

## Análisis de patrones de criminalidad por medio de redes espacio-temporales de eventos

Miguel Quintero, Juan Sosa y Martha Bohórquez<sup>1</sup>

Universidad Nacional de Colombia

Rafael Rentería-Ramos

Universidad Nacional Abierta y a Distancia

### RESUMEN

En este trabajo se presenta una metodología para el análisis espaciotemporal de eventos utilizando redes. Se construye una red de eventos considerando cada evento como un vértice y las relaciones espaciotemporales entre ellas como aristas. Se estudia la estructura de la red mediante la identificación de patrones recurrentes que describen comportamientos emergentes del sistema, conocidos como *motifs*. Para la detección de *motifs*, se diseña un algoritmo que permite aproximar la distribución empírica de los conteos de *motifs*, combinando métodos de simulación de redes espacio temporales, algoritmos de agrupamiento y la fórmula de Stirling por radio espacial y ventana de tiempo definidos. Se observa que una distribución log-normal se ajusta adecuadamente a la distribución del grado de la red de eventos, permitiendo definir la distancia en la que dos sucesos pueden estar relacionados en el espacio dentro de una ventana de tiempo. La metodología se ilustra aplicándola al caso de hurtos a personas reportados al departamento de policía local entre 2018 y 2021 en la ciudad de Pereira, Colombia. Los resultados evidencian que la metodología propuesta es eficaz en la identificación de *motifs* que capturan patrones espaciotemporales generando información que permite desarrollar estrategias de prevención de la delincuencia.

**Palabras clave:** Patrón espaciotemporal, criminalidad, *motifs*, red de eventos.

## A unified approach to the study of spatio-temporal relational patterns

### ABSTRACT

This work presents a methodology for the spatio-temporal analysis of events using complex networks. A network of events is constructed considering each event as a vertex and the spatio-temporal relationships between them as edges. The network structure is studied by identifying recurring patterns that describe emerging behaviors of the system, which are known as motifs. For the detection of motifs, we design an algorithm that allows us to approximate the empirical distribution of motif counts, combining simulation methods of spatiotemporal networks, clustering algorithms and the Stirling formula for defined spatial radius and time window. It is observed that a log-normal distribution adequately fits the degree distribution of the event network, allowing to define the distance in which two events can be related in space within a time window. The methodology is illustrated by applying it to the case of thefts from people reported to the local police department from 2018 to 2021 in the city of Pereira, Colombia. The results show that the proposed methodology is effective in identifying motifs that capture spatio-temporal patterns, generating information that allows developing crime prevention strategies.

**Key words:** Spatio-temporal pattern, crime, motifs, network of events.

<sup>1</sup> Contacto con los autores: Martha Bohórquez, [mpbohorquezc@unal.edu.co](mailto:mpbohorquezc@unal.edu.co)

## INTRODUCCIÓN

Las redes (las redes analizadas en este trabajo presentan características estructurales y dinámicas no triviales) han cobrado especial importancia gracias a su gran capacidad de adaptación en diferentes temáticas (Estrada, 2012). Sus aplicaciones van desde la física (Dorogovtsev, Goltsev, & Mendes, 2008) hasta la formación de opinión (Sznajd-Weron, 2020) o cualquier sistema que requiera ser estudiado en términos relacionales (Barabási A.-L. &, 1999) y (Newman M. , 2018). Si bien sus orígenes se remontan a principios del siglo XIX (resultados y metodologías fundamentales se pueden consultar en los trabajos recopilatorios de (Wernicke, 2006) y (Boccaletti, y otros, 2014), solo en los últimos 20 años el estudio de la complejidad ha cobrado especial importancia. Así, el análisis de redes tiene raíces interdisciplinarias y ha ganado especial relevancia desde el siglo XX. A finales de los años 90, evolucionó hacia el estudio de sistemas dinámicos con propiedades emergentes, superando un enfoque restringido a conexiones estáticas. Este cambio dio lugar a una desafortunada separación conceptual entre la complejidad, centrada en dinámicas globales, y el análisis de redes, que desarrolló herramientas específicas para abordar estructuras, funciones y dinámicas en sistemas como redes sociales, biológicas y tecnológicas, consolidándose como una disciplina independiente.

