Spatial-temporal patterns of aggressive behaviors. A case study Bogotá, Colombia

"Jorge Victorino Department of Computer Science Universidad Nacional de Colombia Bogotá, Colombia jevictorinog@unal.edu.co

"Cristian Pulido Department of Mathematics Universidad Nacional de Colombia Bogotá, Colombia cpulido@unal.edu.co "Jorge Rudas Institute of Biotechnology Universidad Nacional de Colombia Bogotá, Colombia jerudass@unal.edu.co

[•]Luisa Fernanda Chaparro Department of Mathematics Universidad Nacional de Colombia Bogotá, Colombia luchaparros@unal.edu.co "Ana Maria Reyes Department of Mathematics Universidad Nacional de Colombia Bogotá, Colombia amreyesp@unal.edu.co

[•]Luz Ángela Narváez Department of Mathematics Universidad Nacional de Colombia Bogotá, Colombia lanarvaezn@unal.edu.co

"Darwin Martinez Department of Mathematics Universidad Nacional de Colombia Bogotá, Colombia demartinezr@unal.edu.co "Francisco Gómez Department of Mathematics Universidad Nacional de Colombia Bogotá, Colombia fagomezj@unal.edu.co

Abstract-Understanding of crime patterns is paramount important for citizen's security planning. In particular, the comprehension of the Spatio-temporal dynamics related to aggressive behaviors is fundamental for deploying policial resources and devising mitigation actions. Currently, a significant number of approaches to find patterns, characterize dynamics, and predict crime have been proposed in state of the art. However, the operation of these approaches is strongly adapted to the specific conditions of a city. In this paper, we propose a novel approach to finding spatio-temporal crime patterns in the city of Bogotá (Colombia), particularly aggressive behaviors reported to the emergency line. We characterize aggressive behaviors through rhythm and tempo based on the theory of routine activity. Finally, we show that the dynamics of aggressive behaviors in the city are shared by several spatial units in which specific strategies can be applied to mitigate it.

Index Terms—Aggressive behaviors, Temporal patterns, Data analytics, Spatial-temporal data, Spectrograms.

I. INTRODUCTION

Recently, there has been an increase in the development of computational models to predict and characterize crime [1]–[5]. Among the motivations is to take prevention actions, optimize available resources, improve indicators and the perception of security, and define more effective public policies based on safety [4]. However, differences in environmental conditions, culture, size of cities, and type of crime change behavior patterns and crime dynamics, which means that many

of the models designed for a given city do not work in other cities [6]. In particular, for Bogotá, some approaches have been advanced for crime prediction, understanding of criminal behavior, and efficient allocation of resources for security [7]– [9]. Nevertheless, as far as we know, the dynamics of aggressive behavior in the city have not yet been characterized by historical data.

In general, these models work with the records of incidents where the public force intervenes or incidents reported by citizens on the emergency phone number. However, most of the proposed models are subject to certain conditions that are not always true and may affect the convergence or the certainty of the results [10]. For example, in the self-exciting point process, when crimes are considered personal injury, it is not clear that these incidents produce more personal injury aftershocks. On the other hand, the predict hotspots models are often susceptible to parameter selection in incident counts, such as bandwidth, outliers, and zeros [10].

There are a series of events reported to the emergency phone number when there are fights between individuals or different types of confrontations that shock the public peace, which is classified as aggressive behaviors [11]. Usually, these incidents correspond to acts between individuals that, through physical, verbal, or psychological violence, cause harm, injury, and/or pain [11]. Usually, aggressive behaviors are considered minor incidents because, most of the time, they do not have serious consequences. However, in Bogotá, about 70 percent of registered crimes of personal injury and homicide originate from aggressive behavior. Since 2014, Bogotá has reported

Supported by Bank of National Investment Programs and Projects, National Planning Department, Government of Colombia (BPIN: 2016000100036) IEEE/ACM ASONAM 2020, December 7-10, 2020 978-1-7281-1056-1/20/\$31.00 © 2020 IEEE



Fig. 1. Process to identify spatio-temporal patters of aggressive behaviors. The process begins with depured dataset of aggressive behavior reported to NUSE. The incidents are then divided and analyzed into space units. Incidents aggregated every six hours are using to construct time series in each space unit. Each series is analyzed in frequency using the Kaiser window and the Fourier transform. Finally, a topological analysis is used to clustering using the Mahalanobis distance and spatial units patterns are obtained.

almost half a million incidents of aggressive behavior per year [12].

policies by decision-makers.

