

Agrupamiento basado en densidad para la detección automática de hot spots delictivos en la CDMX

Jorge Carrasco-Islas, Edgar Moyotl-Hernández

Benemérita Universidad Autónoma de Puebla,
Facultad de Ciencias Físico Matemáticas,
México

jorge.carrascoi@alumno.buap.mx, emoyotl@fcfm.buap.mx

Resumen. Este trabajo presenta la aplicación del agrupamiento basado en densidad a un conjunto de datos de delitos sexuales, cometidos en la Ciudad de México (CDMX) de 2019 a 2021, para detectar automáticamente zonas con incidencia delictiva elevada (hot spots delictivos). La metodología propuesta consiste en aplicar el método de agrupación OPTICS al conjunto de datos espaciales para localizar zonas con las agrupaciones más densas de puntos en los mapas y en estimar el área y la magnitud de incidencia delictiva encontrada en dichas zonas. Los resultados obtenidos indican que los hot spots no necesariamente coinciden con divisiones territoriales como podrían ser las alcaldías o las colonias de la ciudad. Así mismo, se observó que ciertos hot spots se repiten o presentan desplazamientos menores en distintos años.

Palabras clave: Datos espaciales, puntos críticos, mapas del delito, agrupamiento, OPTICS, patrones delictivos.

Density-Based Clustering for Automatic Detection of Criminal Hot Spots in Mexico City

Abstract. This work presents a density-based clustering approach for automatic detection of criminal hot spots in Mexico City, focusing on sexual crimes committed during the 2019 – 2021 period. The proposed methodology consists of applying the OPTICS clustering algorithm to the data set in order to locate the regions with densest clusters of points on the maps and estimating the area and magnitude of criminal incidence in said regions. The results suggest that hot spots do not necessarily match with territorial divisions of Mexico City such as districts or neighborhoods. Moreover, it was observed that for different years, certain hot spots consistently reappeared or presented minor position shifts.

Keywords: Spatial data, hot spots, crime maps, clustering, OPTICS, crime patterns.

1. Introducción

La violencia y la inseguridad son un problema grave en México, de acuerdo con la última edición de la ENVIPE se estima que, en 2021, a nivel nacional el 58.9 por ciento de la población de 18 años y más considera la inseguridad como el problema más importante que aqueja actualmente, le siguen el desempleo con 41.5 % y la salud con 40.2 %.

Respecto a los crímenes sexuales, por cada 8 delitos sexuales cometidos contra mujeres, hay 1 delito sexual cometido a hombres [7]. La Unicef reconoce a la violencia sexual como una de las formas de violencia más extrema que sufren las personas en especial las mujeres y las niñas. En el mismo documento la define como cualquier acto sexual cometido en contra de la voluntad de otra persona [18].

Por otro lado, en la mayoría de las ciudades, el crimen está altamente concentrado en un número pequeño de lugares generalmente conocidos como puntos críticos o puntos calientes (hot spots) del crimen, es decir, mientras que algunos lugares apenas registran incidentes, en otros se acumulan los delitos. El estudio de [11] encontró que la mitad de los delitos de cinco grandes ciudades de Colombia se concentran entre el 3 % y 5 % de las calles.

Por esta razón, una de las estrategias para resolver el problema de la inseguridad es dirigir los recursos policiales hacia estos lugares. En general, el término hot spot hace referencia a un lugar donde existe alta concentración de un fenómeno, en este caso el fenómeno delictivo¹. Un hot spot delictivo es una zona donde se comete un número de delitos superior a la media de una ciudad, o una zona donde el riesgo de ser víctima de un delito supera al promedio [6].

El análisis de hot spots mediante mapas del delito se ha convertido en una técnica analítica comúnmente utilizada por la policía para identificar visualmente los lugares donde el número de delitos tiende a ser más alto, facilitando así la toma de decisiones respecto a dónde dirigir recursos [4]. En el presente trabajo se emplea el agrupamiento basado en densidad con un conjunto de datos de crímenes sexuales, cometidos en la Ciudad de México (CDMX) de los años 2019 a 2021, para la detección de hot spots delictivos.

Los datos contienen la longitud y latitud de geolocalización lo que permite localizar el lugar donde se cometió el delito y visualizar geográficamente la actividad criminal en mapas. La metodología se centra, por un lado, en aplicar el método OPTICS a los datos espaciales para determinar los patrones en el tiempo y el espacio encontrando las zonas con agrupaciones más densas de puntos en los mapas; y por otro, en estimar el área y la magnitud de incidencia delictiva presentada en dichas zonas.

El artículo está organizado de la siguiente manera. En la sección 2 se exponen los trabajos relacionados. La sección 3 presenta la propuesta de solución donde se explica la metodología que se siguió. En la sección 4 se describen las pruebas con el modelo propuesto. Posteriormente, en la sección 5 se analizan y discuten los resultados obtenidos de los experimentos. Finalmente, la sección 6 muestra las conclusiones y el trabajo futuro.

¹ También existe el término cold spot o punto frío que se refiere a sitios con baja concentración de un fenómeno.

2. Trabajo relacionado

Desde que se creó el primer mapa de hot spots criminales en [15] (citado por [1]) se concluyó que los crímenes no se distribuían de manera homogénea a lo largo de las ciudades, sino que, se concentraban en ciertas zonas; también afirmaron que el crimen está determinado por las características generales de la estructura social (factores socioeconómicos, espaciales y poblacionales).

