The Hidden Connections of Urban Crime: A Network Analysis of Victims, Crime Types, and Locations in Rio de Janeiro

Fernanda C. Ventorim and Vinicius M. Netto

Abstract: Urban crime is a pressing issue in socially unequal regions like Brazil. We investigate the complex relationships between crime, victims, and urban situations in Rio de Janeiro. We analyze the connections between specific types of crime, victim characteristics, and crime locations through complex network analysis. In a large-scale empirical study, we examine 5000 randomly selected crime incidents in Rio between 2007 and 2018. Our analysis reveals a strong association between sex, race, location, and income inequality in the risk of exposure to crime. The results suggest that specific social groups, notably black and brown women, face a higher-than-average vulnerability to particular types of crime in Rio. Our investigation confirms our hypothesis that crime incidents are not random occurrences. Instead, we have discovered meaningful connections between specific types of crime, victim characteristics, and crime locations. These findings indicate a pattern of underlying factors shaping the distribution of crime and vulnerable social groups.

Keywords: complex network analysis; crime types; place-based methods; urban segregation; victims profiles; vulnerable social groups

1. Introduction

Crime is frequently seen as a critical societal challenge, particularly in contexts with significant social disparities. Criminal activities seem to impact citizens indiscriminately, regardless of their location, economic status, or social background [1,2]. As such, cities are particularly susceptible to crime, with increasing per capita incidence rates relative to population size [3]. Many factors contribute to the distribution of crime within urban areas, from socioeconomic conditions that trigger and enable violence to features of the built environment, such as inequalities in residential locations, accessibility, and neighborhood conditions [4–6]. Public security institutions need systematic, evidence-based approaches to address the root causes and distribution of urban crime, aiming to operate more efficiently and promote safer communities, especially for social groups exposed to violent crimes.

Previous research in this field has identified relationships between crime and urban space as functions of spatial characteristics, such as residential density, urban block size, and street accessibility [7]; land uses and architectural features like the presence of street-level commercial activities and visibility at the building-street interface [8,9]; the presence or absence of urban features such as bus stops and parks [10,11]; levels of social appropriation of urban space, such as co-presence and natural surveillance in public space, as well as institutional and police security factors [12,13]. Altogether, these approaches show that crimes do not occur randomly in cities and that different crime types involve specific urban conditions.

Despite the progress made by such empirical approaches, there remains an incomplete understanding of the complex interplay between the key elements of urban crime. Specifically, we refer to the topology that connects (a) victims with their social and economic
characteristics, (b) the types of crimes most likely to be committed against them, and (c) the broader urban conditions and patterns of concentration of specific types of crime. Accordingly, our study addresses the following questions: What are the connections between different types of crimes, the characteristics of the victims, and the locations where crimes occur? We aim to uncover non-random associations between social and locational factors in order to identify which types of crimes victims have been experiencing in urban contexts.

We hypothesize that these associations can reveal distinct patterns of criminality, providing a deeper understanding of the nature of urban crime and the complex interplay between the factors that contribute to its exacerbation. Are there particular social profiles of victims that show higher vulnerability to specific types of crimes? We argue that an uneven distribution of crime implies that certain social groups are at greater risk of experiencing particular types of crime daily. For example, our findings will show that black female victims aged 20–40 living in low-income, spatially segregated areas of Rio de Janeiro are more likely to experience violent crimes than women from other social groups.

We will investigate such crucial issues through a novel approach that has yet to be fully explored in the specialized literature on the urban conditions of crime: complex network analysis. This method will allow us to uncover associations between factors that make up the problem of urban crime. Namely, it will allow us to identify specific characteristics of victims, the types of crime they are more frequently subject to, and the places such crimes more frequently occur. Our method uses measures of centrality in networks and detects “communities” based on similarities in associations between types of crime, victims, and locations, giving us accurate and systemic empirical results. Our study focuses on the city of Rio de Janeiro. We developed a large-scale empirical study involving a sample of 5000 incidents randomly selected from an initial database of 492,305 incidents between 2007 and 2018 obtained from the Public Security Institute of the State of Rio de Janeiro (ISP-RJ). Using network analysis, our aim is to contribute to developing targeted crime prevention strategies.

This paper is organized into seven sections. The first section introduced the research problem and our approach. The second section provides an overview of the state of the art in the field, highlighting some of the main approaches to the relationship between crime and urban space and analyzing some of their limitations. In the third section, we introduce complex network approaches to the study of urban crime. The fourth section contextualizes our approach, bringing general data on crime in Rio de Janeiro. In the fifth section, we develop our method, which involves a metric to evaluate the similarity between criminal incidents, followed by an algorithm for identifying clusters of similar crime incidents based on empirical associations between characteristics of victims, crimes, and locations. In the sixth section, we analyze the results generated for Rio de Janeiro, with a focus on the distribution of incidents and their distances from Rio’s CBD. This distribution analysis allows us to identify patterns of concentration of types of crime. The concluding section discusses the potential contributions and limitations of our approach in relation to the current state of the art on the connections between crime, victims, and urban space.