According to the theory of routine activity, the occurrence of crimes, particularly aggressive behaviors, is related to the social-environmental features [13]. Additionally, when certain circumstances converge in space and time, the probability of incidents increases [13]. For example, when the flows of people around an event or celebration increase, there is alcohol consumption, and the lack of a guard or control mechanism. The theory of routine activity suggests that three elements characterize the occurrence of incidents: 1. the *rhythm* refers to the periodicity with which the events occur, 2. the *tempo*, refers to the volume of incidents that occur. They produce in a spatial unit, and 3. the *timing*, which is how the different circumstances that lead to the occurrence of incidents are coordinated [13].

In this approach, we propose a quantitative analysis of the reported incidents of aggressive behaviors oriented towards the characterization of rhythm and tempo in the city of Bogotá, since there is no information about the flow of activities of people to characterize timing. Our approach describes the rhythm through a measure of periodicity and the tempo through a trend measure based on spectrogram analysis [14], which is constructed by aggregating incidents in predefined units of space and time.

In this work, we seek if the different spatial units share a temporal dynamic in aggressive behaviors. We analyze how this is related to the characteristics of the social ecosystem that condition this occurrence. This analysis provides insights and a better understanding of aggressive behavior in the city of Bogotá in order to design mitigation strategies and intervention

II. MATERIALS AND METHODS

The process begins with all incidents of aggressive behavior reported to the emergency phone number in Bogotá, see Figure 1. The incidents are then divided and analyzed into space units. A monthly time series of incidents is constructed in each space unit, which are aggregated every six hours and with an overlap average of 53.56 days. Each series is analyzed in frequency using the Kaiser window and the Fourier transform. Finally, they are taken to a topological space where they are grouped by similarity using the Mahalanobis distance and spatial units patterns are obtained.

A. Data acquisition and preprocessing

2.235.552 reports of aggressive behaviors were collected in the Unique Number of Security and Emergency (Número Único de Seguridad y Emergencia - NUSE) system from 1st January 2014 and 24th November 2019 years in the city of Bogotá. The information collected contains the details and nature for each incident and the Spatio-temporal location. This data set was preprocessed by using the cleaning methodology suggested by [15]. This step included: merging dataset, rebuilding missing data, data standardization and normalization, records de-duplicate, and enrichment data.

Although the data on aggressive behaviors of NUSE indeed have disadvantages, such as that there are areas of the city where people do not make reports to the emergency phone, some of the incidents have duplicates, and there are unconfirmed incidents, which could be false positive. NUSE data has the advantage of being closer to a real scenario by having a



Fig. 2. PCA and TDA analysis. Left: UPZ's distribution in the representation space generated by spectrograms. The space was projected to two dimensions using PCA. Right: Persistent homology diagram of spectrograms per UPZ. Results for H_0, H_1, H_2 .

record 10 times greater than the data on complaints and penalties provided by the police. This is because the underreporting of aggressive behaviors reported to the emergency telephone number is much lower than those collected by the police, in contrast to what happens with the report of homicides and robberies.

B. Spatial-temporal analysis

Spatial-temporal analysis of aggressive behavior and personal injury events were represented by predefined spatial units and spectrograms [14]. Spectrograms use overlapping time windows that do a full frequency analysis of events grouped by a time unit and a spatial unit. The spatial unit chosen to study the dynamics of aggressive behaviors in Bogotá is the Zonal Planning Units (UPZ by its acronym in Spanish) [16]. These zones are delimited by main roads or geographical features and group neighborhoods with similar characteristics in land use, culture, and socio-economic conditions. The city has 112 UPZs in the urban part and 5 Rural Planning Units (UPR for its acronym in Spanish) in the rural part. Spectrograms were calculated each month, where 84-day windows (12 full weeks) were used, which overlap with the monthly time unit. The reported events accumulate every six hours, generating four values per day, starting at zero hours. This scheme provides a complete frequency analysis in the range: [(1/84)hz, 2hz]

