Consistent spatial decomposition of temporal occurrence of aggressive behaviors: A case study in Bogotá, Colombia

'Jorge Rudas Institute of Biotechnology Universidad Nacional de Colombia Bogota, Colombia jerudasc@unal.edu.co

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

"Jorge Victorino Department of Computer Science Universidad Nacional de Colombia Bogota, Colombia jevictorinog@unal.edu.co "Cristian Pulido Department of Mathematics Universidad Nacional de Colombia Bogota, Colombia cpulido@unal.edu.co

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

[•]Darwin Martinez Department of Mathematics Universidad Nacional de Colombia Bogota, Colombia demartinezr@unal.edu.co [.]Francisco Gómez Department of Mathematics Universidad Nacional de Colombia Bogota, Colombia fagomezj@unal.edu.co

Abstract-Aggressive behaviors are triggers of personal injuries and homicides in modern cities. Their origin is multicausal, but some phenomena, such as agglomerations and alcohol use, exacerbate these events. Understanding how the occurrence of aggressive behaviors is spatially and temporally distributed in cities will potentially allow making better governmental policies to mitigate them. Traditionally, the occurrence of aggressive behaviors in cities results from an aggregation of the reporting count of this kind of event over some range of time. Recent work suggests that this counting pattern can be obtained from a combination of independent sources of aggressive behavior occurrence. This paper explores the existence of underlying shared sources of activity in the occurrence of aggressive behaviors in the city of Bogotá, Colombia, by using source decomposition. Our results suggest that these behaviors are related to consistent and reproducible independent sources of the activity through different spatial scales in the city Bogotá.

Index Terms—Aggressive behaviors; crime; matrix factorization; patterns of crime; crime data mining.

I. INTRODUCTION

Aggressive behavior is an individual or collective social interaction with a hostile behavior with the intention of inflicting damage or harm [1], [2]. These social events are significant triggers of personal injuries and homicides in cities [3]. Therefore, they represent a fundamental problem in

the design of safety and health public policies [3]. Different studies suggest a multicausal origin of aggressive behaviors linked mainly to individual and environmental features [4]–[6]. For instance, citizens agglomerations, excessive alcohol consumption, and a history of family violence can exacerbate these events [7]. Understanding the dynamics associated with the appearance of aggressive behaviors in cities from a spatial and temporal perspective will potentially allow better policies to reduce them and mitigate their effects.

In quantitative criminology for large cities, aggressive behaviors are commonly studied by using spatio-temporal occurrence records of these events [3], [4]. Particularly by measuring the occurrences as counts of events at specific temporal scales, for example, weeks, days, months, and a particular spatial scales, for example, blocks, neighborhoods, localities, states, and others. This strategy has identified several patterns of occurrence of these crimes over time, for instance, discovering crime trends and to make short-term forecasting of crimes [8], detecting crime hotspots from raw crime data [9] and others [10]-[12]. More recently, it has been suggested that these patterns of occurrence may emerge from a linear combination of independent sources of crime [13]. This finding represents evidence pointing to the existence of "independent sources" of activity underlying the occurrence of aggressive behaviors in large cities, which can be understood alternatively as latent factors with common temporal occurrence patterns shared across different spatial units. This approach has recently

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

been used for the analysis of time series associated with crimes in the city of Sao Paulo, resulting in the identification of particular spatial patterns of interest in the occurrence of crimes [13]. Importantly, the existence of these patterns may be related to environmental and ecological features, as hypothesized by modern theories in criminology, for instance, the routine activity theory [14]. Nevertheless, despite the importance of this pattern, their robustness and consistency of these patterns across different spatio-temporal scales are still poorly understood for the crime dataset.