2. Crime and Urban Space in the Literature

To gain insight into the conditions under which crimes occur, several lines of study have been developed: from theories that investigate criminal behavior [14,15] to theories of crime as a consequence of social disorganization or a deficient social system [16]. Criminology, traditionally focused on offender and victim studies, has evolved in recent decades to include environmental criminology, which emphasizes the impact of the physical and urban environment on crime. Crime affects social groups in different ways in terms of social and spatial dimensions. This implies that different social groups tend to be exposed to different degrees of violence in their daily lives. Becker [17] was one of the first to highlight this relationship, focusing on the criminal’s decision-making process. Cohen and Felson [18] developed the criminal pattern theory, which associates incidents with variables such as location and specific victims. Crimes would be spatially concentrated but changeable and dynamic
over time. The theory of routine activities [19] suggests that crime incidents involve the convergence of three elements: the aggressor, the victim, and the absence of policing—forming a favorable environment for its realization. Finally, opportunity theory [20] deals with crime vectors related to the victim, environment, type, and time when the crime is committed. Each of these theories contributes to our understanding of how crime occurs in specific spatial and social contexts, allowing for the development of targeted crime prevention strategies.

More recently, the spatial pattern analysis of crime has provided insights into the impacts of violence on cities and their populations [16], including qualitative analyses of the perception of violence [21]. In turn, place-based methods in criminology have evolved, focusing increasingly on smaller scales—from neighborhoods to street networks, segments, and intersections. These features contribute to the spatial patterning of crime. Understanding spatial conditions at these fine-grained scales can explain crime concentrations [6], with implications for prevention initiatives and reducing crime incidence. Spatial concentration and crime hotspots have been linked to small-scale spatial features [4,22], empirically grasped through high-resolution analysis. Additional research areas focus on paths that include the analysis of neighborhood-level criminal opportunity [23], crime concentration and place diversity [24], and crime location choice. Studies also aim to predict crime rates, spatiotemporal variation [25], crime diversity, and environmental design [26,27] in different regional contexts. In the context of our research, some studies in Brazil have focused on how architectural elements like walls and fences can influence crime propensity [9] and the perception of insecurity in public spaces [28].

Methodologically, many of these approaches have utilized geolocation techniques to track incidents of various types of crimes, supplemented by qualitative and quantitative analysis of urban environments. Aligning with our objectives, research has examined the relationship between the attributes of potential crime locations and offenders’ choices [29,30]. Some empirical studies highlight differences in the data used and divergences in analyzing the relationship between incidents and urban factors. For example, analyses include simulations to track the evolution of gang rivalry in Los Angeles [12], and studies on the behavior and spatial locations of offending gang members [31]. Recent, more sophisticated spatial approaches have explored how the topological structure of street networks and their characteristics influence crime distribution [7,32]. However, these analyses have rarely included systemic connections to victim types.

In our research, we suggest that an approach based on analyzing similarity networks connecting types of crimes, victims, and locations is a promising way to fill this gap. Mathematically and computationally driven, network analysis can be used to search for and identify complex associations between a plethora of factors that constitute the conditions and features of crime in large empirical samples, allowing for a precise examination of connections between them. By representing these connections through networks, we can recognize association patterns and gain a deeper understanding of crime dynamics. For a better picture of network analysis, let us examine the approach from its fundamental principles to some of its practical applications.

3. Complex Network Approaches and the Study of Urban Crime

Complex network approaches refer to an area of study that involves modeling phenomena by analyzing entities and their relationships. Networks are useful abstraction tools that can be applied to large databases to generate statistically reliable results. Complex networks have found applications in various fields, such as the dissemination of fake news on social media (social networks), the causes and behaviors of epidemics like COVID-19 (biological networks), and the functioning of the internet (technology networks).

In this study, investigating patterns of association between crime types, victims, and locations in a vast number of cases would be nearly impossible without the computational resources to exhaustively verify similarities between crime incidents. While complex networks are not widely explored in urban crime studies, they originate in classical quantitative
methods such as graph theory. The representation of society and the natural environment through the model of complex networks has been present since Euler’s studies in the 18th century and continued with works focused on spatial urban patterns by Alexander [33], social structures by Freeman [34] and others, and more recent studies on biological structures, such as the work by Girvan and Newman [35]. Highlighting its topological nature, Boccaletti et al. [36] (p. 180) argued that “graph theory is the natural framework for the exact mathematical treatment of complex networks and, formally, a complex network can be represented as a graph”. In turn, Szwarccliter et al. [37] defined a graph by the following:

\[ G = (V, E), \]

given the following:

- \( G = \) denotes the graph composed of a finite nonempty set \( V \) and a set \( E \) of unordered pairs of distinct elements of \( V \);
- \( V = \) denotes the vertices of the graph \( G \);
- \( E = \) denotes the edges of the graph \( G \).

Graphs can be represented in various ways, with three common types identified by Boccaletti et al. [36] (p. 180): undirected, directed, and weighted. In undirected graphs, the edges appear when two vertices have a relationship between them. In directed graphs, vertices are connected by arrows that indicate the direction of each relationship. Weighted graphs are represented by the thickness of the relationships between the vertices, indicating their weights. While several studies have investigated the topological characteristics of networks [38,39], the heterogeneity of relationships between vertices (edges) has increasingly become a topic of interest. In many real-world networks, such as social networks in workplaces or digital information diffusion networks, relationships are associated with weights that differentiate them in terms of strength, intensity, or capacity [40–42].