To locate each spatial unit's events S_i , for each $i \in \{UPZ\}$, the centroid of latitude and longitude were calculated using all the event values that occurred in that unit. For each month T_j , j = 1, 2, ..., N, a time series was constructed starting at zero hours on the first of each month and with data aggregated every six hours for 84 days. Each time series T_j was multiplied by Kaiser window, and the Fourier transform was calculated. To obtain a Kaiser window [17] that represents an FIR filter:

$$w[n] = \frac{I_0 \left(\beta \sqrt{1 - \left(\frac{n - N/2}{N/2}\right)^2}\right)}{I_0(\beta)}, \quad 0 < n < N$$

where I_0 is the zeroth-order modified Bessel function [17], and β is the sidelobe attenuation parameter. The Discrete Fourier Transform on T_i was calculated as:

$$T_{k} = \frac{a_{o}}{2} + \sum_{n=0}^{N} a_{n} e^{2n\pi t_{k}/T}$$

where a_n , n = 0, 1, 2, ..., N - 1 are discrete Fourier coefficients.

Since the time series has real values, the frequency spectrum is symmetric. Therefore only half the spectrum was taken. Finally, a standardization process was applied to each spectrum. Whitening was first performed on the spectrograms: the mean was subtracted and divided by the standard deviation.

$$T_k^{(1)} = \frac{(T_k - \mu)}{\sigma}$$

where μ is the average of T_k and σ the standard deviation of T_k . Then it is scaled in the interval [0, 1] based on the sigmoid function and normalization factor f:

$$T_k^{(2)} = \frac{1}{1 + e^{-T_k^{(1)}f}}$$

The spectrogram of each UPZ E_i is constructed by combining all the normalized spectra $T^{(2)}k$ for each month j.

C. Topological analysis

The topological analysis was carried out for each spatial unit, the spectrogram was used as features of each unit. The distance between spectrograms evidence the relationships between the spatial units, and similar behavior. Given a matrix



Fig. 3. UPZ's with the highest degree of similarity. The figure shows the name of UPZ, its respective spectrograms, and degree of similarity with the others UPZ's.

 $A^{n,m}$ where n = 1, 2, ..., N, with N the number of spatial units, and m = 1, 2, ..., M, with M as the number of features, that results from $M = T \times K$, with T as the number of months and K the number of harmonics. To measure the distance, Mahalanobish was used between two rows x_i, x_j of the matrix A, such as: $d_{i,j} = \sqrt{(x_i - x_j)^T / S(x_i - x_j)}$, where S is the covariance matrix [18].

Homological persistence analysis was performed with the distance matrix D from the characteristics matrix A [19]. To build the persistence diagram based in the Vietoris Rips filtration scheme [20], was used the library known as Ripser [21]. For a filter value, the connected components, loops, and voids were shown from the persistence diagram. Additionally, a Principal Components Analysis (PCA) was applied in the representation space generated by the spectrograms, to make a description of the dimensionality and distribution of each

UPZ.

III. RESULTS

The data used for this analysis reported incidents in 112 UPZ and five UPR spatial units. For each unit, a spectrogram was constructed, as shown with some UPZ of the Figure 3. Each point in the spectrogram's horizontal axis is a month, beginning in 2014 and ends in 2019. The vertical axis of the spectrogram corresponds to the frequency by days. Therefore, the value of 1/7 is the weekly frequency, and the value of 1 corresponds to the daily frequency. The spectrograms of the spatial units selected in Figure 3 show a predominant weekly and daily frequency and a significant increase of incidents in all December.

The spectrograms were used as the characteristics of each spatial unit. In this way, all the spatial units were mapped



Fig. 4. Interactive map of UPZ's Bogotá. The UPZ's with a similar incident dynamic is highlighted in red. Additionally, the spectrogram of the selected UPZ and the remaining UPZ's is shown in yellow.

to the feature space defined by the spectrogram. Each spatial unit is represented by a point of 2016 dimensions in the characteristics space (2016 = 168 frequencies by 12 monthly averages). Figure 2 left side, shows the distribution of the points in the feature space, which was reduced to 2 dimensions by PCA for presentation purposes. The first two components of PCA represented 16.6% of the information. To obtain 90% of the information, at least 99 components are required.

