The Contribution of Leiden Algorithm Approach in Spatial Mapping of Violence Against Women

Maria Eduarda T. Souza[1,*], Thabatta M. A. de Araújo[1] and Michel P. da Silva[1]

[1]*Department of Computing, CEFET-MG, Divinópolis, Brazil*

Abstract

Violence against women in Brazil remains a significant issue, even with existing laws like the Maria da Penha Law and the Femicide Law. This study presents a computational approach for conducting a descriptive spatio-temporal analysis of such violence against women, using Divinópolis (Brazil) as a case study and utilizing public data from the Department of Justice and Public Security of Minas Gerais. The Leiden algorithm was used to identify spatially coherent clusters, which helps avoid fragmented groupings that could compromise the analysis. Additionally, this study analyzed temporal variations in standard deviations based on factors such as education level, race/color, and case risk classifications through scatter plots. The results indicate that neighbourhoods with a high incidence of violence show consistently elevated deviations, suggesting that victims come from diverse social and demographic backgrounds. These neighbourhoods tend to be located in areas with higher crime rates and greater socioeconomic vulnerability. The proposed methodology can be used for ongoing monitoring, allowing for the rapid identification of changes in violence patterns and supporting evidence-based preventive public policies.

Keywords

Violence against women, Geospatial Analysis, Clusters, Pattern Detection, Leiden Algorithm

1. Introduction

Violence against women remains a serious human rights violation across various societies, including Brazil. Despite the implementation of laws like the Maria da Penha Law [1] and the Femicide Law [2], rates of femicide and gender-based violence are still alarmingly high [3]. In 2023, Brazil recorded 1,463 femicides, translating to 1.4 women per 100,000 inhabitants—a 1.6% rise from the previous year, according to the Brazilian Public Security Forum.

Research highlights factors contributing to violence against women, including structural inequality, ineffective protective measures, and insufficient institutional support [4, 5]. Studies show that socioeconomic vulnerability correlates with domestic violence, while regions with greater economic instability experience higher rates of gender-based violence [6].

Institutional inadequacies are also critical. Only 8.3% of Brazilian municipalities have specialized Women's Police Stations, limiting access to support from [7]. The justice system often

Proceedings XVII Congress of Latin American Women in Computing 2025, October 27–30, 2025, Valparaíso, Chile.

*Corresponding author.

maria@aluno.cefetmg.br (M. E. T. Souza); thabatta@cefetmg.br (T. M. A. d. Araújo); michel@cefetmg.br (M. P. d. Silva)

0009-0009-4184-2355 (M. E. T. Souza); 0000-0003-3031-9840 (T. M. A. d. Araújo); 0000-0002-1343-1331 (M. P. d. Silva)

underestimates risks, leading to high levels of impunity. Cultural factors can further discourage reporting. This paper hypothesizes that violence against women displays specific spatial patterns shaped by structural factors, and that geospatial analysis of socioeconomic data can inform targeted public policies.

To address the issue, this study applies the Leiden algorithm, an unsupervised graph-based optimization method, to analyze the spatial distribution of cases in a case study of Divinópolis, a Brazilian city located in the state of Minas Gerais, which has one of the highest crime rates related to violence against women. A case study can contribute as a framework for developing strategies to identify and monitor violence against women in other contexts that integrate temporal and spatial analyses with official crime data. Therefore, this work aims to fill this gap by applying the findings to provide a complementary approach to identifying critical areas based on electronic monitoring, as suggested [8, 6].

1.1. Related Work and Feminist Framing

Violence against women comprises a wide range of physical, psychological, sexual, and patrimonial aggressions that occur in a continuum that can culminate in death by homicide, a fact that has been termed femicide [3]. The findings of studies demonstrate that physical and sexual partner violence against women is widespread [6] and common in patriarchal systems, in which women are subjected to the control of men. The causes of these crimes are not attributable to pathological conditions of the offenders, but rather to the desire for possession of women, who are often blamed for not fulfilling the gender roles designated by the culture [3]. What is reinforced by social norms and institutional systems that perpetuate gender inequality and violence [9, 5].