Más tarde, la revisión sistemática de [3] que incluyó 19 estudios, 17 de los cuales se realizaron en Estados Unidos, uno en Australia y otro en Argentina, demostró una reducción general de la delincuencia cuando se implementaron acciones policiales en lugares de alta actividad criminal; la revisión también sugiere que invertir los recursos humanos limitados de las policías en el control de hot spots evitará la delincuencia en estas áreas y las áreas circundantes, reduciendo la delincuencia en su conjunto.

Sin embargo, los resultados del análisis en [12] evidenciaron diversos factores que evitaron que el modelo de seguridad ciudadana basada en los hot spots redujera los niveles de delincuencia en la ciudad de Bogotá, además, señala que los resultados obtenidos (buenos o malos) de la aplicación del modelo dependen en gran medida de la forma en que se implementen las tácticas policíacas en las zonas críticas.

En el caso de México, en [5] se realizó un análisis exploratorio de datos espaciales para conocer la distribución geográfica de los delitos pecuniarios² en la Zona Metropolitana del Valle de México donde explorando distintos factores se corroboró la relación de las variables socioeconómicas y urbanas con este tipo de crímenes.

Adicionalmente, los mapas comprobaron que los crímenes no se distribuyen de forma aleatoria en el espacio por lo que se pueden identificar patrones espaciales (hot spots y cold spots). Para verificar la significancia de los clústers encontrados se hizo un análisis de indicadores locales de asociación espacial (LISA) para la variable de delitos pecuniarios y un modelo espacial de regresión para conocer las variables que influyen en los delitos cometidos.

Por su parte, en [8] se identificaron hot spots de delitos callejeros en la CDMX mediante el uso de la técnica de análisis espacial densidad de kernel encontrando patrones espacio temporales del delito callejero; especialmente las zonas críticas, se asocian con la actividad comercial y laboral de la centralidad de la ciudad, pues en la periferia, en promedio, se presenta una menor actividad de este tipo de hechos.

De igual manera, los horarios de incidencia se relacionan con la salida y regreso a casa, es decir, por las mañanas y por la noche. Otro trabajo, es el de [9], el cual identificó patrones de agrupamiento espacial (hot spots) de las denuncias por violencia familiar en Ciudad Victoria (del estado de Tamaulipas en México) e ilustró su relación con las desventajas del vecindario.

También, desarrolló un análisis mediante las técnicas: densidad de kernel, promedio del vecino más cercano, indicadores locales de autocorrelación espacial y regresión logística; confirmando que la cantidad de ocupantes por vivienda y los hogares con jefatura femenina tienen una relación positiva con la probabilidad de estar en un hot spot de violencia familiar.

² Los crímenes pecuniarios son los que se realizan para obtener una retribución económica [5].

En contraste con los enfoques presentados en los trabajos citados, este trabajo, se centra en la aplicación del aprendizaje automático (machine learning, en Inglés) para la detección de hot spots, específicamente del aprendizaje no supervisado donde el algoritmo por sí mismo intentará aprender de los datos con el fin de encontrar patrones o tendencias.

Una categoría del aprendizaje no supervisado es el agrupamiento o clustering cuyo objetivo es encontrar cómo agrupar naturalmente los datos de forma que los elementos del mismo grupo sean más similares entre ellos que aquellos elementos de diferentes grupos [13]. Se experimenta con el método de agrupamiento OPTICS, basado en densidad, para determinar patrones de delitos sexuales: acoso sexual, abuso sexual y todos los tipos de violación sexual, cometidos en la CDMX en el periodo 2019-2021.

3. Propuesta metodológica

Entre los métodos de agrupamiento, los algoritmos basados en densidad enfocan el problema de dividir el conjunto de datos en grupos o clusters teniendo en cuenta la distribución de densidad de los puntos. Estos métodos son importantes debido a su capacidad para encontrar grupos de formas arbitrarias.

DBSCAN (Density Based Spatial Clustering of Applications with Noise) fue el primer algoritmo de agrupamiento basado en densidad [10], su idea básica es: dado un conjunto de puntos en algún espacio, agrupa los puntos que forman una región densa en un cluster, marcando como puntos de ruido los puntos que se encuentran en regiones de baja densidad. El método OPTICS es una generalización o mejora de DBSCAN que aborda una de sus principales debilidades que es el problema de detectar agrupaciones significativas en datos de densidad variable.

3.1. Método de agrupamiento OPTICS

El algoritmo OPTICS (Ordering Points To Identify the Clustering Structure), presentado en [2], genera un orden de los grupos (clusters) basado en la densidad, en el cual los grupos con mayor densidad están completamente contenidos en otros grupos menos densos. Este algoritmo recibe dos parámetros, $\max Eps$ y $\min Pts$, los cuales representan la distancia máxima de generación de grupos y el número mínimo de puntos que debe contener la vecindad de radio $\epsilon < \max Eps$ con centro en un punto p , como se puede ver en la figura 1, para que este tenga posibilidad de ser parte de un grupo.

Lo que diferencia a este algoritmo de otros métodos de agrupación es que no asigna etiquetas de grupo a los puntos, en cambio, almacena el orden en el que los puntos son procesados y dos características clave de cada punto p : la distancia a sus puntos vecinos ($core(p)$) y la distancia de alcance ($r(p)$), las cuales indican qué tan condensados están los vecinos de p y qué tan cerca está p de su predecesor o , respectivamente (ver figura 1). Esta información después puede ser utilizada para generar grupos de forma automática.