In Figure 2 right side, the homological persistence diagram using Mahalanobis distance for H_0 , H_1 and H_2 is shown. The persistence diagram shows that the elements connect almost uniformly over the entire distance range (0, 50). The results of homological persistence H_0 suggest the optimum is to divide the spatial units into two groups: the first group of twelve units that are relatively close, and the second group the rest units, which are highly dispersed. As can be seen in the persistence diagram H_1 , H_2 does not present information, suggesting that there are no cycles in the organization of spatial units.

Figure 3 shows how the 12 UPZ with the highest measure of similarity were connected with their respective spectrogram. The color intensity of the line connections means the degree of similarity among differents UPZ. The darker corresponds higher the similarity value. The figure 3 shows the 44 most robust connections between UPZ. Figure 4 shows how the 12 UPZ were distributed with the most significant measure of similarity in the city. In this UPZ group (Figure 3), it is shown that the spectrograms have a well-defined pattern. Analyzing these spectrograms shows that all of them present

a high frequency of daily, weekly, inter-weekly, and monthly incidents. Furthermore, December's month presents a high rate of incidents due to many traditional celebrations in Colombia. In opposition to the low rate presented in January because many people go on vacation outside the city. This contrasts with spectrograms such as "Paseo De Los Libertadores, Jardín Botánico, Parque Salitre, La Academia, Country Club, San Isidro" where the entropy measure is high, which implies a high degree of uncertainty, which suggests a random behavior of the incidents.

The configuration of spatial units in Figure 3 presents similar temporal dynamics and a high rate of incidents, but they also coincide with other fundamental aspects. For example, UPZs correspond to highly populated residential areas with low economic resources and high social needs since these areas host a high percentage of the population displaced by the violence of the entire national territory and the immigrant population of Venezuela, in addition to the social challenges mentioned [22]. According to the theory of routine activity, several conditions meet to increase the probability of aggressive behavior in these areas, such as large flows of people on weekends around different cultural activities accompanied by alcohol consumption. Despite the police's greater presence in this area [23], there is still a high rate of incidents, suggesting that preventive social intervention programs that promote self-control may be more effective than increased police surveillance.

IV. CONCLUSIONS

The results of the work have shown that there is no clearly established pattern of the occurrence of incidents of aggressive behavior. However, it is possible to establish similar dynamics of incident occurrence between UPZ groups, except for those spatial units that show a random behavior of incidents.

The spatial-temporal characterization of the dynamics of incidents of aggressive behavior and personal injury in Bogotá is a relevant finding for decision-makers to design and test the success of prevention strategies.

UPZ with similar behavior also show a degree of similarity with respect to socioeconomic conditions and land use. This can be an important finding that must be statistically evaluated. In the same way, it is expected to be able to relate the incidents with other variables such as demographic variables and social, cultural and sports events.

ACKNOWLEDGEMENTS

This work was funded by the project "Diseño y validación de modelos de analítica predictiva de fenómenos de seguridad y convivencia para la toma de decisiones en Bogotá", at Bank of National Investment Programs and Projects, National Planning Department, Government of Colombia (BPIN: 2016000100036).