In Brazilian contexts, although violence and femicide are criminalized by law [2] (Feminicide Law) and law [1] (Maria da Penha Law), such acts persist at alarming rates [10]. In this way, overcoming domestic violence remains one of the central causes of the feminist movement, which challenges a patriarchal structure that has historically relegated women to the private sphere of the home [8]. To support the overcoming, several approaches have been employed to identify patterns of violence and to guide targeted actions aimed at combating domestic violence [3, 8, 6]. From this perspective, spatial analysis has been increasingly approached through a feminist, intersectional lens that takes into account factors such as race/ethnicity, education, and economic dependency [6].

In this way, this study builds on geospatial and network-based approaches to map unequal risks across neighbourhoods and to inform prevention strategies. Prior work has applied machine learning approaches based on clustering and density techniques to identify patterns of the geospatial descriptive datasets [4, 11]. This methodological enhancement allows for more precise identification of territorial patterns and can support the design of targeted and evidence-based public policies to combat violence against women.

2. Contextual Background on Violence Against Women in Minas Gerais (Brazil)

This section presents an exploratory analysis of violence against women in Minas Gerais, aiming to identify patterns in documented cases over time and locations using data from the Department of Justice and Public Security. Findings are summarized in Table 1, highlighting variations in crimes against women between 2022 and 2023.

Table 1
Comparison of crimes against women recorded in Minas Gerais between 2022 and 2023.

Crime Type	2022	2023	Variation (%)
Violent Crimes	19,199	16,375	-14.71
Femicide	175	183	+4.57
Domestic Violence	141,441	154,742	+9.40

The results presented in the Table 1 key findings include a 14.71% decrease in female victims of violent crimes between 2022 and 2023, with completed robbery being the most common crime at 59.77%. In contrast, domestic violence victims increased by 9.40% during the same period. Domestic violence has risen, while overall violent crimes have decreased.

In addition to crimes against women by category, the analysis also ranks the top 15 municipalities in the state for the highest incidents of violence against women, as illustrated in Figure 1. Belo Horizonte has the highest absolute number of cases, followed by Juiz de Fora, Uberlândia, and Contagem.

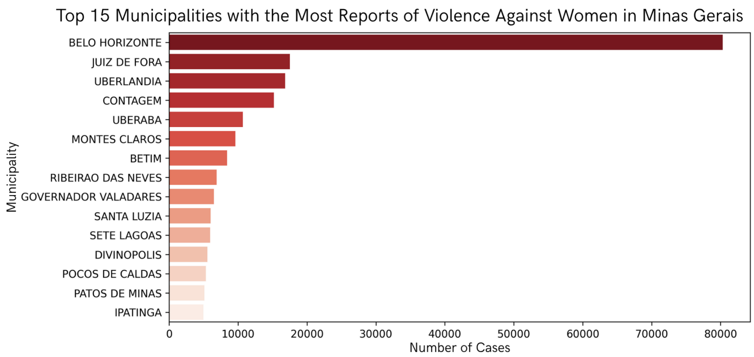


Figure 1: Distribution of records of violence against women in the 15 municipalities with the highest number of occurrences in Minas Gerais [10]

Divinópolis, with a population of 231,091, has a femicide rate of 0.021636 per capita, placing it among the cities with the highest rates in the state. This rate is higher than Ipatinga's (0.0131) and similar to those of Betim (0.02185) and Sete Lagoas (0.02198). Unlike the metropolitan area of Belo Horizonte, Divinópolis serves as a medium-sized urban setting, allowing for a better analysis of local patterns and causes of violence against women, while still reflecting the structural challenges of similarly sized municipalities.

Therefore, the municipality's rate of about 0.002 femicides per inhabitant, similar to critical cases in other areas, highlights its potential to guide targeted public policies and prevention strategies. This study's methodological approach can be applied to other medium and small municipalities, aiding in the creation of evidence-based strategies to combat femicide.

Then, Divinópolis displays distinct internal spatial patterns regarding femicide records, as shown in Figure 2.