A primary objective of complex networks is to uncover patterns within physical, biological, and social phenomena. Mathematical algorithms play a crucial role in solving this. Girvan and Newman [35] developed an algorithmic method to grasp such structures as communities in unweighted networks, seeking to identify more densely connected vertices leading to network clustering and strong similarities between vertices and relationship patterns. Blondel et al. [43] proposed an algorithm to detect communities in graphs by evaluating the extent to which vertices of a community are connected compared to a random network. More recently, Cazabet and Rossetti [24] (p. 2) provided a concise definition of a community within a graph as a cluster of vertices with similar topological characteristics: “A (static) community in a graph \( G = (V, E) \) is (i) a cluster (i.e., a set) of nodes \( C \subseteq V \), (ii) having relevant topological characteristics as defined by a community detection algorithm”.

We briefly look at the complex network approach and some basic concepts and applications. We can now relate this topological approach to the associations between crime events, victims, and situational conditions. Urban crime is a multifaceted phenomenon that involves various elements, such as the personal, social, and behavioral characteristics of both victims and criminals, as well as temporal and spatial conditions that interact in potentially complex ways, i.e., not in an entirely predictable and linear fashion. In addition to this, high crime rates and the resulting volume of criminal data make the complex network’s method particularly attractive. In this study, we treat criminal incidents as vertices of a graph and the relations between them as edges. Stronger relationships will indicate greater similarity between the criminal incidents.

To understand the associations between types of crime, victims, and urban situations in Rio de Janeiro, we will apply a mathematical algorithm that can identify communities within the graph representing criminal incidents and their features. Clustering criminal incidents in communities will reveal their similarities. We hope to identify patterns and relationships that may not be immediately apparent through traditional analysis. To do so, let us first understand the nature of crimes in the specific context at hand.
4. The Empirical Context: Crime Rates in Rio de Janeiro

Rio de Janeiro is a city that has gained international recognition for several reasons, including its problems with crime. Since the 1960s, high murder rates have been prevalent, particularly in the favelas or informal settlements within the city. These problems have also extended to the south zone, an affluent area by the sea, to the vast north and west zones, and into Baixada Fluminense, a conurbation of cities in the metropolitan area. Recently, militias have taken control of large territories in North and West Rio and have implemented surveillance practices on commercial and social activities, adding to the challenge of organized crime in the city. Despite such ongoing issues, data show that the general incidence of violent crimes has decreased in the period under study (2007–2018), possibly in connection with public policies. According to the Atlas of Violence 2019, there has been a reduction in the number of murders in the city since 1991, when the historical series of crimes began to be recorded [44]. Moreover, this type of crime in the state of Rio de Janeiro decreased by 19.3% between 2018 and 2019. Theft incidents reached an annual average of approximately 18,000 incidents between 2010 and 2015 and have declined since then. Despite these declines in the last few years, other types of crime have increased, such as robbery, which surpassed 210,000 incidents reported in 2018 (Figure 1a).

Another piece of evidence to be highlighted is the similarity in the evolution of three types of crime: threat, aggravated assault, and vehicular assault/manslaughter. These three crime types showed slight increases between 2007 and 2013, followed by decreases until 2017. Other types of crime remained stable over the period: murder, rape, missing persons, and attempted murder, each with around 10,000 incidences per year. Finding connections between crime factors in this staggering volume involves penetrating the plethora of potential relations between cases, seeking to identify similarities between them.

5. Method
5.1. Part I: Finding Similarities between Crime Incidents

Our analysis uses data provided by the Public Security Institute of the State of Rio de Janeiro (ISP-RJ). These data were extracted from the records of criminal incident reports between 2007 and 2018 in the city of Rio de Janeiro, the state’s capital. The period was chosen based on the availability and completeness of data. It is important to point out the risk of biases due to the underreporting of crimes. Underreporting might arise from several factors, including fears of retaliation, distrust of law enforcement, and associated cultural or social stigmas. Since under-reported incidents are not captured in the dataset, the analysis might not accurately represent the true extent of criminal activity. This could lead to biases in understanding crime patterns and the effectiveness of prevention strategies. Among the types of crimes available in these records, we will focus on crimes committed against the person, following the ISP-RJ classification. We attempted to translate the terminology used

![Figure 1. (a) Historical series of incidents concerning the nine different types of crime analyzed. (b) Histogram of the distribution of the degree of similarity between criminal incidents.](image)
in the ISP-RJ crime taxonomy to approximate the terminology used in the international specialized literature while preserving the original taxonomic structure used in the Brazilian context. This might imply differences in terminology and scope regarding other contexts:

(a) Assaulting: assault, aggravated assault, domestic violence, abusive behavior, threat, and sexual assault (rape and other forms of sexual violence);
(b) Missing persons: potentially involving unsolved kidnapping and murder cases;
(c) Murders: murder, attempted murder, death by the intervention of a State agent (e.g., police action);
(d) Traffic injuries: vehicular assault/manslaughter.

In addition to the typology of crimes, the database includes data on (i) the victim’s characteristics (sex, age, and race), and (ii) the location of the incident (neighborhood) and time of the incident. Due to the lack of locational data on some types of crimes, we could not include property-related crimes (this term and category are deployed in the ISP-RJ crime taxonomy that constitutes our database), such as theft and robbery, for example. Furthermore, original locational data may include the streets where offenses occurred, but without a precise position, rendering the data unreliable for our purposes. Using neighborhood-level data instead of more granular spatial data imposes limitations on analytic precision. Neighborhood-level data tend to aggregate information, which can mask variations and specific crime hotspots within neighborhoods. While this approach provides a broader spatial context, it might overlook fine-grained spatial patterns critical to understanding crime dynamics. On the other hand, higher-resolution data, such as street or intersection-level information, can offer deeper insights into the spatial distribution of crimes and the localized socio-environmental factors influencing criminal behavior. While previous research has emphasized the significance of micro-level mapping to enhance crime estimates [45], our dataset requires an examination of neighborhoods as the spatial unity of analysis. Naturally, there are other potential causal factors in the relationships between these variables and their implications for committed crimes, such as the presence of militias [46], the effects of public security policies, and practices of policing in different areas of the city. However, these factors are notoriously difficult regarding available information and data and could not be considered in this analysis.

