic regularities. We could thus address the initial generativist question: can this microspecification explain the crime pattern of the city of Milan? From the plot we can conclude that this is not the case. Crimes exhibit a uniform distribution among neighborhoods, with the predicted "hotspots" showing only a marginal increase of 0.4% compared to the safest neighborhoods. Furthermore, crimes tend to concentrate on the outer ring of the city, likely due to the higher prevalence of criminal residences in outlying neighborhoods. However, notable effects emerge from busy nodes such as "Stazione Centrale," "Stazione Porta Garibaldi," and the concentration along the circular loop line. Interestingly, higher crime rates near "Q.re Forlanini" are correctly predicted, in accordance with the MineCrime dataset. The alignment is likely due to to the neighborhood's lower economic status, which affects both real and generated residences for criminals and real crime. Conversely, relatively few crimes are predicted in the "Duomo" area, which was a significant hotspot in the dataset. Several explanations may account for this divergence. Firstly, areas near the city center may be too distant from criminal residences, with current weight assumptions favoring closer, but equally profitable areas based on risk-reward analysis. The result is consistent both

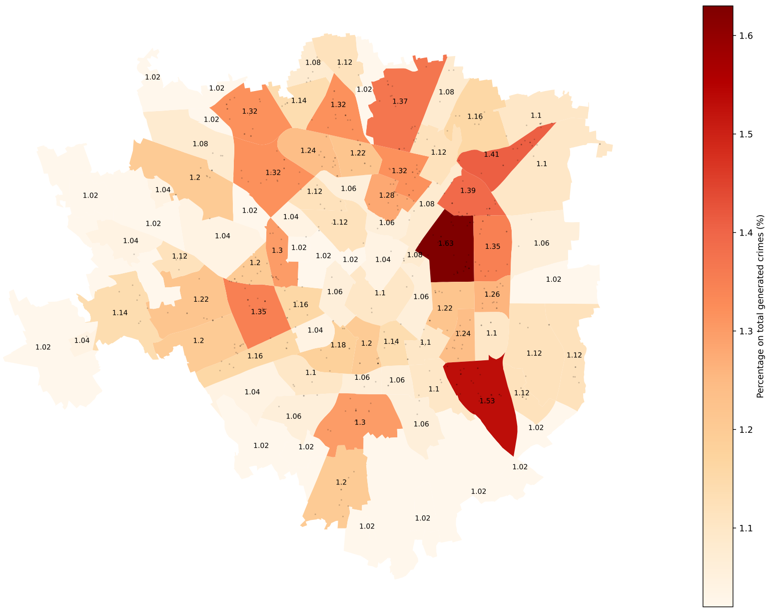

Figure 4.2: Generated Crime Distribution

when comparing linear and logarithmic weights. Secondly, the explanation may lie in the relatively small size of the "Duomo" neighborhood. It may indeed be the case that the neighborhood's central role in Milan's urban life is lost to the generalization of the buildings in the dataset. For the future, there is the possibility of achieving a better approximation of reality with the retrieval of buildings from OpenStreetMap with `osmnx`.

5. Limitations and Future Directions

The preceding findings underscore several critical limitations in the interpretation and application of this model, while also illuminating promising avenues for future research. One of the central limitations of this model pertains to the definition the Criminal class. A definite reevaluation of the activity decision rule is needed, particularly by parametrizing the weight of the distance condition. According to the generated crime distribution, criminals seem to commit crimes too close to their residential homes and might therefore need a different incentive to commit areas in real hotspots. The model also relies on the simplistic assumption of assigning criminal residences based on the income distribution of each neighborhood. While reasonable, it significantly influences subsequent criminal behavior. Finally, the absence of criminal social networks within the model represents a substantial limitation as such networks play a pivotal role in shaping crime motivations and crime influences. Another issue is related to the current level of realism, particularly regarding the fixed activities of agents (work and rest). Additional fixed activities should be considered, as they may exert notable effects on the transportation patterns and consequently on crime dynamics. Finally, the code implementation could benefit from revision and optimization. This would facilitate data collection for longer durations, involve more agents, and better reflect the actual agent proportions. However, with careful consideration and calibration, the model may offer a somewhat accurate estimation of crime hotspots and crime distribution within a city. Indeed, using the current model as baseline, there could be multiple research questions that could be explored on the same basis. For example, a comparative analysis of different criminological theories can be conducted to determine which provides the most accurate depiction of the real crime dynamics. The model could be employed also to explore experimental questions that contribute to the development of crime reduction strategies. For instance, by utilizing the dataset on street lighting present in the Milan dataset [18], one can identify the optimal locations to position street lighting and ebaluate is effectiveness as a crime prevention measures. Other research questions could be related to the effect of policing: building on the work of Weisburd et al. [17] the model could investigate the impact of hotspot policing by considering various levels of information proportion. Collaboration with public institutions, particularly law enforcement agencies, could enhance the model's data sources, for example by providing access to data on criminal residences and conviction rates, already in possession of the Police. In closing, it is essential to underscore that the true significance of this model lies not solely in its presented results but in the establishment of an open-access framework. This framework offers universal accessibility, adaptability to diverse research questions, environments and data reproducibility through unique identifiers. A collective endeavor towards open generative science is pivotal in advancing the field of generative social science.