En [2] también se propone un método para agrupar los puntos de manera automatizada detectando fronteras tentativas de inicio y fin del cluster usando la distancia de alcance y el orden generado por OPTICS.

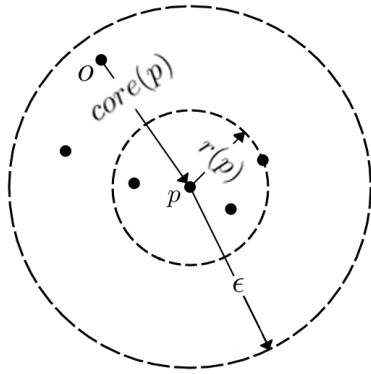


Fig. 1. Distancias a los vecinos y de alcance con $\text{MinPts} = 4$.

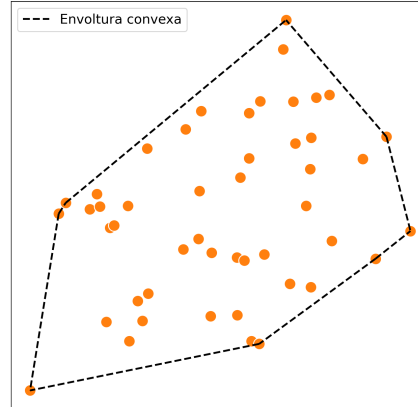


Fig. 2. Envoltura convexa de un conjunto de puntos.

En este algoritmo se introduce un nuevo parámetro, ξ , el cual controla la estrictez de las fronteras entre grupos y entre grupo y ruido. La implementación del lenguaje Python de OPTICS disponible en la biblioteca scikit-learn [16] implementa esta forma de agrupación y recibe los parámetros `min_samples`, `max_eps` y ξ , entre otros, donde `min_samples` es equivalente a MinPts .

En este trabajo, se eligió el algoritmo de agrupamiento OPTICS porque es muy versátil y eficaz al determinar los grupos, puesto que no asume que los clusters tengan una forma predefinida y toma en cuenta la posible existencia de ruido, el cual comúnmente se presenta en conjuntos de datos reales [2].

Así mismo, existen técnicas de visualización basadas en el orden generado por OPTICS y las distancias de alcance, las cuales permiten visualizar claramente los resultados obtenidos.

3.2. Concepto de hot spot

En [6] se define a un hot spot delictivo como una zona geográfica cuya incidencia delictiva es mayor a la del promedio, o una zona donde las personas tienen un riesgo de victimización más alto que la media. También sugiere la existencia de puntos fríos (cold spots) o zonas con menor cantidad de crímenes que el promedio y señala que algunos hot spots pueden ser más calientes que otros, es decir, varían en qué medida por encima de la media están.

El concepto de hot spot en el contexto de la detección de zonas altamente delictivas puede entenderse como una zona donde el número de casos por unidad de área sea mayor al número de casos por área promedio. En trabajos previos, este enfoque suele ser utilizado en conjunto con la división territorial de un mapa para determinar las zonas con mayor actividad criminal [5, 8, 9].

A pesar de que los hot spots obtenidos a partir de una división territorial predeterminada pueden ser útiles para determinar zonas de alta incidencia delictiva, estos están restringidos por la misma división de donde fueron producidos.

En teoría, es completamente posible que existan zonas geográficas de formas no convencionales con un mayor número de casos por área que el promedio, las cuales serían pasadas por alto si se presupone, explícita o implícitamente, que los hot spots tienen relación con una división predefinida del mapa.

Al aplicar OPTICS a un conjunto de coordenadas geográficas de casos de un cierto conjunto de delitos se obtienen grupos o clusters separados por zonas de ruido. Los grupos obtenidos mediante este procedimiento pueden tener una forma arbitraria dado que son detectados a través del cambio en la densidad de los puntos en el mapa.

Así mismo, se puede estimar el área que abarca cada grupo al calcular la envoltura convexa de dicho conjunto de puntos, esto es, al determinar el polígono que contiene a todos los puntos del cluster y cuyos vértices también pertenecen al cluster en cuestión, como se ilustra en la figura 2. Para efectuar el cálculo del área de la envoltura convexa se utilizó el módulo `scipy` [14] de Python, en especial la clase `ConvexHull`.

Al estimar el área de un cluster con un polígono, puede estimarse su número de casos por área y compararse con el número total de casos entre el área total del mapa para precisar si esta zona cumple con la definición convencional de hot spot anteriormente mencionada.

Desde esta perspectiva, si se define el número de casos por unidad de área del cluster como α y al número total de casos entre el área de estudio total como β , el cociente α/β es un indicador de la incidencia delictiva de la zona, puesto que si este es menor a uno, entonces la zona tiene una incidencia delictiva menor al promedio, si es igual a uno, entonces tiene una incidencia delictiva promedio y si es mayor a uno, entonces tiene una incidencia delictiva mayor al promedio, o dicho de otra manera, es un hot spot.

Además de ser un indicador de la incidencia delictiva, el cociente α/β también indica la magnitud del hot spot, la cual puede interpretarse como el número de veces en que la incidencia delictiva del cluster supera al promedio. Esta medida no solo permite detectar los hot spots, también permite compararlos entre sí y clasificar su gravedad o prioridad.