Note that the network analysis of a large number of criminal incidents—a total of 492,305 incidents—demands high computational capacity, which poses difficulties for the empirical examination. One way to overcome this limitation and maintain the representativeness of the set of cases to be examined demands the definition of a sample—a data subset large enough to offer statistical robustness. We solved this issue through a trade-off between a desirable sample size and computational cost. The generation of this sufficient and feasible sample was made from a random selection of a set of 5000 incidents, capable of offering a statistical confidence level above 95%, via an algorithm implemented in Python 3.7.

5.2. Constructing Degrees of Similarity between Criminal Incidents

The next step involves checking the degree of similarity between every pair of criminal incidents. Low degrees of similarity are less interesting than cases with high similarity to detect patterns in the data. In our analysis, the rationale for the weighting of different variables in the similarity metric was based on the importance of each variable in characterizing the similarity between criminal incidents (Table 1). We constructed a metric that integrated variables related to the crime itself (type of crime and time of day) and variables related to the victims (sex, age, and race).

Weights were assigned as follows: the similarity between two identical types of crime was set at 100%, while crimes from the same group were weighted at 50%, reflecting their shared group characteristics despite not being identical. This rationale was similarly applied to other variables. For the time of day, identical times received a 100% similarity score, while different times were weighted lower, such as 33% for neighboring time slots. Sex similarity was set at 100% if the same and 0% if different. Age similarity was calculated as 100 minus
the absolute age difference, and race similarity was weighted at 100% for the same race and varied for different races. This weighting approach enabled the incorporation of various factors when quantifying the degree of similarity between incidents (see Equation (2)). In this experimental design, the spatial variable was not included in the community detection method. Instead, location patterns emerged a posteriori from the comparison between the algorithm’s similarity results and the incident locations. The same approach was applied to the income variable.

Table 1. Metrics of the degree of similarity between incident variables.

<table>
<thead>
<tr>
<th>Theme</th>
<th>Variables</th>
<th>Incident 1</th>
<th>Incident 2</th>
<th>Similarity Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of Crime</td>
<td>Same crimes</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Crimes in the same group</td>
<td>50%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Crimes in different groups</td>
<td>0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crime</td>
<td>Same time</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Morning</td>
<td>33%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Afternoon</td>
<td>33%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Night</td>
<td>33%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dawn</td>
<td>33%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Different time</td>
<td>0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>Same sex</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>Different sex</td>
<td>0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>Examples: 89</td>
<td>43%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>96%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td>Same race</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>White</td>
<td>33%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Brown</td>
<td>33%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Different race</td>
<td>0%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From these percentages, which represent the degree of similarity between variables in each category, we propose calculating the total degree of similarity between two criminal incidents using the following equation:

\[
DS = TC\% + T\% + S\% + (100 - |Ai - Aj|)\% + R\%,
\]

(2)

where:

- \(DS\) = Degree of similarity between two incidents.
- \(TC\) = Similarity of the variable: type of crime.
- \(T\) = Variable similarity: time.
- \(S\) = Similarity of the variable: sex.
- \(Ai\) = Age of victim \(i\).
- \(Aj\) = Age of victim \(j\).
- \(R\) = Similarity of the variable: race.

We tested this measure by checking the distribution of incidents based on degrees of similarity. We did this by examining the histogram of frequencies of incidents (y-axis in the graph in Figure 1b) for different degrees of similarity (x-axis in the graph in Figure 1b). The histogram shows that most incidents presented degrees of similarity between 40% and 70%. Given that low degrees of similarity are less interesting for detecting patterns in the data, we only considered incidents with degrees of similarity above 70% as a threshold to recognize clusters of similar incidents.
5.3. Part II: Exploring Complex Networks to Capture Associations

Our method involves assessing the existence of similarities between incidents by computing relationships between the information that makes up each criminal incident record. To this end, we developed an algorithm to map associations onto the graph representing the incidents. We shall now assess their degrees of similarity. Degrees of similarity between incidents will be used to generate the similarity network between elements of urban crime according to a relationship model structured according to the variables. The analysis of similarities allows the automatic detection of clusters or communities [38].

Constructing a Network of Similar Criminal Incidents

The network of similar criminal incidents was constructed using graph theory, where each incident was treated as a vertex of the graph, and the degree of similarity between them was represented as an edge. The resulting weighted graph demonstrates the relationship between vertices by the thicknesses of the edges between them [36]. Peixoto [47]’s algorithm was used to group entities based on their similarities.

With a sample of 5000 incidents, the method analyzes the numbers of criminal incidents to identify the most frequent associations between types of crime, victims, and locations using the weighted complex network model. We used graph-tool 2.7 [47], a Python library designed for the statistical analysis of graphs, to recognize and generate these connections in the form of graphs. In addition to providing high-quality graph visualizations, this tool also includes algorithms for graph traversal, shortest paths, network flow, clustering, and statistical analysis.