Aggressive behaviors are a result of social problems related to the absence of education, domestic violence, bad affective family environment, and in general, corrupt sociocultural practices [15]–[17]. For instance, in Colombia, many people have been victims of aggressive behaviors, including domestic violence, terrorism, hooliganism, or get physically or psychologically injured [18]. Notably, in Bogotá, the capital of Colombia, even if the number of homicides (per 100000 people) is decreasing during the last ten years, the number of aggressive behaviors events keeps growing [19], [20]. Therefore, understanding the spatial and temporal patterns of occurrences of these incidents represents a critical challenge for decision-makers from the local government in Bogotá. For this reason, this paper proposes exploring a source decomposition strategy of counting the occurrence of aggressive behaviors in the case of the city of Bogotá, Colombia, during the range of years from 2014 to 2018. We aimed to study the existence, robustness, and consistency of these patterns across multiple spatio-temporal scales, which may potentially enable decision-makers at citizen's security departments a rational use of limited resources and develop more effective strategies for crime prevention and mitigation.

II. MATERIALS AND METHODS

Figure 1 illustrates the proposed method. First, the crime data was preprocessed to reduce the internal variability of the data II-A. The dataset information was aggregated by counting incidents in two different spatial scales and a particular temporal scale. After that, data was decomposed into Spatiotemporal components by using five different methods (single and group-level analysis); see Section II-B and II-D. Finally, the similarity between different factorization algorithms was estimated.

A. Data acquisition and preprocessing

A total of 3.024.784 reports of aggressive behaviors were collected in the Unique Number of Security and Emergency (Número Único de Seguridad y Emergencia - NUSE) system from 2014 and 2018 years in the city of Bogotá. This information contains the details and nature for each incident and its spatio-temporal location. This data set was preprocessed by using the cleaning methodology suggested by [21]. This step included: merging dataset, rebuilding missing data, data standardization and normalization, records de-duplicate, and enrichment data. After the cleaning process, a new dataset with 2.268.394 records was consolidated. Additionally, localities

and planning units zones (UPZs for its acronym in Spanish) were used as spatial scales [22]. Therefore, these spatial entities were used to aggregate by counting the daily reports of aggressive behaviors within the range of time from 2014 to 2018. Finally, time series of occurrence of aggressive behaviors were globally normalized to zero-mean subtracting the general average.

B. Matrix decomposition algorithms

In linear algebra, a matrix decomposition or matrix factorization methods are a set of strategies that decomposition a matrix into a product of two lower dimensionality submatrices [23]-[25]. There are several methods to do a matrix decomposition [23]–[25]. In general, a matrix factorization can be considered an approach to split a matrix into its constituent parts that make it easier to calculate more complex matrix operations [26]. Another point of view to understand the matrix factorizations is the idea to discover the latent factors from the data matrix [26]. This last perspective is the approach used for the following analysis. To investigate robustness five different algorithms for matrix factorization were explored in this work: independent component analysis (ICA) [27], Tucker decomposition via Higher-Order Orthogonal Iteration (HOI) [28], CANDECOMP/PARAFAC decomposition via alternating least squares (ALS) [28], Nonnegative PARAFAC/CANDECOMP (CP) decomposition [29] and Nonnegative Tucker decomposition [30].

1) Independent component analysis (ICA): ICA is an extension of the principal component analysis (PCA) method [27]. PCA solves the optimization problem in the covariance matrix of the data, and this one represents a second-order statistic, while ICA optimizes higher-order statistics such as kurtosis [28]. PCA finds uncorrelated and orthogonal components that maximize the representation of variance in the data, while ICA finds components statistically independent [28]. ICA technique can be considered as a blind source separation technique, that means, there is a signal which is resulting from a linear combination of many independent sources [29]. The topic of separating these mixed signals is called blind source separation (BSS). The blind term is related to the fact that the source signals can be separated even if little information is available from the nature of the source signal [?], [29].

2) Tucker decomposition via Higher Order Orthogonal Iteration (HOI): The Tucker decomposition is a form of higherorder PCA. Tucker algorithm decomposes a tensor into a core tensor multiplied (or transformed) by a matrix along each dimension [31]. Thus, each of these matrices (which are usually orthogonal) can be thought of as the principal components for each dimension inside the original matrix [31].

3) CANDECOMP/PARAFAC (CP) decomposition via alternating least squares (ALS): CP decomposition factorizes a tensor into a linear combination of rank one tensors. The Alternating Least Squares (ALS) algorithm [32], [33] is one of the most famous and commonly used algorithms to solve the tensor factorization, which updates one component iteratively at each round, while holding the others constant [32], [33].