Bibliography

[1] J. M. Epstein, "Agent-based computational models and generative social science," *Complexity*, vol. 4, pp. 41–60, May 1999. DOI: 10.1002/(sici)1099-0526(199905/06)4:5<41::aid-cplx9>3.0.co;2-f.

[2] J. M. Epstein, *Generative Social Science*. Princeton University Press, Jan. 2012.

[3] W. Buckley, "Society as a complex adaptive system," *Emergence: Complexity and Organization*, vol. 10, pp. 490–513, Jul. 2008. DOI: 10.4324/9781315130569-67.

[4] B. L. Heath and R. R. Hill, "Some insights into the emergence of agent-based modelling," *Journal of Simulation*, vol. 4, pp. 163–169, Sep. 2010. DOI: 10.1057/jos.2010.16. (visited on 11/16/2020).

[5] T. C. Schelling, "Models of segregation," *The American Economic Review*, vol. 59, pp. 488–493, 1969. [Online]. Available: https://www.jstor.org/stable/1823701.

[6] R. Axtell, R. Axelrod, J. M. Epstein, and M. D. Cohen, "Aligning simulation models: A case study and results," *Computational and Mathematical Organization Theory*, vol. 1, pp. 123–141, Feb. 1996. DOI: 10.1007/bf01299065. (visited on 04/10/2021).

[7] P. Wilcox and F. T. Cullen, "Situational opportunity theories of crime," *Annual Review of Criminology*, vol. 1, pp. 123–148, Jan. 2018. DOI: 10.1146/annurev-criminol-032317-092421.

[8] L. E. Cohen and M. Felson, "Social change and crime rate trends: A routine activity approach," *American Sociological Review*, vol. 44, pp. 588–608, Aug. 1979. [Online]. Available: https://www.jstor.org/stable/2094589.

[9] A. D. Biderman, M. J. Hindelang, M. R. Gottfredson, and J. Garofalo, "Victims of personal crime: An empirical foundation for a theory of personal victimization.," *Contemporary Sociology*, vol. 11, p. 41, Jan. 1982. DOI: 10.2307/2066613.

[10] T. D. Miethe, *Testing theories of criminality and victimization in seattle, 19601990*, 2006. DOI: 10.3886/ICPSR09741.v2.

[11] L. W. SHERMAN, P. R. GARTIN, and M. E. BUERGER, "Hot spots of predatory crime: Routine activities and the criminology of place," *Criminology*, vol. 27, pp. 27–56, Feb. 1989. DOI: 10.1111/j.1745-9125.1989.tb00862.x.

[12] P. Brantingham and P. Brantingham, "Crime pattern theory," *Oxford Research Encyclopedia of Criminology and Criminal Justice*, Jan. 2021. DOI: 10.1093/acrefore/9780190264079.013.8.

[13] J. E. Eck and T. Madensen-Herold, "Place management, guardianship, and the establishment of order," in (Deterrence, choice, and crime : contemporary perspectives), Deterrence, choice, and crime : contemporary perspectives. Routledge, An Imprint Of The Taylor Francis Group, 2018, pp. 269–307.

[14] E. R. Groff, S. D. Johnson, and A. Thornton, "State of the art in agent-based modeling of urban crime: An overview," *Journal of Quantitative Criminology*, vol. 35, pp. 155–193, Feb. 2018. DOI: 10.1007/s10940-018-9376-y. (visited on 05/04/2020).

[15] X. Wang, L. Liu, and J. Eck, "Crime simulation using gis and artificial intelligent agents," in (Artificial Crime Analysis Systems: Using Computer Simulations and Geographic Information Systems), Artificial Crime Analysis Systems: Using Computer Simulations and Geographic Information Systems. IGI Publishing, 2008, pp. 209–225. DOI: 10.4018/978-1-59904-591-7.ch011.