Communities were formed through a branching of incidents, creating a hierarchy with different levels. Figure 2 illustrates the result of a subset of the sample with 1000 vertices and their connections (stylized as curvilinear connections). Each point along the outer circle represents a criminal incident, and each set of points forming a line of the same color is identified as a community, i.e., incidents with similar topological characteristics. The graph of straight lines and gray vertices (little squares) inside the circle illustrates the hierarchy of branches in communities, emerging at the circle’s edge. The colors in the stylized graph represent the intensity of the connections within the hierarchical graph of branches across different communities. The intensity ranges from yellow, indicating weaker connections, to purple and black, which represent stronger connections. In addition to color, the widths of the flows also indicate the strength and intensity of the connections between different clusters of incidents. Thicker flows denote higher degrees of similarity or stronger relationships between incidents, while thinner flows indicate weaker connections. Together, these visual elements provide a comprehensive view of the hierarchical structure and connectivity within the network of criminal incidents.

The algorithm identified a total of 96 communities within the 5000 criminal incidents, with an average of 52 per community. While this level of detail provided analytical precision, it generated numerous communities that were highly internally similar but specific to particular incidents. A smaller number of communities facilitates the identification of similar relationships in a wider range of incidents. In this sense, the community analysis method allows for exploring different arrangements of community numbers from different degrees of similarity. We did so by exploring varying degrees of similarity. We tested five hierarchical levels, with 52, 29, 16, 8, and 3 communities. After analyzing these hierarchies, we determined that the structure with the best balance between the degree of internal similarity in the communities and the number of communities to be internally assessed was 16 communities. These hierarchical clusters combined the specificities of each crime incident with reach in potentially similar topological relationships.
Using the algorithm by Peixoto [47], we identified groupings of similar incidents related to crime and victim variables. Once connection patterns are found, we may move on to the final step of the method: identifying spatial patterns in the distribution of the 16 clusters in Rio’s neighborhoods.

6. Results: Patterns of Criminal Similarity

We analyzed connections between the clusters of incidents and their spatial locations to determine whether relationships exist between groups of similar crimes and their spatial distribution in Rio’s neighborhoods. Our hypothesis points to patterns of association between the components of the typology of crime as defined by Brazilian law and the characteristics of victims. Furthermore, victims are potentially unevenly distributed according to the distance from the city business center (CBD). Note that a heterogeneous distribution of crimes and victims can indicate that specific social groups are more susceptible to certain types of crime.

A broad analysis of the types of crime and victims in Rio in the period from 2007 to 2018 shows that females make up the majority of crime victims, accounting for 56.5% of the total. It must be noted that females make up 53.2% of the total population in Rio, according to the last available census. The victims’ races follow a similar pattern, with 53% of the victims being white, 38% being brown, and 10% being black, relatively close to the population distribution (51.2% white, 36.5% brown, and 11.5% black). In terms of age group,
victims between the ages of 20 and 40 are the most frequent, corresponding to the highest percentage age group in the city’s population.

Further analysis reveals that women are the most susceptible to assaults in Rio, including aggravated assault, domestic violence, and threat, accounting for 71.7% of the annual average of the total number of cases in the period. Assaults and sexual assault (rape and other forms of sexual violence) appear in 9 of the 16 detected communities. Black and brown victims account for 51.7% of these crimes, and 79.8% of all victims are aged between 20 and 40.

In stark contrast, male victims are more likely to experience vehicular assault or manslaughter (63.5% of cases in the period), appearing in 7 out of 16 communities. Additionally, crimes such as murder, attempted murder, death by the intervention of a state agent, or missing persons predominantly affect men (83.8% of victims). Race and age are also significant factors. More than two-thirds of the male victims were black or brown (68.3%). Adult men between the ages of 20 and 40 were the most frequent victims (61.9% of cases).

Our method identified 16 communities within the sample of 5000 criminal incidents. As expected, incidents within each community share similar characteristics, which bring them closer within the topology. Table 2 presents the main results for each identified community:

Table 2. Summary of the 16 communities identified by the algorithm according to variables related to the victim, crime and location of the occurrence.