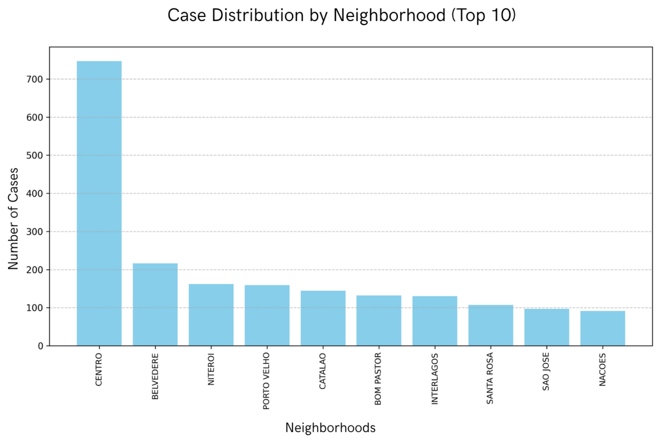


Figure 2: Number of violence against women cases in the 10 neighbourhoods with the highest incidence in Divinópolis/MG [10].

Figure 2 shows the ten neighborhoods the Centro leads in reports, followed by Belvedere, Niterói, and Porto Velho. The concentration in the city center may relate to higher population density and better access to reporting channels. Elevated rates in other neighborhoods reflect social vulnerability and urbanization.

3. Graph-Based Optimization: a Leiden Approach

This section outlines the steps for constructing and analyzing the network of violence against women in Divinópolis, Minas Gerais, Brazil. The study serves as a framework for identifying and monitoring violence against women in other contexts through temporal and spatial analyses with official crime data. All technical information, including source code and datasets, is available in a public repository. [12].

The methodology involved data collection and preprocessing, followed by network modeling based on spatial relationships among neighborhoods. The Leiden algorithm was used for community detection, and a sociodemographic assessment of the clusters was conducted. Notably, the data were descriptive, using aggregated information at the census tract and neighborhood levels instead of precise coordinates or detailed geospatial data. Figure 3 summarizes the workflow by outlining the methodological steps from the approach, detailed in the following sections.

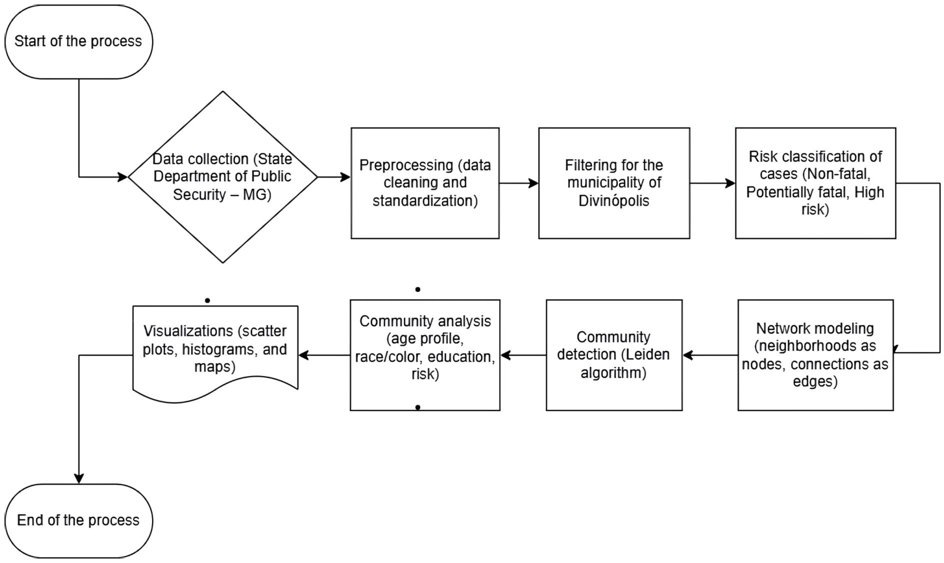


Figure 3: Flowchart of methodological steps adopted in this study.

3.1. Data Access and Description

The data for this study were sourced from the Department of Justice and Public Security of Minas Gerais [10]. All datasets are publicly available and comply with the Brazilian General Data Protection Law, ensuring that no sensitive or personally identifiable information is disclosed. It included detailed records of violence against women, such as location, crime type, and characteristics of victims and aggressors, encompassing all municipalities in the state for 2023 and 2024. The records were obtained via the state Access to Information portal, and the analysis relied exclusively on de-identified, non-sensitive variables reported at the neighbourhood level, in accordance with the LGPD [10].