Existen algunas propuestas para incorporar la dimensión temporal en las redes. Por ejemplo, las redes multicapa consideran el factor temporal de las conexiones (Boccaletti, y otros, 2014), considerando cada conjunto de eventos que ocurren dentro de una ventana de tiempo, como una capa diferente del sistema. Así, desde la perspectiva dinámica del análisis de redes multicapa, hoy en día hay una gran variedad de propuestas para modelar la evolución del sistema a través del tiempo incluyendo los trabajos de (Durante, 2014), (Higdon P. D., 2015), (Sewell & Chen, 2017), (Gupta, Sharma, & Dukkipati, 2019), (Kim, Lee, Xue, & Niu, 2018) y (Betancourt, Rodriguez, & BOYD, 2020), entre otros.

Otra área muy importante en este contexto se refiere a las técnicas para detectar patrones espaciotemporales a partir del descubrimiento de subgrafos que ocurren con una frecuencia particular en una red, lo que contribuye a comprender la emergencia de clústeres o buffers espaciales en ventanas de tiempo particulares que son cruciales para predecir la probabilidad de ocurrencia de nuevos eventos. Estos subgrafos con altas frecuencias de

ocurrencia son conocidos como motifs. Por ejemplo, (Davies & Marchione, 2015), (Atluri G, 2018), (Pasquaretta C, 2021), (Oberoi KS, 2021), (Dang TA, 2018) son una excelente selección de artículos relacionados con métodos para detectar patrones utilizando subgrafos y sus propiedades. (Jazayeri A, 2020) presentan una revisión completa y actualizada de algoritmos de descubrimiento de motifs. Además, los motifs han permitido la generación de importantes aplicaciones para la toma de decisiones en dinámicas poblacionales, entre las que se destacan la epidemiología, la movilidad y la criminología.

(Grubescic TH, 2008) adaptan la idea de la prueba de Knox para construir redes variando el radio espacial y temporal, obteniendo vínculos de similitud de ubicación (Borgatti SP, 2009). El análisis de redes aborda la organización delictiva (Scott J, 2011), un aspecto clave en el estudio de actos ilegales.

El enfoque de redes ha sido útil también para el escrutinio de grupos terroristas con fines de captura de delincuentes (Borgatti SP, 2009).

Su procedimiento permite observar la cercanía espaciotemporal y la forma de los delitos, es decir, los motifs, y encontrar patrones con un enfoque fijo. (da Cunha BR, 2020) presentan una red de inteligencia criminal única para diferentes clases de delitos federales en todo Brasil y analizan su estructura topológica, robustez y respuesta a diferentes estrategias de ataque (da Cunha BR, 2020) exploran la naturaleza interconectada del comportamiento delictivo, construyen una red temática en la web oscura y comparan las estrategias de disruptión de redes con el trabajo policial real.

En el estudio de la propagación de enfermedades en redes espaciales tanto (Salje H, 2016) como (Lang JC, 2018) generan una variedad de estructuras espaciales aplicando suavizamiento a través de una función kernel cuyo argumento es la distancia espacial. La detección de puntos críticos espaciotemporales se ha llevado a cabo mediante enfoques descriptivos solo en el espacio, solo en el tiempo y también evaluando su interacción mediante diversas técnicas de minería de datos (Leong & Sung, 2015) y usando aprendizaje automático en (Zhao, 2017) y (Zhang, 2020) y usando aprendizaje profundo en (U. M. Butt, 2020).

Uno de los desafíos principales en el estudio de redes espaciotemporales, tales como redes de crimen, epidemiológicas, genéticas, físicas, entre otras, es la identificación de patrones relacionales como subredes con topologías específicas. Se han desarrollado trabajos

enfocados en la identificación de tales patrones, con el propósito de explicar dinámicas específicas asociadas con los agentes del sistema (por ejemplo, (Davies & Marchione, 2015); (Lotero, Hurtado, Floría, & Gómez-Gardeñes, 2016) y (Sanabria, Bohorquez, Rentería, & Mateu, 2022). Por lo tanto, el desarrollo y la implementación de algoritmos computacionalmente eficientes para la detección de patrones relacionales espaciotemporales es un reto fundamental en esta área (Sanabria, Bohorquez, Rentería, & Mateu, 2022).