<table>
<thead>
<tr>
<th>Community</th>
<th>Incidents</th>
<th>Sex</th>
<th>Race</th>
<th>Age</th>
<th>Type of Crime</th>
<th>Time</th>
<th>Neighborhood</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>497</td>
<td>100% (M)</td>
<td>100% (W)</td>
<td>23.1%—30–40</td>
<td>Aggravated Assault</td>
<td>47.5% (A)</td>
<td>Lagoa (SZ)</td>
</tr>
<tr>
<td>2</td>
<td>336</td>
<td>100% (F)</td>
<td>60.7% (B)</td>
<td>33.9%—20–30</td>
<td>Aggravated Assault</td>
<td>53.6% (D)</td>
<td>Pavuna (NZ)</td>
</tr>
<tr>
<td>3</td>
<td>115</td>
<td>100% (F)</td>
<td>51.3% (W)</td>
<td>28.7%—10–20</td>
<td>Vehicular Assault</td>
<td>60% (N)</td>
<td>Lagoa (SZ)</td>
</tr>
<tr>
<td>4</td>
<td>487</td>
<td>100% (F)</td>
<td>99.4% (Br)</td>
<td>32.4%—20–30</td>
<td>Aggravated Assault</td>
<td>62.8% (N)</td>
<td>Ricardo de Albuquerque (NZ)</td>
</tr>
<tr>
<td>5</td>
<td>165</td>
<td>100% (F)</td>
<td>99.4% (W)</td>
<td>21.8%—30–40</td>
<td>Traffic injury</td>
<td>50.3% (A)</td>
<td>Bonsucesso (NZ)</td>
</tr>
<tr>
<td>6</td>
<td>415</td>
<td>100% (M)</td>
<td>51.8% (W)</td>
<td>23.1%—30–40</td>
<td>Vehicular Assault</td>
<td>66% (N)</td>
<td>Lagoa (SZ)</td>
</tr>
<tr>
<td>7</td>
<td>239</td>
<td>100% (M)</td>
<td>47.3% (Br)</td>
<td>25.1%—20–30</td>
<td>Aggravated Assault</td>
<td>52.7% (N)</td>
<td>Lagoa (SZ)</td>
</tr>
<tr>
<td>8</td>
<td>792</td>
<td>100% (F)</td>
<td>100% (W)</td>
<td>28.2%—30–40</td>
<td>Threat</td>
<td>42% (A)</td>
<td>Santa Cruz (WZ)</td>
</tr>
<tr>
<td>9</td>
<td>770</td>
<td>100% (F)</td>
<td>56% (W)</td>
<td>30%—30–40</td>
<td>Aggravated Assault</td>
<td>72.7% (N)</td>
<td>Ricardo de Albuquerque (NZ)</td>
</tr>
<tr>
<td>10</td>
<td>25</td>
<td>100% (F)</td>
<td>80% (Br)</td>
<td>48%—10–20</td>
<td>Rape</td>
<td>36% (N)</td>
<td>Santa Cruz (WZ)</td>
</tr>
<tr>
<td>11</td>
<td>262</td>
<td>100% (M)</td>
<td>100% (W)</td>
<td>34.7%—20–30</td>
<td>Vehicular Assault</td>
<td>44.3% (A)</td>
<td>Parada de Lucas (NZ)</td>
</tr>
</tbody>
</table>

Table 2.
Table 2. Cont.

<table>
<thead>
<tr>
<th>Community</th>
<th>Incidents</th>
<th>Victim Profile</th>
<th>Crime</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Sex 1</td>
<td>Race 2</td>
<td>Type of Crime</td>
</tr>
<tr>
<td>12</td>
<td>156</td>
<td>100% (M)</td>
<td>100% (Br)</td>
<td>39.7%—20–30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>55.9%—Aggravated assault</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>261</td>
<td>100% (M)</td>
<td>100% (Br)</td>
<td>26.8%—30–40</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>164</td>
<td>100% (M)</td>
<td>66.5% (B)</td>
<td>31.1%—20–30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>31.7% (Br)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>134</td>
<td>100% (F)</td>
<td>70.9% (Br)</td>
<td>20.9%—30–40</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>28.4% (B)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>182</td>
<td>100% (M)</td>
<td>57.7% (Br)</td>
<td>33%—20–30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>40.7% (B)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24.2%—Missing persons</td>
</tr>
</tbody>
</table>

1 (F): Female; (M): Male. 2 (W): White; (B): Black; (Br): Brown. 3 (M): Morning; (A): Afternoon; (N): Night; (D): Dawn. 4 Number of incidents in relation to total neighborhood population (neighborhoods with the largest rates).

Among the 16 clusters, we observed significant social and spatial patterns. For instance, eight communities predominantly feature crimes such as aggravated assault, threats, and rape. Female victims aged between 30 and 50, mostly black or brown, are prevalent in these communities, constituting 27.36% of the sample. Most victims in these communities reside in low-income neighborhoods, with per capita monthly incomes ranging from USD 367 (BRL 644.82) to USD 774 (BRL 1359.92). Community 2 stands out as an example of such an association pattern.

6.1. Community 2

In Community 2, a total of 336 criminal incidents were identified, exclusively targeting black and brown women. Among the victims in this cluster, 60.7% are black, and 39% are brown. The most prevalent type of crime is aggravated assault, accounting for 58% of the total number of incidents, followed by threats at 40.8% (Figure 3).

The main victims are women aged between 20 and 30 years (33.9% of the total number of victims in this cluster) and between 30 and 40 years (29.5%). Victims in adjacent ranges—10 to 20 years old and 40 to 50 years old—account for 14% of victims. Most incidents occur at dawn (53.6%). Other incidents are divided into two time slots: 25.9% in the afternoon and 20.5% in the morning.

The 336 incidents are mostly concentrated in the West Zone of Rio de Janeiro, with 25 of them in the neighborhood of Santa Cruz, 19 in Campo Grande, and 18 in Bangu, which are major local centralities in that large municipal area. When the number of incidents is relativized by the total population of each neighborhood, the neighborhoods with the highest incidence are Santa Cruz (rate of 229.45 incidents/population of the neighborhood multiplied by 100,000, the annual average in the period), Pavuna (119.54), Ricardo de Albuquerque (56.69), and Rocha Miranda (55.91) (Figure 3).