Fig. 1. Proposed method. First, the crime data was preprocessed to reduce the internal variability of the data. The dataset information was aggregated by counting bases of two different spatial scales (UPZ - unity of zonal planning and localities - groups of neighborhoods) and a defined temporal scale (days). The data was decomposed into Spatio-temporal components by using five different methods (single and group-level analysis). Finally, the similarity level across different factorization algorithms was estimated.

4) Non-negative matrix factorization: This approach constraints the matrix factorization algorithms to products of two matrices W and H, with the property that all three matrices have no negative elements [31]. This kind of algorithm has become popular strategies because of its ability to automatically extract sparse and easily interpretable factors [31].

C. Matrix decomposition in the crime context

Given that the reports of crime in the cities traditionally are estimated from the aggregated counting of these kinds of events, then it can be considered that this aggregated counting variable may have emerged from a combination of independent latent factors of crime in the cities. For applying these techniques to the criminal context, the following approach was used:

Let A^{nxm} a variable which represents the level of occurrence of aggressive behaviors in a particular city with n is the number of spatial units and mis the number of temporal units to describe the crime. A matrix factorization methods over this variable A can be applied in order to identify clusters o latents factors of spatial units with independent temporal behaviors in the city of Bogotá, Colombia. UPZs (Unidad de Planeación Zonal - Units of Zonal Planning) and localities (groups of neighborhoods) of Bogotá, Colombia, were the spatial scales used in this analysis.

On the other hand, days of years were also used as temporal scales. In particular, using 117 spatial unites (UPZs) for Bogotá, and using the years as temporal scale, the variable Ahad this form A117x365. When was used localities as spatial scale (19 localities, Sumapaz locality was discarded) and years as temporal scale, the variable A had this form A^{19x365} . Finally, to determine the best number of components, the level of reproducibility was computed for each case (year by year). Nineteen different numbers of components were validated (from 2 to 20 numbers of components). It is important to note that the number of latent factors expected.

D. Group level analysis

Based on the results of the previous analysis, it was computed by a group matrix factorization analysis to identify the common behavior at the group level during the whole temporal variable. This kind of analysis was inspired by the group ICA analysis traditionally used in the neuroscience field [34]. In particular, the matrix A introduced in the previous section was concatenated over the temporal scale for each year. Thus, a new B^{txm} variable is introduced here, where t is the temporal variable resulting from the temporal concatenation for each year. Then, this B grouped variable was used as an input in the matrix factorization analysis for each algorithm.

III. RESULTS

Figure 2 shows the similarity among components (average across matrix factorization methods) when different numbers of components were used for the matrix factorization analysis at years and group level. As observed, five components produced the highest similarity level at the group level from 2014 to 2018 (absolute Pearson correlation coefficient = 0.818 among components). When comparing components across methods in a single analysis, five components also showed the best similarity between 2016 and 2018. Moreover, 2014 and 2017 achieved the best similarity using the fourth component. Finally, 2015 produced the maximal similarity level using six components.

Figure 3 shows the mean spatial pattern when five components at the group level were explored by using two different spatial scales (localities and UPZs). Kennedy emerged as a particular and isolated latent factor, see plots C in Figure 3 (top). Moreover, the localities of Bosa, Kennedy, and San Cristobal share temporal behavior in certain ways; see plot D in Figure 3 (top). Additionally, localities such as San Cristobal, Usme, Ciudad Bolívar, Bosa, Kennedy, Engativa, and Suba also seem to share a temporal behavior of aggressive behavior B plot in Figure 3 (top). Kennedy, Engativa, and Suba also belong to the same grouping level, see A plot in Figure 3 (top). Finally, Figure 3 (botton) represents the same results as Figure 3 (top), but using UPZs like spatial scale.



Fig. 2. Similarity level versus the number of components across years (single and group level analysis). Five independent components produced maximal similarity level among years (group level). During 2016 and 2018 five components also produce the maximal similarity level cross factorization algorithms.