[16] H. Zhu and F. Wang, "An agent-based model for simulating urban crime with improved daily routines," *Computers, Environment and Urban Systems*, vol. 89, p. 101 680, Sep. 2021. DOI: 10.1016/j.compenvurbsys.2021.101680. (visited on 03/21/2022).

[17] D. WEISBURD, A. A. BRAGA, E. R. GROFF, and A. WOODITCH, "Can hot spots policing reduce crime in urban areas? an agent-based simulation*," *Criminology*, vol. 55, pp. 137–173, Feb. 2017. DOI: 10.1111/1745-9125.12131.

[18] C. di Milano, *Database topografico 2020*, Milano Geoportale, 2020. [Online]. Available: https://geoportale.comune.milano.it/sit/download-utili/.

[19] G. Boeing, "Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks," *SSRN Electronic Journal*, 2016. DOI: 10.2139/ssrn.2865501. (visited on 05/20/2020).

[20] OpenStreetMap, *Overpass turbo*, Overpass Turbo. [Online]. Available: https://overpass-turbo.eu/.

[21] C. di Milano, *Nuclei d'identità locale (nil) vigenti - pgt 2030 - open data - comune di milano*, Comune di Milano, Feb. 2021. [Online]. Available: https://dati.comune.milano.it/dataset/e8e765fc-d882-40b8-95d8-16ff3d39eb7c.

[22] C. di Milano, *Popolazione residente*, Comune di Milano, Sep. 2023. [Online]. Available: https://www.comune.milano.it/aree-tematiche/dati-statistici/pubblicazioni/popolazione-residente-a-milano.

[23] M. dell'Economia e delle Finanze, *Redditi e principali variabili irpef su base comunale - classificazione: Sub-comunale(cap)*, Dipartimento Finanze, Apr. 2023. [Online]. Available: https://www1.finanze.gov.it/finanze/analisi_stat/public/index.php?search_class%5b0%5d=cCOMUNE&opendata=yes.

[24] I. N. di Statistica, *Delitti denunciati dalle forze di polizia all'autorità giudiziaria*, Istat, 1983. [Online]. Available: http://dati.istat.it/Index.aspx?DataSetCode=dccv_delittips#.

[25] A. Caldarini, *Milano crime monitoring*, MapHub, Apr. 2022. [Online]. Available: https://maphub.net/ARINI/map-3.

[26] M. Crime, *Crime dataset of milan (2020 - 2023)*, Mine Crime, Sep. 2023. [Online]. Available: https://www.minecrime.it/search/map?place=Milano.

[27] A. Amer, *Streetcrime: A generative science approach*, GitHub, Sep. 2023. [Online]. Available: https://github.com/adamamer20/streetcrime.

[28] *Mesa-geo: A GIS extension for the mesa agent-based modeling framework in python*, Association for Computing Machinery, 2022, pp. 1–10. DOI: 10.1145/3557989.3566157. [Online]. Available: https://doi.org/10.1145/3557989.3566157.

[29] D. Birks, M. Townsley, and A. Stewart, "Emergent regularities of interpersonal victimization," *Journal of Research in Crime and Delinquency*, vol. 51, pp. 119–140, Jun. 2013. DOI: `10.1177/0022427813487353`. (visited on 12/14/2021).

[30] A. Tseloni and K. Pease, "Population inequality: The case of repeat crime victimization," *International Review of Victimology*, vol. 12, pp. 75–90, Jan. 2005. DOI: `10.1177/026975800501200105`. (visited on 10/08/2019).