In turn, the spatial distribution of violent crimes against black and brown women is relatively heterogeneous, with peaks in low-income, spatially segregated areas. Three of them (Pavuna, Ricardo de Albuquerque, and Rocha Miranda) are approximately 20 km from the CBD. Santa Cruz is the most distant neighborhood and has the highest rate of incidents, which has direct implications with risks to its black and brown female population.
Figure 3. Community 2. Distribution of variables related to the type of crime, victim profile (sex, race, and age), and time of the incident (top). The majority of incidents registered by women are intentional bodily harm or threats and occur mostly at dawn. These victims are mostly black or mixed-race and between 20 and 40 years old. The (bottom) map displays the spatial distribution of such incidents in Rio. Vertical bars represent the relative incident rate in each neighborhood, calculated as the annual average of incidents during the period divided by the population in the neighborhood and multiplied by 100,000. Colors indicate the average per capita monthly income range in each neighborhood, while the concentric circles represent distances to the CBD’s centroid.

Regarding other clusters of crimes and victims, our model identified seven communities in which vehicular assault/manslaughter stands out. In most cases, such victims are white men aged between 20 and 40 living in medium and high-income neighborhoods, where per capita monthly income ranges from USD 1448 (BRL 2544.14) to USD 2896 (BRL 5088.27). In fact, such an association between crime, victim, and location has emerged as a distinct community.

6.2. Community 11

This cluster recorded 262 criminal incidents, predominantly targeting white men, mostly aged between 20 and 40 (59.1% of cases in the cluster). Other age groups include victims between 40 and 50 years old (13.4%) and between 10 and 20 years old (8.4%). The vast majority of these incidents (82.1%) involve vehicular assault or manslaughter. Missing persons and attempted murder also frequently occur in this community, accounting for 7.3% and 5% of crimes, respectively. Crime incidents were most frequent in the afternoon (44.3%), followed by the morning (34.7%) and dawn (21%) (Figure 4).
Figure 4. Community 11. Distribution of variables for the type of crime, characteristics of the victim (sex, race, and age), and time of incident (top). Incidents involve white men aged between 20 and 40, who are mostly victims of bodily injury in traffic, particularly in the afternoon. The (bottom) map illustrates the spatial distribution of incidents in Rio. Vertical bars indicate the relative incident rate in each neighborhood, calculated as the annual average number of incidents during the period divided by the population of the neighborhood and multiplied by 100,000. The colors represent the average per capita monthly income range in each neighborhood. Concentric circles denote distances to the CBD’s centroid.

The spatial analysis of the 262 incidents in Community 11 reveals the neighborhoods with the highest rates: Tijuca and Campo Grande (14 incidents) and Copacabana and Barra da Tijuca (13 incidents). Most of these neighborhoods are known to be predominantly middle and upper-class. To further analyze the spatial distribution of incidents and account for the population size in each neighborhood, Lagoa—a high-income neighborhood in South Rio located less than 10 km from the CBD—has the highest rate of incidents, with 489 incidents per 100,000 population annually. This rate is four times greater than those of the following neighborhoods in this community: Parada de Lucas (119 cases), Penha (63.24), and Pavuna (54.34), which are located in North and West Rio, between 10 and 25 km away from the CBD; these neighborhoods are ‘cut across’ by express motorways and fall into the lowest range of per capita income. Such results underscore the heterogeneous character of the distribution of traffic-related crime, both spatially and in terms of income (see Figure 4).

7. Discussion: Associating Crimes, Victims, and Neighborhoods in Rio

General findings indicate that crime types have certain patterns regarding sex, race, and age in Rio de Janeiro:

- Women were victims of 56.6% of all crimes in the city during the period—a figure slightly higher than their proportion in the overall population (53.2%, according to the 2010 census). Most notably, women are victims of 71.7% of assaults, including aggravated assault, domestic violence, abusive behavior, threats, and sexual assault (rape and other forms of sexual violence). Black or brown victims account for 51.7% of these cases. The vast majority of these women are aged between 20 and 40 (79.8%).
Men are victims of 83.8% of murders, a type of crime that disproportionately affects black men, who make up 68.3% of the victims. Notably, black and brown populations represent 47% of the total population in Rio. Murders most frequently affect individuals between the ages of 20 and 40 (61.9% of the victims). Men also account for 63.5% of vehicular assault/manslaughter incidents, with 51% of cases involving the black and brown populations and 54.6% involving victims aged between 20 and 40.

Our method also identified patterns of connections between types of crimes, victims, their income status, and locations. Types of crimes exhibit well-defined social profiles and, to some extent, spatial profiles. We analyzed 16 communities with similar incidents of similar characteristics, which grouped victims into social groups more susceptible to certain types of crime. The spatial distribution of these clusters was analyzed, along with the average income in the neighborhoods where they occurred. Taken together, our results highlight alarming heterogeneities in the incidence of crimes across Rio:

- Communities 2, 4, and 10 are mostly composed of brown or black female victims (27.4% of the sample) subject to assault. These clusters are associated with predominantly low-income neighborhoods in spatially and socially segregated areas, i.e., far from the main employment center in Rio, and characterized by homogeneity in income (see Figure 3).
- In turn, white women victims of assault are grouped in Communities 8 and 9 and have a slightly lower percentage (24.5%). Overall, victims are distributed differently, residing both in poor, spatially segregated areas such as in the West Zone and affluent neighborhoods in South Rio.
- Community 10 clusters severe cases of sexual assault. Although small in number (25 cases), all incidents have rape reported by all female victims, aged mostly between 0 and 20, and brown (88% of cases).
- Vehicular assault/manslaughter is most strongly related to male victims (Communities 7, 11, 12, 15, and 16). Offenders are likely to have purchasing power commensurate with vehicle ownership. Victims are frequently pedestrians and incidents occur more commonly, in absolute numbers, in neighborhoods with medium and high per capita income close to the CBD. When relativized by the size of neighborhood populations, the victims’ incomes are shown as more diverse, distributed in poorer neighborhoods linked to express motorways.
- We also identify a community subject to particularly diverse intentional criminal incidents, with a high rate of missing persons (24.2% of cases in the community) and murder (19.8%). All victims are young adult men, mostly brown (57.7%) and black (40.7%). In these cases, the victims live in predominantly low-income neighborhoods in spatially and socially segregated areas as well.