Este trabajo se enfoca en caracterizar comportamientos de interés y detectar comportamientos inusuales que puedan ser relacionados con distancias en el espacio y en el tiempo a través de unas herramientas interpretables y computacionalmente eficientes. La propuesta se basa en la búsqueda, construcción y seguimiento de grafos específicos que aportan información relevante acerca de un fenómeno en estudio. Para este fin, se combina la teoría de redes con algunos métodos matemáticos y estadísticos de inferencia y simulación. En este trabajo se hace una correspondencia entre la ubicación espaciotemporal de los eventos y los nodos de una red. Las reglas lógicas que se establecen para la conexión entre los nodos deben tener en cuenta directamente tanto la ubicación como el momento de ocurrencia de los eventos. Se utilizan las ubicaciones de ocurrencia de los eventos como nodos y las conexiones se construyen dirigidas de acuerdo con el orden temporal de ocurrencia de dichos eventos. Así, en primer lugar, se tienen conjuntos de puntos con coordenadas espacio-tiempo a partir de los cuales se generan las redes espaciotemporales. Para este fin, es necesario establecer alcances de la correlación espacial y temporal presentes, así como distancias específicas tanto en el tiempo como en el espacio, a las cuales se observan cambios relevantes en el grado, conexiones y estructuras de la red. Finalmente, se propone un mecanismo inferencial con base en distribuciones exactas en algunos casos y simulaciones de redes aleatorias en otros.

En consecuencia, aquí se presenta una metodología que permite la caracterización de la estructura espaciotemporal de los eventos, mediante el uso de diferentes métodos de agrupamiento de los datos según sus atributos y su configuración espaciotemporal, con fines de detección de subredes importantes en las redes generadas a partir de la interacción entre los eventos, por ejemplo, (Sanabria, Bohorquez, Rentería, & Mateu, 2022). Esto posibilita el descubrimiento de diferencias sustanciales en el comportamiento del fenómeno de interés en diferentes zonas del área afectada y en diferentes momentos del

tiempo. Además, se propone una mejora al generador empírico de eventos de red aleatorios (ERGEN, *empirical random network event generator*) de (Sanabria, Bohorquez, Rentería, & Mateu, 2022), con el fin de incrementar su eficiencia computacional en la identificación de subredes sustantivos. Es así como se encuentran casos de subredes en los que se puede aproximar mediante expresiones cerradas la cantidad total de combinaciones de eventos, facilitando y agilizando la construcción de la distribución empírica asociada a la ocurrencia de estos subredes y así la detección de una cantidad significativa de los mismos.

Durante los últimos años se han desarrollado distintos estudios para abordar el análisis de crímenes, desde muchas perspectivas, entre ellas la ciencia de redes, que permite estudiar sistemas complejos por medio de grafos (Jiménez-García W., 2023). El crimen se ha estudiado principalmente por medio de la criminalística y la criminología, en donde la primera se encarga de estudiar la escena en donde ocurre el hecho, y la segunda de estudiar los aspectos sociales y psicológicos asociados al crimen. No obstante, un crimen es un suceso con una ubicación espaciotemporal, lo cual justifica la identificación de patrones que permitan caracterizar lugares y momentos de acuerdo con el riesgo. La prevención del crimen a través del diseño ambiental (CPTED), es muy importante dado el impacto en la salud pública, la vigilancia, la sostenibilidad y la calidad de vida. En el marco de la prevención de la delincuencia mediante el diseño ambiental, la manera como dos crímenes se relacionan espacialmente depende primordialmente del entorno próximo al lugar del delito o yuxtaposición geográfica (GJ, *geográfica yuxtaposición*). (Cozens, Love, & Davern, 2019), señalan que la GJ sirve para comprender una gama amplia de crímenes que se ven influenciados por algún actor, dependiendo de la cercanía o lejanía al sitio del delito. Para esto, se definen las siguientes formas de GJ: Micro GJ (factores de GJ que actúan en el sitio del delito), Proximal GJ (factores que actúan desde lugares próximos o contiguos al sitio del delito), Meso GJ (factores que actúan en zonas próximas y alejadas del sitio del delito), y finalmente, Macro GJ (factores que actúan como influencias remotas sobre el delito sin importar su ubicación). Así, la criminología ambiental busca estudiar los distintos aspectos y configuraciones del crimen (Boba, 2016), contrario a la criminología tradicional en donde se busca explicar la causa del crimen y por qué las personas se convierten en criminales. Patrones puntuales con ubicación espacial configurados por medio de un sistema coordenado universal, permiten que la criminología ambiental adquiera mayor

relevancia. Como mencionan (Felson & Boba, 2012), hay configuraciones que contienen mayores oportunidades para que un crimen ocurra y la criminología ambiental permite entender cómo estas oportunidades son creadas y luego se agrupan.