A. Appendix

A.1 City Data

Area	Proportion of the population (%)	Income distribution - Skewed $\mathcal{N}(\xi, \omega, \alpha)$
ADRIANO	1.29	$(\xi^e = 4645.0, \omega^e = 34635.0, \alpha^e = 4329242.0)$
AFFORI	1.84	$(\xi^e = 4753.0, \omega^e = 35038.0, \alpha^e = 18327383.0)$
ASSIANO	0.02	$(\xi^e = 4733.0, \omega^e = 28895.0, \alpha^e = 10345851.0)$
BAGGIO - Q.RE DEGLI OLMI - Q.RE VALSESIA	2.18	$(\xi^e = 4733.0, \omega^e = 28895.0, \alpha^e = 10345851.0)$
BANDE NERE	3.15	$(\xi^e = 4538.0, \omega^e = 47964.0, \alpha^e = 22854982.0)$
BARONA	1.19	$(\xi^e = 4812.0, \omega^e = 31863.0, \alpha^e = 17153785.0)$
BICOCCA	0.63	$(\xi^e = 4544.0, \omega^e = 40700.0, \alpha^e = 5846416.0)$
BOVISA	1.0	$(\xi^e = 4721.0, \omega^e = 35776.0, \alpha^e = 3654839.0)$
BOVISASCA	0.52	$(\xi^e = 4857.0, \omega^e = 28279.0, \alpha^e = 6547274.0)$
BRERA	1.31	$(\xi^e = 4046.0, \omega^e = 123543.0, \alpha^e = 1251604.0)$
BRUZZANO	0.91	$(\xi^e = 4805.0, \omega^e = 31968.0, \alpha^e = 810556.0)$
BUENOS AIRES - PORTA VENEZIA - PORTA MONFORTE	4.47	$(\xi^e = 4275.0, \omega^e = 86562.0, \alpha^e = 6468726.0)$
CANTALUPA	0.04	$(\xi^e = 4812.0, \omega^e = 31863.0, \alpha^e = 17153785.0)$
CASCINA MERLATA	0.07	$(\xi^e = 4831.0, \omega^e = 27022.0, \alpha^e = 17570863.0)$
CHIARAVALLE	0.07	$(\xi^e = 4713.0, \omega^e = 38693.0, \alpha^e = 14988505.0)$
CIMIANO - ROTTOLE - Q.RE FELTRE	1.4	$(\xi^e = 4420.0, \omega^e = 44677.0, \alpha^e = 16589086.0)$
CITTA' STUDI	2.59	$(\xi^e = 4209.0, \omega^e = 68103.0, \alpha^e = 7436163.0)$
COMASINA	0.7	$(\xi^e = 4805.0, \omega^e = 31968.0, \alpha^e = 810556.0)$
CORSICA	1.4	$(\xi^e = 4486.0, \omega^e = 45662.0, \alpha^e = 14200374.0)$
DE ANGELI - MONTE ROSA	1.55	$(\xi^e = 4329.0, \omega^e = 79670.0, \alpha^e = 9550020.0)$
DERGANO	1.71	$(\xi^e = 4695.0, \omega^e = 38005.0, \alpha^e = 18236228.0)$
DUOMO	1.19	$(\xi^e = 4023.0, \omega^e = 149765.0, \alpha^e = 1335788.0)$
FARINI	0.27	$(\xi^e = 4524.0, \omega^e = 54439.0, \alpha^e = 19500924.0)$
FIGINO	0.18	$(\xi^e = 4686.0, \omega^e = 30118.0, \alpha^e = 23352035.0)$
FORZE ARMATE	1.8	$(\xi^e = 4681.0, \omega^e = 36823.0, \alpha^e = 1247771.0)$
GHISOLFA	1.29	$(\xi^e = 4387.0, \omega^e = 70835.0, \alpha^e = 21190185.0)$
GIAMBELLINO	2.21	$(\xi^e = 4545.0, \omega^e = 50357.0, \alpha^e = 24083240.0)$
GIARDINI P.TA VENEZIA	0.0	$(\xi^e = 4091.0, \omega^e = 122468.0, \alpha^e = 2854700.0)$
GORLA - PRECOTTO	2.12	$(\xi^e = 4539.0, \omega^e = 40240.0, \alpha^e = 5713894.0)$
GRATOSOGLIO - Q.RE MISSAGLIA - Q.RE TERRAZZE	1.33	$(\xi^e = 4812.0, \omega^e = 31863.0, \alpha^e = 17153785.0)$
GRECO - SEGNANO	1.16	$(\xi^e = 4596.0, \omega^e = 41086.0, \alpha^e = 5101684.0)$