4. Experimentación

4.1. Área de estudio

El área de estudio en el que se desarrolla este trabajo corresponde a la Ciudad de México (CDMX), la entidad federativa con la mayor densidad demográfica. La CDMX, capital del país, ocupa una superficie de 1495 km², equivalente al 0.1 por ciento del territorio total nacional y se estima que en ella existen más de 2.6 millones de viviendas en las que residen casi 9 millones de personas, esto significa que aproximadamente 8 de cada 100 viviendas de las que hay en el país se ubican en esta ciudad y 7 de cada 100 habitantes residen también aquí [17].

La ciudad también cuenta con el número más alto de policías por habitante del país (9.6 policías por cada mil habitantes). La ciudad de México está dividida en dieciséis demarcaciones territoriales.

Tabla 1. Delitos sexuales geolocalizados cometidos por año.

Tipo de delito	2019	2020	2021	Total
Abuso sexual	3,451	2,758	3,162	9,371
Acoso sexual	831	934	1,219	2,984
Violación	1,263	1,366	1,462	4,091
Violación equiparada	239	388	516	1,143
Tentativa de violación	16	18	29	63
Violación tumultuaria	32	22	16	70
Violación equiparada por conocido	21	21	17	59
Violación equiparada y robo de vehículo	3	2	4	9
Violación tumultuaria equiparada por conocido	1	0	0	1
Violación tumultuaria equiparada	1	1	1	3

4.2. Conjuntos de datos

Para los experimentos se utilizaron los datos de las víctimas de los delitos en las carpetas de investigación de la Fiscalía General de Justicia (FGJ) de la Ciudad de México³, datos con coordenadas geográficas (longitud y latitud) de crímenes cometidos en las 16 delegaciones de la CDMX en los años 2019, 2020 y 2021⁴.

De acuerdo con los datos, en dicho periodo se cometieron 683,129 delitos del fuero común y de estos hechos 20,742 fueron clasificados como delitos sexuales: acoso, abuso y violación sexual. Sin embargo, el número de crímenes sexuales con coordenadas geográficas disponibles, utilizados en los experimentos, es de sólo 17,794.

La Tabla 1 muestra que en todos los años el delito que más se cometió fue el abuso sexual con un 52.66 % del total, en segundo puesto se encuentran los delitos de violación con un 22.99 % y en tercer lugar los delitos de acoso sexual con un 16.77 %. Es importante señalar que el delito de violación se clasifica en varios tipos los cuales representan el 7.58 % del total de los crímenes de este tipo cometidos en la CDMX.

Respecto a las alcaldías, el número total de delitos de tipo sexual denunciados en cada una de las delegaciones en el periodo 2019-2021 es el siguiente: Iztapalapa (IZTP), primer lugar, con 2,892; luego Cuauhtémoc (CUAU) que tiene 2,341 casos; sigue Gustavo A. Madero (GAM) con 2,090; Álvaro Obregón (AO) 1,284; Tlalpan (TLAL) 1,194; Venustiano Carranza (VC) 1,124; Coyoacán (COYO) 1,073; Miguel Hidalgo (MH) 939; Benito Juárez (BJ) 932; Xochimilco (XOCH) 848; Iztacalco (IZTC) 724; Tláhuac (TLAH) 658; Azcapotzalco (AZ) 655; Magdalena Contreras (MC) 411; Cuajimalpa (CUAJ) que suma 355 y Milpa Alta (MA) registra 274.

La información de las carpetas de investigación también permite calcular la frecuencia o número de casos ocurridos, a nivel de demarcación, de los distintos tipos de crímenes sexuales.

³ Actualización del 23 de marzo del 2022, disponibles en el Portal de Datos Abiertos de la CDMX en <https://datos.cdmx.gob.mx>

⁴ Sólo se consideran estos años porque son los únicos que tienen un registro completo de actos delictivos, los años del 2018 hacia atrás tienen menos de 575 eventos y el 2022 es el año en curso.

En el mapa de calor mostrado en la figura 6b se puede observar la frecuencia de incidencia para el delito de abuso sexual, de acuerdo con la intensidad del color, para todo el periodo 2019-2021. Con este mapa se puede ver que la frecuencia es especialmente alta (polígonos color rojo y naranja) en Iztapalapa, Cuauhtémoc y Gustavo A. Madero, considerando que la ocurrencia promedio de este tipo de delito es de 585.6 casos por alcaldía.

Del otro extremo, en Magdalena Contreras, Cuajimalpa y Milpa Alta, la frecuencia es baja (polígonos color azul oscuro) e incluso es inferior al promedio. Cabe señalar que los datos utilizados poseen un sesgo, pues no reflejan la situación real del delito en la CDMX, ya que provienen de los delitos denunciados en los Ministerios Públicos y no tienen en cuenta los delitos cometidos que no se denuncian o denuncias que no derivan en el inicio de carpetas de investigación.

En [7] se estima que la cifra negra de delitos no denunciados a nivel nacional en 2020 fue de 93.3 por ciento, aproximadamente sólo se denuncian 7 de cada 100 delitos. Entre las razones para no denunciar destaca la pérdida de tiempo con 33.9 % y la desconfianza en las autoridades con 14.2 %.