3.2. Preprocessing

The dataset preprocessing was four steps: (i) **Removal of irrelevant columns** (e.g., internal codes); (ii) **Standardization of text**, including accent removal and uppercase conversion; (iii) **Correction of city name inconsistencies** for geospatial data accuracy; and (iv) **Elimination of invalid records** labeled ”NOT INFORMED” or ”INVALID.”

3.2.1. Territorial Scope

The official urban map of Divinópolis was the primary source for constructing the spatial graph [13]. Manual mapping of connections between adjacent neighborhoods ensured geographic accuracy and adhered to the city’s territorial boundaries. Following the isolation of municipal data, neighborhood names were standardized, and connections were mapped according to the official map, reflecting spatial relationships stored in a CSV file.

3.2.2. Risk Classification of Case

Violence cases against women were classified into three risk levels: *Non-Fatal*, *Potentially Fatal*, and *High Risk of Fatality*. This classification was based on severity factors as crime nature, method, and completion status. Each factor was weighted to determine the final risk category. Along with victim age, race/color, and education level, these criteria facilitated a descriptive analysis by neighborhood, revealing areas with higher concentrations of severe cases and increased vulnerability.

3.2.3. Variable selection rationale

Age, race/colour, education, and case risk classification are analyzed because these variables capture intersectional dimensions repeatedly linked to victimization and barriers to protection, as suggested by [3, 5], and they are consistently available in the administrative records used here. In addition, the age is retained as a control for demographic structure across neighbourhoods and for potential cohort effects. Even when mean-based plots were inconclusive, dispersion measures (standard deviations) revealed within-neighbourhood heterogeneity that is policy-relevant.

3.3. Network Modeling

To analyze the spatial distribution of violence cases against women in Divinópolis, a geographic network was created with neighbourhoods as nodes and geographic proximity as edges. The number of cases defined the nodes, allowing for the application of community detection algorithms to identify clustering patterns. Although initial algorithms like Girvan-Newman, Louvain, and Label Propagation had limitations, the Leiden algorithm was ultimately chosen for its stability and ability to create well-connected communities through node movement and refinement, enhancing cluster quality.4.

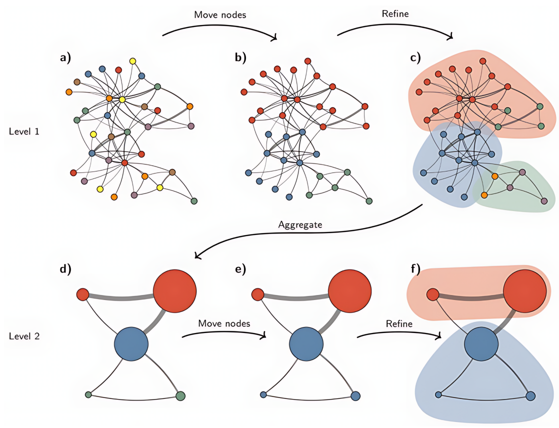


Figure 4: Illustrative diagram of the Leiden algorithm's operation, showing local refinement, aggregation into supernodes, and refinement of new communities to ensure internal connectivity [14].

3.3.1. Why Leiden over alternatives?

After testing Girvan-Newman, Label Propagation, and Louvain, the Leiden approach was selected because it guarantees well-connected communities and improved stability through refinement and aggregation steps, reducing the fragmentation observed with Louvain in heterogeneous, weighted networks, as suggested by [14]. Then, the Leiden approach optimized modularity by reassigning nodes, refining communities for connectivity, and aggregating them into super-nodes. This method prevents isolated communities, enhances stability, and identifies strong structures, especially in weighted heterogeneous networks. This study constructed the network using graph analysis tools and visualized it with a modified ForceAtlas2 layout to highlight neighborhood connections and internal cohesion.