GUASTALLA	1.11	($\xi^e = 4087.0, \omega^e = 118923.0, \alpha^e = 5323498.0$)
ISOLA	1.62	($\xi^e = 4410.0, \omega^e = 64179.0, \alpha^e = 9702015.0$)
LAMBRATE - ORTICA	0.87	($\xi^e = 4376.0, \omega^e = 48082.0, \alpha^e = 9163787.0$)
LODI - CORVETTO	2.64	($\xi^e = 4663.0, \omega^e = 38444.0, \alpha^e = 21133233.0$)
LORENTEGGIO	1.02	($\xi^e = 4585.0, \omega^e = 43117.0, \alpha^e = 23744669.0$)
LORETO - CASORETTO - NOLO	3.15	($\xi^e = 4466.0, \omega^e = 58796.0, \alpha^e = 13079773.0$)
MACIACHINI - MAGGIOLINA	1.87	($\xi^e = 4489.0, \omega^e = 58085.0, \alpha^e = 11716726.0$)
MAGENTA - S. VITTORE	1.29	($\xi^e = 3861.0, \omega^e = 138332.0, \alpha^e = 9918969.0$)
MAGGIORE - MUSOCCO - CERTOSA	0.63	($\xi^e = 4711.0, \omega^e = 43210.0, \alpha^e = 3802405.0$)
MONCUCCO - SAN CRISTOFORO	0.99	($\xi^e = 4466.0, \omega^e = 56469.0, \alpha^e = 19083845.0$)
MONLUE' - PONTE LAMBRO	0.38	($\xi^e = 4462.0, \omega^e = 35269.0, \alpha^e = 3140621.0$)
MORIVIONE	0.55	($\xi^e = 4517.0, \omega^e = 46620.0, \alpha^e = 4550916.0$)
MUGGIANO	0.22	($\xi^e = 4768.0, \omega^e = 27955.0, \alpha^e = 6200608.0$)
NIGUARDA - CA' GRANDA - PRATO CENTENARO - Q.RE FULVIO TESTI	2.59	($\xi^e = 4635.0, \omega^e = 39741.0, \alpha^e = 16248150.0$)
ORTOMERCATO	0.3	($\xi^e = 4663.0, \omega^e = 38444.0, \alpha^e = 21133233.0$)
PADOVA - TURRO - CRESCENZAGO	2.71	($\xi^e = 4542.0, \omega^e = 41261.0, \alpha^e = 27115585.0$)
PAGANO	1.28	($\xi^e = 3987.0, \omega^e = 120993.0, \alpha^e = 5862956.0$)
PARCO BOSCO IN CITTA'	0.05	($\xi^e = 4650.0, \omega^e = 41994.0, \alpha^e = 8570395.0$)
PARCO DEI NAVIGLI	0.03	($\xi^e = 4812.0, \omega^e = 31863.0, \alpha^e = 17153785.0$)
PARCO DELLE ABBAZIE	0.03	($\xi^e = 4673.0, \omega^e = 38000.0, \alpha^e = 3333409.0$)
PARCO FORLANINI - CAVRIANO	0.17	($\xi^e = 4384.0, \omega^e = 44614.0, \alpha^e = 17873046.0$)
PARCO NORD	0.01	($\xi^e = 4738.0, \omega^e = 37073.0, \alpha^e = 7512152.0$)
PARCO SEMPIONE		($\xi^e = 4115.0, \omega^e = 133096.0, \alpha^e = 9147581.0$)
PORTA GARIBALDI - PORTA NUOVA	0.42	($\xi^e = 4300.0, \omega^e = 100956.0, \alpha^e = 3771577.0$)
PORTA GENOVA	1.09	($\xi^e = 4294.0, \omega^e = 64595.0, \alpha^e = 17819238.0$)
PORTA MAGENTA	1.92	($\xi^e = 4067.0, \omega^e = 99073.0, \alpha^e = 28071059.0$)
PORTA TICINESE - CONCA DEL NAVIGLIO	1.43	($\xi^e = 4315.0, \omega^e = 84059.0, \alpha^e = 1935245.0$)
PORTA TICINESE - CONCHETTA	1.21	($\xi^e = 4354.0, \omega^e = 72701.0, \alpha^e = 3090050.0$)
PORTA VIGENTINA - PORTA LODOVICA	0.97	($\xi^e = 4358.0, \omega^e = 76284.0, \alpha^e = 13857703.0$)
PORTELLO	0.65	($\xi^e = 4340.0, \omega^e = 83307.0, \alpha^e = 7331109.0$)
PTA ROMANA	1.18	($\xi^e = 4388.0, \omega^e = 72881.0, \alpha^e = 6415644.0$)