8. Conclusions

This research aimed to identify patterns in the associations between types of crime, victim profiles related to sex, income, and age, and the locations of incidents based on neighborhoods in Rio. We sought to understand how such associations can make certain social groups more susceptible to specific crimes. To achieve this, we proposed a method in three stages: (i) we developed a metric to analyze the degree of similarity between criminal incidents in terms of types of crime and victims; (ii) we utilized a complex network model to group these incidents into ‘communities’ based on similarities; and (iii) we analyzed the spatial dimension of the incidents, focusing on location characteristics, the distance to the CBD as a proxy for how spatially segregated different types of victims can be, and the average per capita income in the neighborhoods where incidents occurred.

In the experimental design explored, the spatial variable was not included in the community detection method, allowing locational patterns to emerge from the analysis. Once we identified patterns of connection between crimes and victims, we analyzed their distribution in the city, i.e., concentrations in specific neighborhoods and their proximity to the CBD. The same approach was applied to the role of average per capita income
in neighborhoods. This method does not establish causal roles for location or income (a challenge for other methods as well), but it highlights associations between spatial and social characteristics that are useful for understanding and addressing the problem of crime.

The hypothesis guiding this research was that patterns exist in the associations between types of crime, victim profiles, and the locations of criminal incidents based on spatial characteristics such as distance and income. An uneven distribution could make certain social groups more susceptible to specific types of crime, rendering black, female, and young groups more likely to be victims of certain types of violent crimes in spatially segregated and low-income areas of the city, or white male groups more likely to be victims of vehicular assault/manslaughter. Due to computational limitations, we randomly selected a sample of 5000 occurrences from the total database of approximately 500,000 records in Rio between 2007 and 2018.

In sum, the results indicated that victims with specific profiles frequently connect to types of crime and particular places, underscoring distinct territorial patterns of crime in Rio de Janeiro—whether in central or spatially distant areas of the city. Notably, these findings suggest that certain social groups, especially women, and among them, black and brown women, are disproportionately susceptible to specific types of crime in their urban environments. The results also reveal a strong association between sex, race, location, and income inequality in the risks of exposure to urban crime. Methodologically, our approach aims to contribute to systemic quantitative methods that incorporate victim profiles into place-based strategies, typically focused on the spatial and temporal distributions of crime and their correlations with spatial features.

By understanding that crimes do not occur randomly and are influenced by specific urban conditions, urban planners can develop interventions to mitigate conditions most strongly associated with crime. These could include improving urban design to enhance natural surveillance, increasing visibility in public spaces, and creating mixed-use areas that promote social cohesion. Additionally, the identification of crime hotspots and patterns of concentration can help policymakers allocate resources more efficiently, such as deploying more police patrols in high-risk areas or installing better lighting and cameras in crime-prone locations.

Despite this, the empirical results reported here must be considered in light of some limitations. First, using neighborhood-level data rather than more granular spatial data limits the accuracy of our analysis. This data limitation, imposed by the data sources, may have masked more subtle spatial patterns and crime hotspots that more detailed data could reveal. Future research would benefit from incorporating higher-resolution spatial data, when available, to capture a more precise distribution and dynamics of criminal events. Secondly, future research should explore advanced computational techniques and larger samples to further test our findings and uncover more detailed patterns.

Author Contributions: F.C.V.: Research problem definition, Conceptualization, Methodology, Software, Data curation, Writing Original draft preparation, Visualization, Investigation, Writing Reviewing and Editing. V.M.N.: Conceptualization, Writing Original draft preparation, Investigation, Writing Reviewing and Editing. All authors have read and agreed to the published version of the manuscript.

Funding: The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: F.C.V. acknowledges the support of the National Council for Scientific and Technological Development (CNPq), Brazil. V.M.N. acknowledges the support of the Research Centre for Territory, Transports and Environment (CITTA), Faculty of Engineering, University of Porto, Porto, Portugal; the National Council for Scientific and Technological Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq, Brazil), grant number 315086/2020-3, and the Foundation for Research Support (Fundação de Amparo a Pesquisa do Rio de Janeiro | FAPERJ), grant number SEI-260003/015789/2021. The authors thank Urban Science’s editorial team for waiving processing charges.

Data Availability Statement: Data will be made available on request.
Conflicts of Interest: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References


7. Hillier, B.; Sahbaz, O. An Evidence Based Approach to Crime and Urban Design, or, Can We Have Vitality, Sustainability and Security All at Once; Bartlett School of Graduate Studies University College London: London, UK, 2008; pp. 1–28.


11. He, Z.; Deng, M.; Xie, Z.; Wu, L.; Chen, Z.; Pei, T. Discovering the joint influence of urban facilities on crime occurrence using spatial co-location pattern mining. Cities 2020, 99, 102612. [CrossRef]


46. Pope, N. Surviving and dying through the urban frontier: Everyday life, social brokerage and living with militias in Rio de Janeiro’s West Zone. *Urban Stud.* 2023, 60, 343–359. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.