3.4. Visualization of Patterns

After conducting the Leiden community detection, descriptive analyses were conducted to characterize victim profiles by neighborhood, focusing on demographic patterns related to violence. Considering the intersectionality of feminicide [3], four main variables were analyzed: age, race/color, education level, and case risk classification. Initial graphs showed high variability and no significant patterns, prompting a shift to dispersion measures, particularly standard deviation, for deeper insights into victim profile diversity. Moreover, a scatter plots correlating the total number of cases per neighborhood with the standard deviation of each variable were applied for achieved neighborhoods with higher incidences of violence and greater heterogeneity in victim profiles. What can indicate that multiple vulnerabilities coexist within the same area. LOWESS smoothing was used to clarify trends among the variables.

3.5. Operational Continuous Monitoring

The proposed pipeline is designed for continuous operation by public entities. Each newly recorded case triggers an incremental update: (i) the neighbourhood graph is refreshed; (ii) the Leiden community detection is re-executed to reorganize communities if needed; and (iii) dispersion metrics (standard deviations) for victims' education, race/colour, and case risk classification are recomputed and plotted. If statistically meaningful shifts are detected in these dispersion curves or community membership, an alert is raised to prompt review by decision-makers.

In practical terms, this enables near-real-time supervision of territorial dynamics: a steady rise in dispersion within a high-incidence cluster may signal profile diversification or changing modalities of violence, prompting rapid coordination of legal, social, and policing services in the most affected group. The dashboard can refresh at fixed intervals or be event-driven upon data insertion, ensuring that prioritization and resource allocation reflect the current state of the phenomenon.

4. Results and Discussion

This section presents the results obtained for descriptive spatial analysis of violence against women in Divinópolis, Minas Gerais, Brazil. First, territorial groupings generated by the Leiden algorithm are discussed, highlighting geographical patterns and neighbourhoods with the highest incidence of cases. Then, variations in victims' sociodemographic characteristics—such as education, race/colour, and case risk classification—are analyzed, focusing on the dispersion of these factors in the most affected neighbourhoods, aiming to identify areas with greater diversity and complexity in victim profiles.

Thus, the Leiden algorithm mapped violence against women by detecting clusters of spatial units as neighborhoods with similar characteristics. Through iterative refinement and modularity optimization, the algorithm groups areas based on the density and similarity of indicators (e.g., education level, race/color, and case classification), effectively highlighting regions with higher concentrations of reported cases. Each cluster was assigned a distinct color, visually distinguishing zones with shared social and geographic patterns. One of the key advantages of this unsupervised method is its ability to reveal latent structures in the data without prior labeling or predefined categories, making it particularly suitable for exploratory spatial analysis in contexts where patterns are complex and emergent, such as gender-based violence

4.1. Clustered Communities by Violence

To identify spatial clusters of neighbourhoods with similar violence profiles, the Leiden algorithm was applied to the network built based on geographical proximity and the number of recorded cases. Among tested methods, Leiden showed superior performance in terms of internal cohesion, cluster stability, and visual clarity. In this dataset, the outcome was clearer, more contiguous clusters aligned with neighbourhood adjacency and case intensities, as demonstrated in Figure 5, a property that is crucial for territorial policy applications.

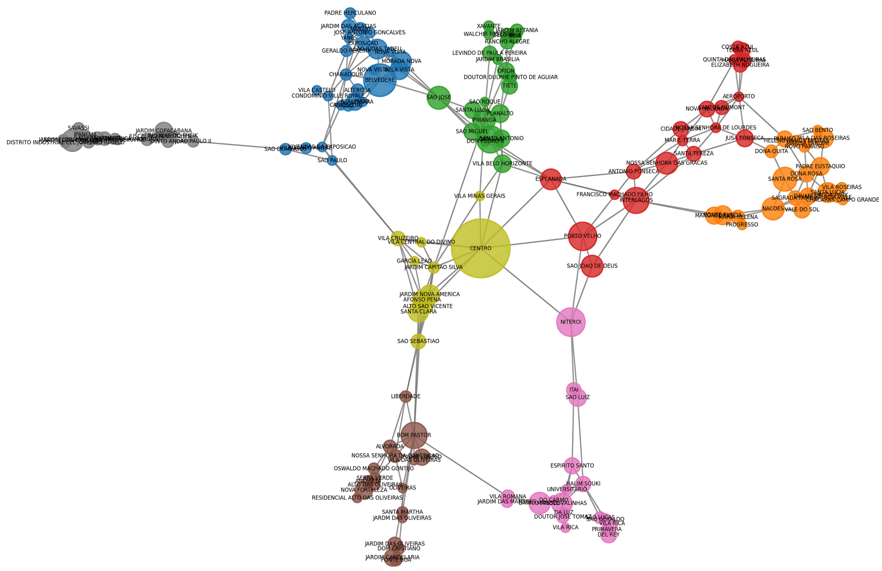


Figure 5: Graph of communities generated by the Leiden algorithm, with nodes representing neighborhoods in Divinópolis and edges indicating geographical adjacency.