REFERENCES

- J. Lukasczyk, R. Maciejewski, C. Garth, and H. Hagen, "Understanding hotspots: a topological visual analytics approach," in *Proceedings of the* 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 1–10, 2015.
- [2] D. Yang, T. Heaney, A. Tonon, L. Wang, and P. Cudré-Mauroux, "Crimetelescope: crime hotspot prediction based on urban and social media data fusion," *World Wide Web*, vol. 21, no. 5, pp. 1323–1347, 2018.
- [3] C. Catlett, E. Cesario, D. Talia, and A. Vinci, "A data-driven approach for spatio-temporal crime predictions in smart cities," in 2018 IEEE International Conference on Smart Computing (SMARTCOMP), pp. 17– 24, IEEE, 2018.
- [4] R. Yadav and S. K. Sheoran, "Crime prediction using auto regression techniques for time series data," in 2018 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE), pp. 1–5, IEEE, 2018.
- [5] R. A. Silva, J. M. Pires, N. Datia, M. Y. Santos, B. Martins, and F. Birra, "Visual analytics for spatiotemporal events," *Multimedia Tools and Applications*, vol. 78, no. 23, pp. 32805–32847, 2019.
- [6] M. Oliveira, E. Ribeiro, C. Bastos-Filho, and R. Menezes, "Spatiotemporal variations in the urban rhythm: the travelling waves of crime," *EPJ Data Science*, vol. 7, no. 1, p. 29, 2018.
- [7] G. Escobar, "Using social disorganization theory to understand the spatial distribution of homicides in bogota, colombia," *Revista Invi*, vol. 27, no. 74, pp. 21–85, 2012.
- [8] M. Dulce, S. Ramírez-Amaya, and Á. Riascos, "Efficient allocation of law enforcement resources using predictive police patrolling," *arXiv* preprint arXiv:1811.12880, 2018.
- [9] M. Dulce, "Predicting criminal behavior with lévy flights using real data from bogotá," *Documento CEDE*, no. 2019-11, 2019.
- [10] M. Adepeju, G. Rosser, and T. Cheng, "Novel evaluation metrics for sparse spatio-temporal point process hotspot predictions-a crime case study," *International Journal of Geographical Information Science*, vol. 30, no. 11, pp. 2133–2154, 2016.
- [11] J. Liu, "Concept analysis: aggression," *Issues in mental health nursing*, vol. 25, no. 7, pp. 693–714, 2004.

- [12] A. M. Secretaría de Seguridad de Bogotá, "Análisis de información estudios estratégicos, "boletín v mensual convivencia".," seguridad de indicadores de 2018. v url:http://www.scj.gov.co/sites/default/files/documentos_oaiee/boletin_ bogota_2018_12.pdf Accessed 30th March, 2020.
- [13] L. E. Cohen and M. Felson, "Social change and crime rate trends: A routine activity approach," *American sociological review*, pp. 588–608, 1979.
- [14] W.-k. Lu and Q. Zhang, "Deconvolutive short-time fourier transform spectrogram," *IEEE Signal Processing Letters*, vol. 16, no. 7, pp. 576– 579, 2009.
- [15] O. Elgabry, "The ultimate guide to data cleaning," tech. rep., Towards to data science, Feb 2019. url:https://towardsdatascience.com/the-ultimateguide-to-data-cleaning-3969843991d4.
- [16] A. B. Secretaría Distrital de Planeación, "Norma urbana. actos administrativos reglamentarios del pot para las unidades de planeamiento zonal," 2020. url:http://www.sdp.gov.co/gestion-territorial/normaurbana/normas-urbanisticas-vigentes/upz Accessed 14th April, 2020.
- [17] J. F. Kaiser, "Nonrecursive digital filter design using the i_0-sinh window function," in Proc. 1974 IEEE International Symposium on Circuits & Systems, San Francisco DA, April, pp. 20–23, 1974.
- [18] R. De Maesschalck, D. Jouan-Rimbaud, and D. L. Massart, "The mahalanobis distance," *Chemometrics and intelligent laboratory systems*, vol. 50, no. 1, pp. 1–18, 2000.
- [19] G. Carlsson, "Topology and data," Bulletin of the American Mathematical Society, vol. 46, no. 2, pp. 255–308, 2009.
- [20] H. Edelsbrunner and J. Harer, Computational topology: an introduction. American Mathematical Soc., 2010.
- [21] U. Bauer, "Ripser: a lean c++ code for the computation of vietoris-rips persistence barcodes," *Software available at https://github.* com/Ripser/ripser, 2017.
- [22] A. B. Secretaría Distrital de Planeación, "Caracterización socioeconómica, encuesta sisben iii, boletín 64," 2015. url:http://www.sdp.gov.co/sites/default/files/caracterizacion_s isben – 04032015.pdf Accessed14thApril, 2020.
- [23] A. M. Secretaría de Seguridad convivencia y justicia de Bogotá, "Plan integral de seguridad, convivencia ciudadana y justicia 2017-2020.," 2017. url:https://scj.gov.co/sites/default/files/planeacion/PISCJ.pdf Accessed 30th March, 2020.