La metodología propuesta es desarrollada e ilustrada a través de su aplicación al caso de estudio de la criminalidad en la ciudad de Pereira, Colombia, caracterizando hurtos por medio de redes de eventos que permiten relacionar tanto la dimensión espacial como la temporal. La seguridad ciudadana es un tema de gran interés para el desarrollo y bienestar de las comunidades. En este contexto, es de sumo interés determinar si existen diferencias sustanciales en la criminalidad en diferentes zonas de la ciudad, ya que los crímenes tienden a agruparse de acuerdo con el contexto socioeconómico de la zona de ocurrencia (Boba, 2016). Por medio de la herramienta presentada en este trabajo, es posible hacer recomendaciones a las autoridades locales para implementar políticas públicas efectivas que mitiguen la criminalidad.

El resto de este trabajo está organizado como sigue: una Sección con la metodología en la que se explican los conceptos fundamentales para desarrollar el trabajo, así como los detalles asociados con la metodología propuesta; una Sección con la aplicación de la metodología propuesta a la criminalidad de la ciudad de Pereira, Colombia; finalmente, se discuten los hallazgos más relevantes junto con el trabajo venidero para investigaciones futuras.

## METODOLOGIA

En esta sección se presentan los conceptos fundamentales y se construye la metodología propuesta. Se utiliza el conjunto de coordenadas espaciales de los puntos en los que ocurren eventos de interés, por ejemplo, crímenes, para construir mapas de su distribución espacial y la de sus atributos asociados. Este conjunto de puntos es aleatorio ya que no es posible anticipar donde va a ocurrir el evento de interés y se estudia con base en los conceptos de procesos puntuales que están enmarcados en la teoría de geometría estocástica.

### Procesos puntuales

Un proceso espacial puntual en el plano corresponde a una colección de puntos

$$\{s_i \in R^2 : i = 1, \dots, N\}$$

donde  $s_i$  señala la ubicación espacial del  $i$ -ésimo evento en una región de observación determinada, para  $i = 1, \dots, N$ . Además, si se tiene el momento del tiempo en el que ocurrió el evento, se tiene un patrón puntual espaciotemporal de la forma

$$\{(s_i, t_j) \in R^2 \times R : i = 1, \dots, N, j = 1, \dots, T\}$$

conformado por la totalidad de ubicaciones y tiempos en los que han ocurrido eventos.

La georreferenciación es inherente a la ubicación espacial  $s_i$  de los eventos. En este trabajo se establecen las ubicaciones espaciales en principio por medio de longitudes y latitudes que posteriormente se proyectan a coordenadas planas.

### Patrón espacial de hurtos

Se ilustra el concepto de patrón puntual a través de la investigación detallada del hurto en la ciudad de Pereira, Colombia, entre el 25 de marzo de 2018 y el 30 de junio de 2021. A continuación, se identifican aquellos patrones que contribuyen a la ocurrencia del crimen por medio de una caracterización exhaustiva de los datos.

Pereira es una ciudad colombiana, capital del departamento de Risaralda. La ciudad se encuentra ubicada en el Valle de los ríos Cauca y Quíto, en el centro-oeste de Colombia. De acuerdo con el Censo Nacional de Población y Vivienda 2018, la población de Pereira es de 409,670 personas, lo que corresponde al 49% de la población del departamento. El porcentaje de hombres en la ciudad es 47.1% y el porcentaje de mujeres es 52.9% y, en consecuencia, se tiene una razón de masculinidad de 89.1% por cada 100 mujeres. Además, el porcentaje de la población entre 0 y 14 años es 22.6%, entre 15 y 64 años es 68.3% y mayor a 65 años es 9.1%, lo que significa que el 68% de la población son mayores de 18 años.