Figure 5 demonstrates that neighborhoods with similar case counts tend to cluster together, reflecting coherent spatial patterns. The red community groups' neighborhoods, such as Esplanada, Porto Velho, São João de Deus, and Interlagos, all exhibit high incidence and physical proximity, suggesting common vulnerabilities. The yellow community includes neighborhoods such as Jardim Nova América, Afonso Pena, and Santa Clara, which also have similar occurrence profiles.

The Del Rey neighborhood, located in the vicinity of the Niterói neighborhood, exhibits a high incidence of cases despite its small territorial area, indicating possible local vulnerabilities linked to socioeconomic factors or inadequate infrastructure. This pattern is consistent with the femicide in Table 2 by neighbourhood reported by the [10]. Conversely, the Alvorada neighbourhood—larger and centrally located—shows fewer cases, potentially associated with better structural conditions or a stronger institutional presence.

This spatial segmentation clarifies critical hotspots and suggests important differences even among neighboring areas. The analysis underscores that violence against women, though present throughout the urban territory, exhibits concentration patterns in certain regions, justifying localized interventions.

4.2. Analysis of Variation in Standard Deviation of Victim Characteristics

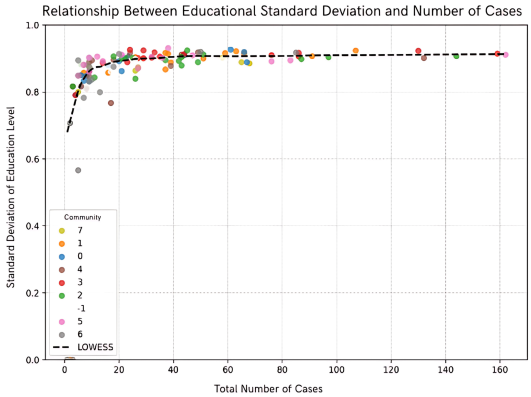
To understand victim profiles across neighbourhoods in Divinópolis, three variables were analyzed: education level, race/colour, and case risk classification. Initially, mean-based graphs for each variable did not reveal significant patterns due to high data dispersion. Therefore, the analysis shifted to the use of standard deviation, which proved to be a more effective metric for capturing the heterogeneity of victim profiles within each neighbourhood

Figure 6a shows the relationship between total cases and the standard deviation of victims' education levels. Neighborhoods with high violence incidence (over 100 cases) exhibit stabilized high standard deviations, indicating considerable diversity in educational levels among victims. Figure 6b illustrates Leiden algorithm-generated communities highlighting neighborhoods with 100 to 160 violence reports.

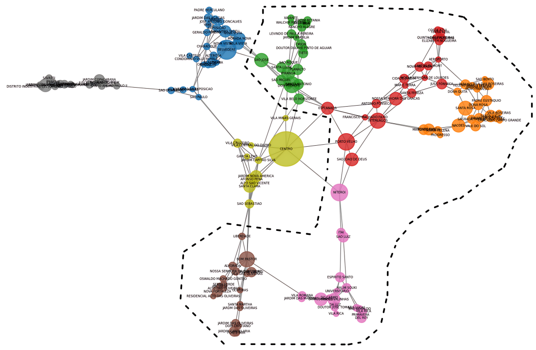
The graph in Figure 6a shows that neighborhoods with high violence incidence have a stabilized, relatively high standard deviation in educational levels, indicating significant diversity among victims. However, among neighborhoods with the highest number of cases (between 100 and 160 total), some communities, like 1, 3, 4, 2, and 5, exhibit a predominantly uniform educational pattern.