4.3. Pruebas

Los resultados obtenidos al aplicar OPTICS al conjunto de datos de coordenadas geográficas filtrado por el tipo de delito pueden visualizarse en dos gráficas: un mapa en el que se puede observar la división territorial, los clusters encontrados y sus respectivas envolturas convexas y, la gráfica de alcance correspondiente, en la cual se resaltan los clusters con distintos colores para diferenciarlos.

La figura 5 muestra los resultados para casos de acoso sexual en el año 2020. La gráfica de alcance representa la estructura generada por OPTICS en el orden en que se procesaron los puntos; el eje de las abscisas indica el orden de procesamiento de los puntos y el eje de las ordenadas representa la distancia de alcance.

De acuerdo con [2], en esta gráfica los clusters pueden visualizarse como muescas o depresiones, las cuales, mientras más pronunciadas sean, indican una mejor definición del cluster asociado. Del mismo modo, mientras más bajo sea el alcance de los puntos de un determinado cluster, este es más denso.

Para poder realizar pruebas con los datos se deben asignar valores específicos a los parámetros de OPTICS, es decir, se deben especificar los valores de min_samples , max_eps y ξ . Se utilizará el valor por defecto de max_eps indicado en la documentación de [16], indefinido, el cual es el valor más adecuado según [2] y [16] si no se tiene un conjunto de datos muy grande.

Dado que ξ representa la estrictez de las fronteras entre clusters, se hicieron pruebas con diferentes valores para ξ : 0.005, 0.01, 0.02, 0.03, 0.04 y 0.05, en las cuales se encontró que un valor bajo de ξ detecta grupos pobremente definidos mientras que un valor alto solamente detecta grupos extremadamente bien definidos, situación poco común al tratar con datos reales.

Las observaciones anteriormente mencionadas, ilustradas en la figura 3 para delitos de acoso sexual en el 2020, coinciden con lo encontrado en [2], donde se muestra una aplicación de OPTICS en un conjunto de datos real en la que se utiliza el valor $\xi = 0.02$ para encontrar clusters razonablemente bien definidos.

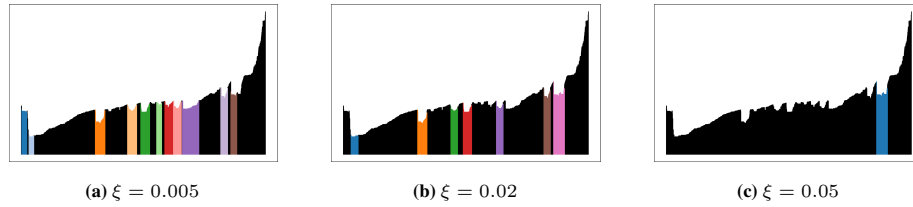


Fig. 3. Gráficas de alcance con distintos valores de ξ para acoso sexual en 2020.

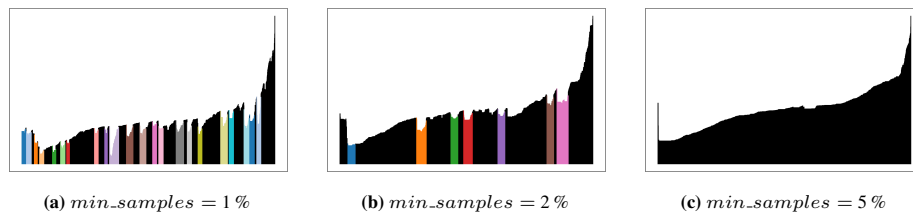


Fig. 4. Gráficas de alcance con distintos valores de $min_samples$ para acoso sexual en 2020.

Por las razones anteriormente mencionadas se usará $\xi = 0.02$ para obtener los resultados finales. En cuanto al valor de $min_samples$, se cree que lo más adecuado es asignar un valor en proporción al número de observaciones en el conjunto de datos por analizar, de manera que se obtengan resultados comparables entre los diferentes conjuntos de delitos estudiados; también, se debe notar que el tamaño mínimo de cluster es mayor o igual al valor de $min_samples$ utilizado, por lo cual $min_samples$ no debe tener un valor muy alto si se quiere evitar pasar por alto algunos hot spots menores, es decir, zonas con una razón de delitos por área no muy alta.

En este trabajo se usó el 2% del total de las observaciones como valor para $min_samples$ debido a que este valor fue el que dio mejores resultados entre los valores 1%, 2%, 3%, 4% y 5%, como se ilustra en la figura 4, de nuevo, para delitos de acoso sexual en el año 2020.

5. Análisis y resultados

En la figura 5a se puede apreciar la geolocalización de los casos de acoso sexual reportados en el año 2020 junto con las envolturas convexas de los clusters encontrados sobre el mapa de las alcaldías de la Ciudad de México, mientras que la gráfica de alcance correspondiente puede observarse en la figura 5b.

Los cocientes α/β de los clusters 1 al 9 de la figura 5 son: 2.4037, 39.8669, 5.6725, 5.6116, 1.7247, 2.9072, 2.1555, 1.5048 y 1.2387, respectivamente, por lo que todos los clusters también son clasificados como hot spots.

Como puede verse en la figura 5a, a simple vista es difícil determinar la existencia de hot spots y delimitar sus respectivas fronteras. Por ejemplo, se puede observar que la región de puntos alrededor del cluster 2 de la figura 5 es más densa que cualquier otra zona del mapa, sin embargo, no toda la región es detectada como hot spot.