Pereira ha enfrentado problemas de violencia y delincuencia, en donde los delitos más comunes incluyen robos, hurtos, extorsiones y homicidios. La ciudad tiene un Plan Integral de Seguridad y Convivencia Ciudadana (PISCC), en el que se pueden encontrar las principales conclusiones y líneas de acción para combatir las distintas modalidades de delitos que ocurren en la ciudad. En este plan se menciona como Pereira ha sufrido problemas relacionados con el narcotráfico y las bandas criminales, junto con factores como la cultura de la ilegalidad, oportunismo, indiferencia, entre otros, lo que contribuye en gran medida a la perpetuación de la delincuencia. La base de datos contiene los delitos reportados a la

*Policía Nacional* que se utilizan para los análisis y reportes llevados a cabo por la misma entidad. Esta base contiene 6149 registros que comprenden entre el 25 de marzo de 2018 y el 30 de junio de 2021. Cada registro corresponde a un crimen con su respectiva ubicación espacial (latitud y longitud), y temporal (Hora y fecha en que ocurre el hurto) junto con una serie de variables que permiten caracterizar el tipo de crimen. Estas variables son:

- *Conducta*: Hurto de personas.
- *Modalidad*: Modalidad del hurto.
- *Municipio*: Pereira, Dosquebradas y La Virginia.
- *Zona*: Rural o urbana.
- *Comunas*: Comuna donde ocurre el hurto.
- *Barrio y dirección*: Barrio y dirección donde ocurre el hurto.
- *Sitio*: Descripción del sitio donde ocurre el hurto.
- *Armas*: Indicadora de si se usaron armas durante el hurto.
- *Género*: Género de la víctima.
- *Edad*: Rango de edad de la víctima.

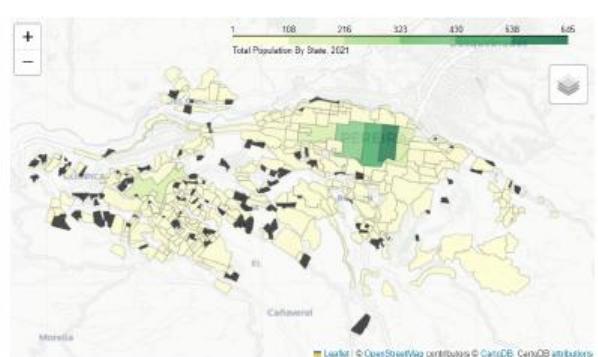
El hurto puede ser agravado o simple. Es agravado si hubo violencia y se utilizaron armas. En otro caso es simple en modalidad cosquilleo si ocurre el hurto y la víctima no se

da cuenta o es simple en modalidad raponazo cuando el ladrón ve la oportunidad y toma las pertenencias para huir de inmediato sin que haya tiempo de que la víctima reaccione. A continuación, se describen las características principales del hurto a personas en la ciudad de Pereira y sus municipios aledaños.

En la **Figura 1** se presentan el patrón puntual junto con un mapa temático por barrios correspondiente a los crímenes en la ciudad de Pereira y el municipio de Dosquebradas. Es evidente una cercanía geográfica entre ambas regiones y que muchos hurtos ocurren en la frontera correspondiente, lo que proporciona información valiosa sobre la relación entre la criminalidad en Pereira y Dosquebradas (Jiménez-García, 2014). Además, se hace evidente que el centro de Pereira corresponde a una zona de alta densidad de hurtos, mientras que en Dosquebradas no se distinguen zonas de alta densidad. Específicamente, los hurtos se concentran principalmente en tres barrios centrales de Pereira: Galería Central (12.51%), Plaza de Bolívar (10.01%) y Lago Uribe (6.89%).