Figure 6b indicates how these neighborhoods cluster in the shaded area, suggesting similar educational diversity characteristics that may signify common local features. Figures 7a and 7b exhibit similar behavior concerning victims' race/color and case risk classification. Standard deviations remain high in neighborhoods with greater violence incidence, highlighting significant racial diversity among victims. The consistency of this measure can serve as a monitoring reference, with abrupt changes potentially signaling shifts in victim profiles over time.

Overall, standard deviation analysis proved more informative than mean-based analysis, providing essential insights. Stability in internal variations within high-incidence neighborhoods can serve as a useful indicator for continuous monitoring systems, enabling tracking potential changes in victim profiles or violence types.

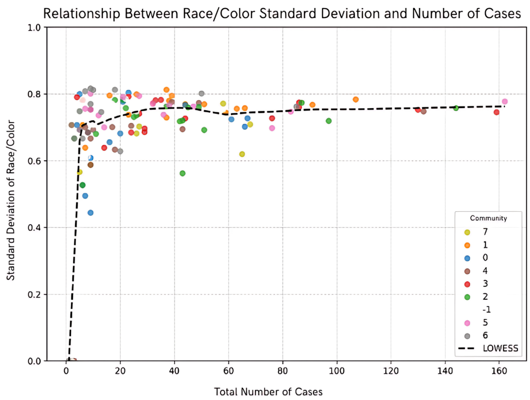


(a) Relationship between number of cases and standard deviation of victims' education by neighborhood.

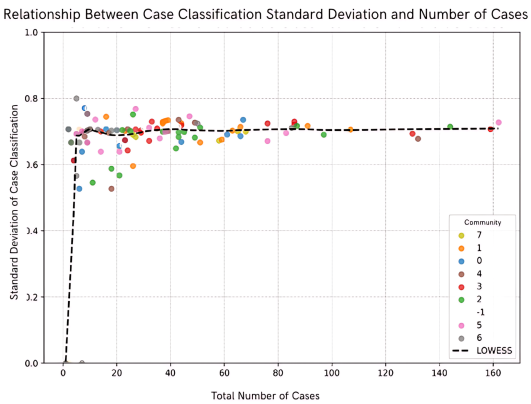


(b) Graph of Leiden-detected communities highlighting neighborhoods with 100 to 160 violence reports.

Figure 6: Graphical analysis of violence patterns based on community and victims' education.



(a) Relationship between number of cases and standard deviation of victims' race/color by neighborhood.



(b) Relationship between number of cases and standard deviation of case risk classifications by neighborhood.

Figure 7: Standard deviation distributions by neighborhood considering various victim characteristics.

4.3. Policy Implications

Spatially cohesive high-incidence clusters with stable internal heterogeneity indicate that prevention should combine territorial targeting (e.g., coordinated action in high-risk clusters) with tailored services that address diverse victim profiles within the same area (legal aid, shelters, income support, and safer mobility), as suggested by [6]. Monitoring shifts in dispersion over time can act as an early-warning indicator for changes in profiles or modalities of violence, informing the allocation of patrols and social assistance in the most affected communities.

5. Final Remaks

This study investigated violence against women in Divinópolis, Brazil, from a descriptive geographic and spatial perspective. It proposed a methodology based on complex networks to identify territorial patterns related to violence. By constructing a neighborhood network among districts and applying the Leiden algorithm, the study identified spatially cohesive communities with similar rates of violence, highlighting significant geographical clusters that can inform targeted public policies.

The analysis of variations in the standard deviation of victim characteristics, such as education level, race/colour, and case risk classification, indicated that neighborhoods with a higher number of incidents tend to have diverse victim profiles while exhibiting stable dispersion patterns. These findings support the hypothesis that violence against women is linked to structural and contextual factors, suggesting that its territorial distribution is not random.

Methodologically, the Leiden algorithm proved to be a powerful, unsupervised, graph-based tool capable of revealing latent spatial structures and forming well-connected clusters through modularity optimization and iterative refinement. The visualization of these clusters using color-coded communities provided valuable insights into areas of concentrated risk. However, the scope of this study was limited to the application of the algorithm in an exploratory, descriptive context, and the approach remains dependent on reporting practices. Cluster quality is sensitive to graph construction and parameterization, and a more comprehensive evaluation—considering metrics such as modularity gain, cluster stability, and sensitivity to parameter settings—is necessary to validate its effectiveness in spatial analyses of gender-based violence.