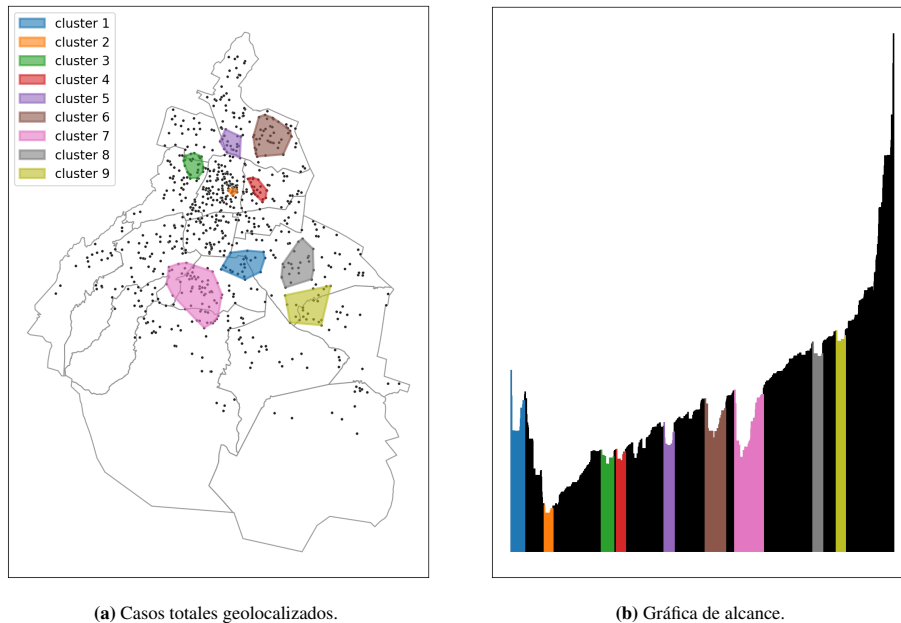


Fig. 5. Resultados de la metodología propuesta en casos de acoso sexual del 2020.

Esta situación ilustra otra característica de los hot spots encontrados mediante la metodología propuesta, los casos que componen la región del hot spot tienden a agruparse significativamente en comparación con sus alrededores, propiedad que no es posible detectar al usar solamente la noción clásica de hot spot.

Las figuras 5a y 5b muestran que aunque los datos sean difíciles de tratar y visualmente no se detecten posibles hot spots, mientras la estructura de los datos en la gráfica de alcance presente muescas relativamente bien definidas y del tamaño mínimo, entonces los datos tienden a agruparse en ciertas zonas del mapa.

De igual manera, si la estructura de la gráfica de alcance no presenta las características anteriormente descritas, entonces la metodología propuesta no detectará indiscriminadamente cualquier hendidura como hot spot.

Con relación a los casos de abuso sexual reportados en los años 2019, 2020 y 2021 se presentan las figuras 6a, clusters geolocalizados en un mapa de la CDMX, y 6b, un mapa de calor de la frecuencia de delitos por alcaldía. Los valores α/β de los clusters 1, 2 y 3 son: 1.9818, 1.3877 y 1.5619, respectivamente, indicando que los clusters encontrados efectivamente son hot spots.

El cluster 1 (en color azul), correspondiente al 4.11 % de los casos, se presenta en la zona este de las alcaldías Venustiano Carranza e Iztacalco; el cluster 2 (en color naranja), correspondiente al 2.29 % de los casos, está completamente contenido en la alcaldía Gustavo A. Madero y localizado en la zona norte de dicha alcaldía; finalmente, el cluster 3 (en color verde) corresponde al 2.08 % de los casos reportados y la región abarcada por este atraviesa la frontera entre las alcaldías Álvaro Obregón y Magdalena Contreras.

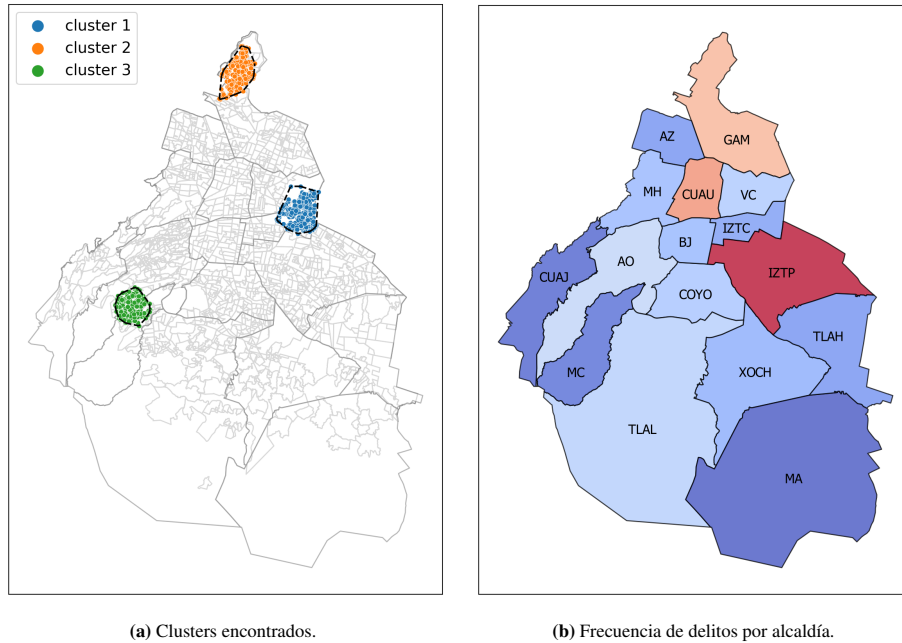


Fig. 6. (a) Resultados de aplicar OPTICS y (b) Mapa de calor de delitos por alcaldía, ambos, para casos de abuso sexual reportados en 2019, 2020 y 2021.