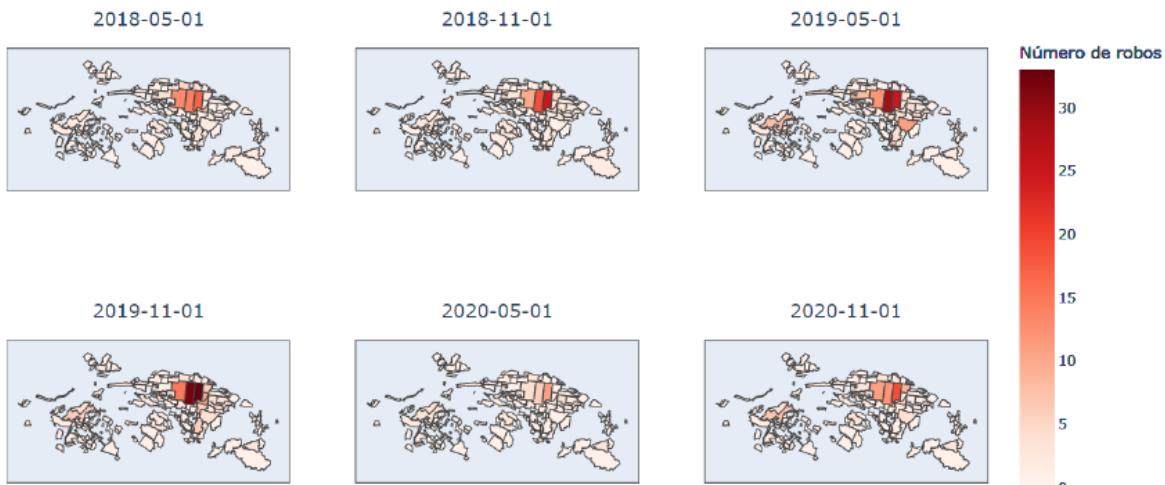


(a) Patrón puntual.



(b) Mapa temático por barrios.

**Figura 1.** Crímenes en la ciudad de Pereira y el municipio de Dosquebradas



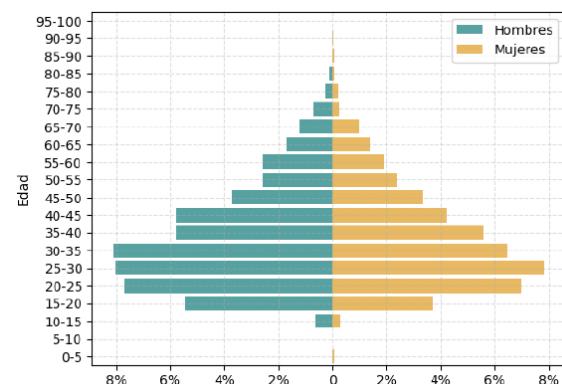
**Figura 2.** Mapas temáticos de los crímenes en la ciudad de Pereira por fecha

En la **Figura 2** se presentan los mapas temáticos de los delitos para distintos meses. Aunque en algunos meses parece decrecer la criminalidad, se observa que a través del tiempo la zona centro junto con algunas zonas periféricas llaman la atención por el bajo o alto número de robos, lo que es evidencia de barrios con posibles observaciones atípicas en términos de la delincuencia. Tales patrones sugieren la conformación de agrupaciones de eventos posiblemente debidas a circunstancias socio-económicas de algunos sectores que contribuyen a la perpetuación de la problemática. Finalmente, la variabilidad del crimen a través del tiempo y del espacio indica que existen relaciones espacio-temporales que se pueden estudiar por medio de redes de eventos.

En la **Figura 3** se presenta la pirámide poblacional por grupos quinquenales de edad. Los hombres alcanzan un mayor porcentaje de robos que las mujeres en general y por cada grupo de edad. Esto sugiere que los hombres son más susceptibles de ser víctimas de un hurto que las mujeres. Este resultado es llamativo porque la razón de masculinidad en la ciudad de Pereira es de 89,1 por cada 100 mujeres. Por lo tanto, ha de haber factores sociales, económicos o culturales que están influyendo en la distribución de los hurtos por género. Identificar tales factores sería una contribución fundamental para la mejoría e implementación del PISCC.

Adicionalmente, se estudia el número de hurtos por hora, día de la semana y parte del día cuando se cometen los hurtos. Para poder hacer el análisis de la temporalidad más adecuado, se crearon las siguientes franjas horarias para tener menor cantidad de categorías, las cuales son: entre las 05:00 y las 08:59 (mañana temprana), entre las 09:00 y las 12:59 (mañana), entre las 13:00 y las

15:59 (tarde temprana), entre las 16:00 y las 17:59 (tarde), entre las 18:00 y las 20:59 (noche temprana), entre las 21:00 y las 04:59 (noche).