For future research, it is advisable to replicate this methodology in various cities to assess its transferability and examine the influence of educational institutions and supportive organizations in the region. Investigating the correlation between economic dependency and incidents of violence could provide further insight into structural drivers. Additionally, analysing neighborhoods with low incidence rates is crucial to determine whether the absence of reported cases reflects actual safety or results from underreporting.

Acknowledgments

The authors thank the Federal Center of Technology of Minas Gerais (CEFET-MG) for their support during this study, particularly the Research and Graduate Studies Director and the Department of Computing at the Divinópolis campus. We also acknowledge the State Judiciary of Minas Gerais, Brazil, for providing essential official data for our research.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT, Gemini, and Grammarly to assist with translation from Portuguese to English and to perform grammar correction. After using these tools, the authors carefully reviewed and edited the manuscript as necessary and take full responsibility for the content of this publication.

References

[1] Brasil, Law No. 11.340, August 7, 2006, Presidência da República, Casa Civil, 2006.

[2] Brasil, Law No. 13.104, March 9, 2015, Presidência da República, Casa Civil, 2015.

[3] S. N. Meneghel, A. P. Portella, Feminicides: concepts, types and scenarios, Ciência & Saúde Coletiva 22 (2017) 3077–3086. doi:10.1590/1413-81232017229.11412017.

[4] R. S. Silva, M. A. Souza, P. R. Gomes, Machine learning applied to predicting gender violence risk areas, Revista de Inteligência Artificial 19 (2022) 45–60. doi:10.1234/ria.v19i3.5678.

[5] L. L. Heise, Violence against women: An integrated, ecological framework, Violence Against Women 4 (1998) 262–290. doi:10.1177/1077801298004003002.

[6] C. Garcia-Moreno, A. F. M. Jansen, M. Ellsberg, L. Heise, C. Watts, Prevalence of intimate partner violence: findings from the who multi-country study on women's health and domestic violence, The Lancet 368 (2006) 1260–1269. doi:10.1016/S0140-6736(06)69523-8.

[7] Brazilian Institute of Geography and Statistics, Municipal basic information survey (munic) 2018, 2019.

[8] E. Oliveira, M. Â. D. Wermuth, "Here you don't come in anymore, I say I don't know you": electronic monitoring and the protection of women victims of domestic violence, in: Proceedings of the VII Virtual Meeting of CONPEDI – Criminologies and Criminal Policy II, National Council for Research and Graduate Studies in Law (CONPEDI), Florianópolis, Brazil, 2024, pp. 1–40. URL: http://www.conpedi.org.br/.

[9] R. Jewkes, M. Flood, J. Lang, From work with men and boys to changes of social norms and reduction of inequities in gender relations: A conceptual shift in prevention of violence against women and girls, The Lancet 385 (2015) 1580–1589. doi:10.1016/S0140-6736(14)61683-4.

[10] Governo de Minas Gerais, Access to information portal of the state of minas gerais, 2025. URL: https://acessoainformacao.mg.gov.br/, accessed on: Feb. 8, 2025.

[11] D. H. M. P. da Silva, A. M. de Morais, A. M. de Morais, Nossa voz: a digital platform for visualizing data on violence against women in brazil, in: Women in Information Technology (WIT), SBC, 2021, pp. 315–319.

[12] M. E. T. Souza, Women violence graphs, GitHub repository, 2025. URL: https://github.com/dudatsouza/women-violence-graphs, accessed on: Feb. 8, 2025.

[13] Prefeitura Municipal de Divinópolis, Map of divinópolis city, 1999. URL: https://divinopolis.mg.gov.br/arquivos/42_mapadacidade.pdf, accessed on: Feb. 8, 2025.

[14] V. A. Traag, L. Waltman, N. J. van Eck, From louvain to leiden: guaranteeing well-connected communities, Scientific Reports 9 (2019) 5233. doi:10.1038/s41598-019-41695-z.