La figura 6 muestra un escenario en el cual la metodología propuesta podría no detectar hot spots pobremente definidos pero altamente densos, como lo son las alcaldías Cuauhtémoc e Iztapalapa, claramente reconocibles en la figura 6b por sus colores naranja oscuro y rojo, respectivamente.

Así mismo, un mapa de calor no es capaz de identificar las aglomeraciones de casos encontradas por OPTICS entre delegaciones, como los clusters 2 y 3, puesto que estas regiones son fragmentadas por la división territorial predefinida de las alcaldías.

Este resultado sugiere que el uso de OPTICS para detectar hot spots se beneficiaría de un contraste con un análisis tradicional para lograr considerar un panorama más amplio del contexto de los datos. De igual manera, la situación ilustrada por la figura 6 muestra que el enfoque menos restrictivo de la metodología propuesta al detectar hot spots ilumina zonas de interés posiblemente pasadas por alto al efectuar únicamente un análisis más tradicional.

Por otra parte, la figura 7 presenta una aplicación interesante del procedimiento propuesto en este trabajo, incorporando el factor temporal al problema de detectar hot spots y zonas de interés. En esta figura se muestran los clusters encontrados al aplicar OPTICS a los casos de violación y sus variantes en cada año por separado, tiñendo los hot spots de un color distinto para cada uno de ellos, azul para 2019, naranja para 2020 y verde para 2021.

Las parejas (min, max) de valores α/β para los clusters encontrados en cada año son: (1.3802, 8.7994), (1.1024, 6.0884), y (1.2885, 6.6621) para 2019, 2020 y 2021, respectivamente, confirmando que todos los clusters encontrados son hot spots.

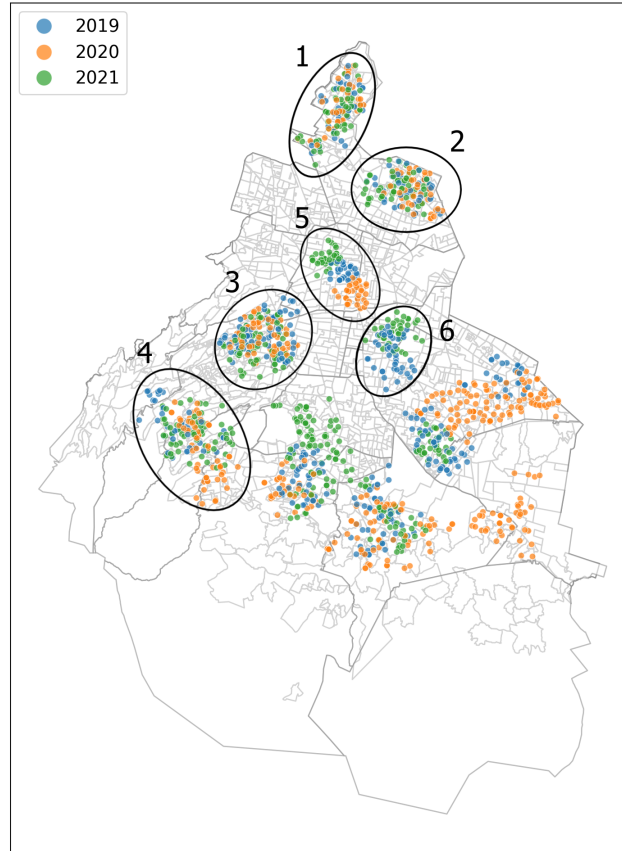


Fig. 7. Gráfica de dispersión con clusters sobrepuestos encontrados al aplicar OPTICS a casos de violación y sus variantes en 2019, 2020 y 2021 por separado.

Las regiones resaltadas y enumeradas en la figura 7 ejemplifican tres distintas evoluciones de interés que un hot spot podría exhibir al paso del tiempo. Los hot spots de las regiones 1, 2 y 3 son muy similares para los tres años de estudio, razón por la cual estos hot spots podrían considerarse recurrentes; la repetición de eventos en los mismos lugares indica que las zonas que concentran la mayor incidencia de violencia sexual en la ciudad permanecen como zonas peligrosas durante los tres años.

Las regiones 4 y 5 ilustran lo que se conoce como un desplazamiento, el cual puede ser más o menos pronunciado; en el caso del hot spot de la región 5, el desplazamiento es fácilmente distinguible incluso a simple vista, este desplazamiento significa que los crímenes de violación se han dispersado o distribuido a otras zonas de la CDMX que antes no registraban este tipo de delito.

La situación presentada en la región 6 es el resurgimiento de un hot spot mínimamente desplazado, este se detecta en el año 2019, en el año 2020 desaparece y en el año 2021 reaparece ligeramente más al norte.

Las interpretaciones de las tres situaciones anteriormente mencionadas pueden ser variadas dependiendo del contexto en el que se presenten, no obstante, posibles explicaciones para estos fenómenos podrían ser la falta de estrategias para asignar recursos a zonas cuya incidencia delictiva sea mayor al promedio, la adaptación de los criminales a la estrategia de seguridad pública vigente y el descuido de regiones previamente consideradas de atención después de reducir la actividad delictiva exitosamente.