A. Discussions and Conclusions

This work explored the possibility to characterize common temporal behaviors in the occurrence of aggressive behaviors at different spatial scales in the city of Bogotá (Colombia). Main hypothesis here was that a mixture of independent latent factors of occurrence of aggressive behaviors produces the quantifiable measure of this phenomenon. Results suggest that there are common patterns of temporal occurrence of aggressive behaviors shared across different spatial units. Recent evidence showed that crime patterns persistently change through several spatial units in the cities [35], [36]. These changes are tightly related to the kind of crime event, the social conditions, organization of crime and other causes [37]. Many studies have been trying to identify these complex relationships among crime occurrences and socio-economic variables [35], [38]. Our result for the spatial pattern analysis (see Figure 3) tried to understand how the spatial pattern of crime, in particular aggressive behavior, emerges across several localities in Bogotá city. Five latent factors seem to underlie this kind of crime event. Even if this work is not possible to infer what variables correspond to these latent factors, previous works have shown that this kind of approach produces consistent results in the identification of spatial patterns of crime in the cities [35]. Several independent components identified in Figure 3 show localities in Bogotá which have similar socio-economics and cultural aspects. For instance, components B, D and E in Figure 3, shows that localities such as Bosa, Kennedy and Ciudad Bolivar share a pattern of occurrence of aggressive behavior from 2014 to 2018. This aggregation level correlates with the fact that these three localities are the most poverty localities in the city of Bogotá according to the unmet basic needs percentage (NBI - Necesidades básicas insatisfechas, by its acronym in Spanish) [39]. Additionally, components A, B, E show that localities such as Engativa, Suba and Kennedy are inside the same aggregation level. This evidence correlates with the fact

that these localities are some of the most dense spatial unities, as well as, these are some of the least green areas (measured as the number of trees inside them) in the city of Bogotá [40], [41].

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

- R. M. M. de Almeida et. al., "Behavioural, hormonal and neurobiological mechanisms of aggressive behaviour in human and nonhuman primates," 2015.
- [2] K. A. e. a. Miczek, "Neurobiology of escalated aggression and violence," J. Neurosci., vol. 27, pp. 11803–11806, Oct. 2007.
- [3] R. Greenberg, "Aggressiveness, violence, homicidality, homicide, and lyme disease," *Neuropsychiatr. Dis. Treat.*, vol. 14, pp. 1253–1254, May 2018.
- [4] M. e. a. Zahid, "Predicting risky and aggressive driving behavior among taxi drivers: Do Spatio-Temporal attributes matter?," 2020.
- [5] A. e. a. Takahashi, "Behavioral and pharmacogenetics of aggressive behavior," *Curr. Top. Behav. Neurosci.*, vol. 12, pp. 73–138, 2012.
- [6] M. Molina-Borja, "Spatio-temporal distribution of aggressive and courting behaviors in the lizardgallotia galloti from tenerife, the canary islands," 1987.
- [7] M. E. Flanigan and S. J. Russo, "Recent advances in the study of aggression," 2019.
- [8] B. e. a. Chandra, "A multivariate time series clustering approach for crime trends prediction," 2008.
- [9] K. Kianmehr and R. Alhajj, "Crime Hot-Spots prediction using support vector machine," 2006.
- [10] Y. e. a. Zhuang, "Crime hot spot forecasting: A recurrent model with spatial and temporal information," 2017.
- [11] J. Corcoran and A. Ware, "Crime hot spot prediction."
- [12] Y. e. a. Lee, "A Theory-Driven algorithm for Real-Time crime hot spot forecasting," 2020.
- [13] G. G. e. a. Zanabria, "CrimAnalyzer: Understanding crime patterns in são paulo," 2019.
- [14] M. Felson, Social Change and Crime Rate Trends: Routine Activity Approach. 1977.



Fig. 3. Group level matrix factorization estimating the spatial patterns of occurrence of aggressive behaviors among 2014 - 2018 in the city of Bogotá, Colombia. Five components were explored in this analysis. A, B, C, D, E show the spatial pattern for each component. Top part shows the spatial pattern using localities as spatial scale and the bottom part represents the spatial pattern for each component using UPZs. The color inside each image represents the z statistical value related to each component.