Q.RE GALLARATESE - Q.RE SAN LEONARDO - LAMPUGNANO	2.33	($\xi^e = 4559.0, \omega^e = 52682.0, \alpha^e = 11414354.0$)
QT 8	0.29	($\xi^e = 4433.0, \omega^e = 69294.0, \alpha^e = 5347234.0$)
QUARTO CAGNINO	0.71	($\xi^e = 4733.0, \omega^e = 28895.0, \alpha^e = 10345851.0$)
QUARTO OGGIARO - VIALBA - MUSOCCO	2.25	($\xi^e = 4831.0, \omega^e = 27022.0, \alpha^e = 17570863.0$)
QUINTO ROMANO	0.34	($\xi^e = 4686.0, \omega^e = 30118.0, \alpha^e = 23352035.0$)
QUINTOSOLE	0.07	($\xi^e = 4654.0, \omega^e = 37735.0, \alpha^e = 13688732.0$)
ROGOREDO - SANTA GIULIA	0.84	($\xi^e = 4680.0, \omega^e = 36206.0, \alpha^e = 6459355.0$)
RONCHETTO DELLE RANE	0.05	($\xi^e = 4812.0, \omega^e = 31863.0, \alpha^e = 17153785.0$)
RONCHETTO SUL NAVIGLIO - Q.RE LODOVICO IL MORO	1.04	($\xi^e = 4545.0, \omega^e = 50357.0, \alpha^e = 24083240.0$)
ROSERIO	0.02	($\xi^e = 4921.0, \omega^e = 21801.0, \alpha^e = 20748440.0$)
SAN SIRO	1.91	($\xi^e = 4457.0, \omega^e = 60721.0, \alpha^e = 68780786.0$)
SARPI	2.22	($\xi^e = 4301.0, \omega^e = 101351.0, \alpha^e = 16487871.0$)
SCALO ROMANA	0.87	($\xi^e = 4502.0, \omega^e = 50896.0, \alpha^e = 8537449.0$)
STADERA - CHIESA ROSSA - Q.RE TORRETTA - CONCA FALLATA	2.11	($\xi^e = 4654.0, \omega^e = 37735.0, \alpha^e = 13688732.0$)
STADIO - IPPODROMI	0.94	($\xi^e = 4613.0, \omega^e = 44914.0, \alpha^e = 8753466.0$)
STAZIONE CENTRALE - PONTE SEVESO	1.32	($\xi^e = 4381.0, \omega^e = 86030.0, \alpha^e = 4529038.0$)
STEPHENSON	0.01	($\xi^e = 4831.0, \omega^e = 27022.0, \alpha^e = 17570863.0$)
TALIEDO - MORSENCHIO - Q.RE FORLANINI	1.39	($\xi^e = 4434.0, \omega^e = 44601.0, \alpha^e = 11933917.0$)
TIBALDI	0.82	($\xi^e = 4575.0, \omega^e = 43472.0, \alpha^e = 37942888.0$)
TRE TORRI	0.17	($\xi^e = 3971.0, \omega^e = 118886.0, \alpha^e = 9425571.0$)
TRENNO	0.28	($\xi^e = 4519.0, \omega^e = 57817.0, \alpha^e = 6749406.0$)
TRIULZO SUPERIORE	0.12	($\xi^e = 4680.0, \omega^e = 36206.0, \alpha^e = 6459355.0$)
UMBRIA - MOLISE - CALVAIRATE	1.63	($\xi^e = 4520.0, \omega^e = 53836.0, \alpha^e = 7109952.0$)
VIGENTINO - Q.RE FATIMA	1.1	($\xi^e = 4591.0, \omega^e = 41300.0, \alpha^e = 3989369.0$)
VILLAPIZZONE - CAGNOLA - BOLDINASCO	2.97	($\xi^e = 4587.0, \omega^e = 54246.0, \alpha^e = 2136750.0$)
XXII MARZO	2.26	($\xi^e = 4225.0, \omega^e = 72834.0, \alpha^e = 16564034.0$)
CITY	100.0	($\xi^e = 3861.444324975754, \omega^e = 66345.3891270837, \alpha^e = 9187324.879473206$)

A.2 GeoAgents Generated Attributes

Class	Attribute Name	Distribution
`Resident`	Income	See Appendix A.1
`Resident`	Car	$\begin{cases} False \text{ if } (\frac{\text{income}}{30000} + \mathcal{N}(0, 0.1)) < 0.5 \\ True \text{ otherwise} \end{cases}$
`Resident`	Resting start time	$\mathcal{N}(21, 2)$
`Resident`	Resting end time	$\mathcal{N}(7.5, 0.83)$
`Worker`	Work start time	$\mathcal{N}(8, 2) > 5$
`Worker`	Work end time	$\mathcal{N}(18, 2) < 21$
`Worker`	Self defence	$\mathcal{U}(0, 1)$
`Worker`	Crime attractiveness	ϕ(income) where ϕ is the cdf of the city income distribution
`Criminal`	Crime motivation	$1 - \phi$(income) where ϕ is the cdf of the city income distribution