6. Conclusiones y trabajo futuro

Los resultados mostrados confirman que los delitos no se distribuyen uniformemente en el espacio y que, en el periodo de estudio, las denuncias por violencia sexual presentan un patrón de agrupamiento espacial.

Estos hallazgos son importantes porque el estudio se ha abordado desde la perspectiva del aprendizaje automático no supervisado a través de técnicas de agrupamiento basado en densidad. Aunque no sea un enfoque frecuentemente utilizado, en este trabajo se obtuvieron buenos resultados al aplicar el algoritmo de agrupamiento OPTICS con los valores de los parámetros elegidos.

Se comprobó que los hot spots encontrados mediante la metodología propuesta presentan las características deseadas por la definición clásica de hot spot así como una agrupación significativa de los casos en comparación con sus alrededores. También se encontró que un estudio híbrido entre los enfoques tradicional y de aprendizaje automático podría ser más informativo que cualquiera de los dos por separado, ya que cada uno obtiene resultados con una interpretación distinta y complementaria a los del otro tipo de estudio.

El trabajo actual se concentró en explorar las características espaciales de los delitos sexuales en la CDMX. No obstante, los factores que influyen en la espacialidad del fenómeno delictivo requiere un análisis más detallado en un trabajo futuro. Así mismo, a futuro, se buscaran más fuentes de información para considerar no sólo la cifra oficial de denuncias sino también la cifra negra.

Finalmente, la información obtenida podría ser evaluada por los departamentos de policía y servir como apoyo para tomar decisiones más informadas en materia de seguridad pública, permitiendo optimizar la asignación de recursos humanos y financieros; asimismo, la metodología propuesta se podría utilizar para detectar hot spots de interés con formas no convencionales y dar seguimiento a la evolución de dichas zonas al paso del tiempo.

Referencias

1. Anderson, T. L.: Understanding deviance: connecting classical and contemporary perspectives. pp. 106–127 (2014) doi: 10.4324/9781315879635
2. Ankerst, M., Breunig, M., Kriegel, H. P.: Optics: Ordering points to identify the clustering structure. *ACM Sigmod Record*, ACM, vol. 28, no. 2, pp. 49–60 (1999) doi: 10.1145/304181.304187
3. Braga, A., Papachristos, A., Hureau, D.: Hot spots policing effects on crime. *Campbell Systematic Reviews*, vol. 8, no. 1, pp. 1–96 (2012) doi: 10.4073/csr.2012.8

4. Chaney, S., Tompson, L., Uhlig, S.: The utility of hotspot mapping for predicting spatial patterns of crime. *Security Journal*, vol. 21, no. 1, pp. 4–28 (2008) doi: 10.1057/palgrave.sj.8350066
5. Corzo, T. J. A.: Criminalidad y sus determinantes socioeconomicos en la Zona Metropolitana del Valle de México (2011-2014): Protocolo de Investigación. Facultad de Estudios Superiores Acatlán, UNAM (2018)
6. Eck, J. E., Chainey, S., Cameron, J. G., Leitner, M., Wilson, R. E.: Mapping crime: Understanding hot spots. National Institute of Justice (2005)
7. INEGI: Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública (ENVIPE). Instituto Nacional de Estadística y Geografía (2021) <https://www.inegi.org.mx/programas/envipe/2021/>
8. Flores-Gutiérrez, S.: Análisis espacial del delito callejero en Ciudad de México, 2018. *Quivera Revista de Estudios Territoriales*, UNAM, vol. 23, no. 1, pp. 25–47 (2021) doi: 10.36677/qret.v23i1.15072
9. Jurado-Flores, V. D., Fuentes-Rios, F. J.: Concentración espacial de las denuncias por violencia familiar en Ciudad Victoria: un acercamiento mediante la detección de hot spots. *CienciaUAT*, vol. 16, no. 2, pp. 126–140 (2022) doi: 10.29059/cienciauat.v16i2.1551
10. Ester, M., Kriegel, H. P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96). AAAI Press, vol. 96, no. 34, pp. 226–231 (1996)
11. Mejía, D., Ortega, D., Ortiz, K.: Un análisis de la criminalidad urbana en Colombia. CAF Banco de Desarrollo de América Latina - Documento de trabajo, no. 06 (2015)
12. Ramírez, A.: Los hot spots: una estrategia fallida para Bogotá. *Desafíos*, vol. 28, no. 2, pp. 255–296 (2016) doi: 10.12804/desafios28.2.2016.07
13. Raschka, S., Mirjalili, V.: Python machine learning: Aprendizaje automático y aprendizaje profundo con Python, scikit-learn y TensorFlow. Marcombo (2019)
14. Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser W., Bright, J., et al.: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, vol. 17, no. 3, pp. 261–272 (2020) doi: 10.1038/s41592-019-0686-2
15. Shaw, C. R., McKay, H. D.: Juvenile delinquency and urban areas. University of Chicago Press (1942)
16. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al.: Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830 (2011)
17. EVALÚA: Ciudad de México 2020: Un diagnóstico de la desigualdad socio territorial. Consejo de Evaluación del Desarrollo Social de la Ciudad de México (EVALÚA) (2020)
18. Unicef: ONU México hace un llamado a eliminar todas las formas de violencia sexual contra las mujeres y las niñas: Comunicado de prensa. Unicef México (2019)