- [15] I. Ibabe, "Adolescent-to-Parent violence and family environment: The perceptions of same reality?," *Int. J. Environ. Res. Public Health*, vol. 16, June 2019.
- [16] Y. e. a. Cao, "Effects of exposure to domestic physical violence on children's behavior: a chinese community-based sample," 2016.
- [17] E. A. Olusegun, "Domestic violence, risky family environment and children: A bio-psychology perspective," 2014.
- [18] G. D. Paredes Zapata, "Terrorism in colombia," *Prehosp. Disaster Med.*, vol. 18, pp. 80–87, Apr. 2003.
- [19] J. Y. Paz, "Hubo mil casos más de agresiones por riñas reportadas en el 2018," Jan. 2019. Accessed: 2020-2-7.
- [20] Bogotá, "Las claves de bogotá para lograr histórica reducción de homicidios," May 2019. Accessed: 2020-2-7.
- [21] O. Elgabry, "The ultimate guide to data cleaning," Feb. 2019. Accessed: 2020-1-15.
- [22] "Datos abiertos bogotá." Accessed: 2020-1-15.
- [23] S.-O. et. al, "Enter the matrix: Factorization uncovers knowledge from omics," *Trends Genet.*, vol. 34, pp. 790–805, Oct. 2018.
- [24] "Chapter 3 matrix factorization for recovery of biological processes from microarray data," in *Methods in Enzymology*, vol. 467, pp. 59– 77, Academic Press, Jan. 2009.
- [25] "Max-margin non-negative matrix factorization," *Image Vis. Comput.*, vol. 30, pp. 279–291, May 2012.
- [26] "Matrix factorization an overview ScienceDirect topics." Accessed: 2020-7-12.
- [27] E. Bingham and A. Hyvärinen, "A fast fixed-point algorithm for independent component analysis of complex valued signals," *Int. J. Neural Syst.*, vol. 10, pp. 1–8, Feb. 2000.
- [28] G. R., "Introduction: Independent component analysis," 2012.
- [29] J. V. Stone, Independent Component Analysis: A Tutorial Introduction. MIT Press, 2004.
- [30] Y. Kim and S. Choi, "Nonnegative tucker decomposition," in 2007 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8, 2007.

- [31] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," 2009.
- [32] R. A. Harshman, Foundations of the PARAFAC Procedure: Models and Conditions for an "explanatory" Multi-modal Factor Analysis. 1970.
- [33] J. D. e. a. Carroll, "Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition," 1970.
- [34] e. a. Jafri, Madiha J, "A method for functional network connectivity among spatially independent resting-state components in schizophrenia," *Neuroimage*, vol. 39, pp. 1666–1681, Feb. 2008.
- [35] G. G. Z. et. al., "Crimanalyzer: Understanding crime patterns in são paulo," *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2019.
- [36] B. P. Manso, *The Structure of Crime in São Paulo*, pp. 107–111. Cham: Springer International Publishing, 2016.
- [37] J. H. Ratcliffe, Understanding Crime Patterns, pp. 75–99. Reducing Crime.
- [38] J. e. a. Caplan, Risk Terrain Modeling Manual: Theoretical Framework and Technical Steps of Spatial Risk Assessment for Crime Analysis. Rutgers Center on Public Security, 2010.
- [39] S. D. de Planeación Alcaldía Mayor de Bogotá, "Indice distribucion de recursos localidades (documento técnico)," 2020.
- [40] "Número de árboles por localidad datos abiertos colombia." https://www.datos.gov.co/dataset/N-mero-de-rboles-por-localidad/9b9k-4hhk. (Accessed on 11/16/2020).
- [41] "Zonas que tienen más habitantes por km cuadrado en bogotá bogotá - eltiempo.com." https://www.eltiempo.com/bogota/zonas-quetienen-mas-habitantes-por-km-cuadrado-en-bogota-190254. (Accessed on 11/16/2020).