**Figura 3.** Pirámide poblacional de los crímenes en la ciudad de Pereira

Primero, el número de casos por hora evidencia dos picos distintivos a las 10:00 y a las 19:00 horas. Así mismo, los viernes, sábados y domingos son los días de la semana cuando se cometen más hurtos durante la noche. Además, la tarde temprana y la tarde son los horarios más seguros de lunes a domingo, mientras que la mañana temprana es el horario más inseguro de lunes a jueves. También, siempre se observa que el centro de la ciudad es el área con la mayor cantidad de hurtos, sin importar el día de la semana y la parte del día cuando se cometen los hurtos. Como la distribución de los hurtos en la ciudad no está necesariamente relacionada con los horarios, entonces tal distribución podría estar influenciada por otros factores como la densidad poblacional y la presencia de actividades económicas, que cambian sustancialmente durante los días laborales y los fines de semana.

Como se observa en la **Tabla 1**, se tienen 27 modalidades diferentes de robo, y tan solo siete de estas modalidades acumulan la mayoría de casos (98.11%). Un estudio de la distribución de estas categorías durante el día cuando se cometen los hurtos señala que hay algunas diferencias importantes en algunas partes de la distribución. Por ejemplo, la modalidad de engaño ocurre casi siempre en la mañana, mientras que las modalidades de rompimiento de vidrios y violación de cerradura en la noche temprana. Si bien se podría intuir que las modalidades de hurto son independientes entre sí, es importante resaltar que algunas podrían estar relacionadas y, en consecuencia, llevar a cabo análisis espaciotemporales independientes para cada modalidad no sería la mejor opción.

**Tabla 1.** Número y porcentaje de casos para las distintas modalidades de crímenes reportados en la ciudad de Pereira y el municipio de Dosquebradas.

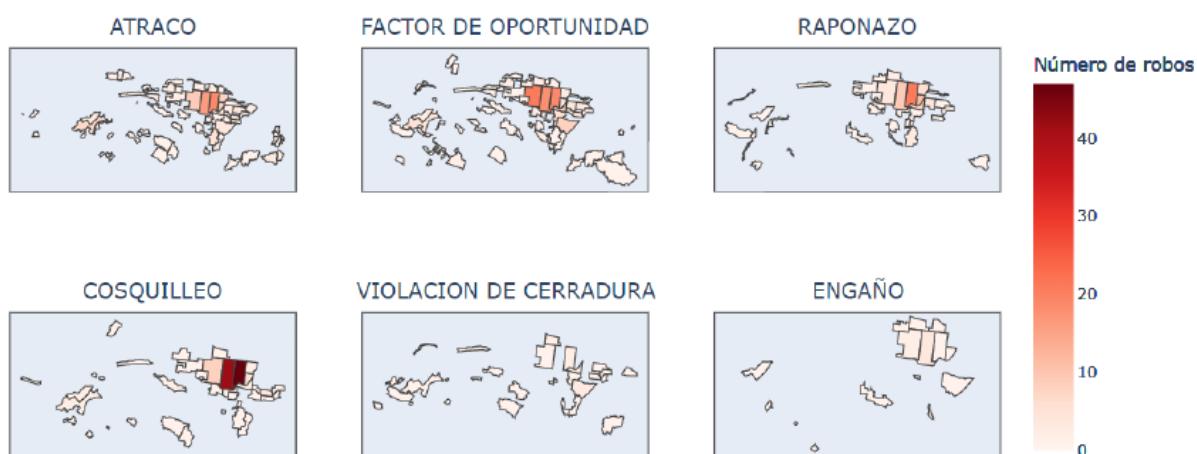
Modalidad	Número de casos	Porcentaje de casos
Atraco	3,109	38.61
Factor de Oportunidad	1,991	24.72
Raponazo	1,055	13.10
Cosquilleo	766	9.51
Violación de Cerradura	514	6.38
Engaño	284	3.52
Rompimiento de Vidrios	182	2.26
Otros (20 categorías)	152	1.89

Finalmente, en la **Figura 4** se presenta la distribución espacial de los crímenes durante el último trimestre del año 2019 por modalidad de hurto. Se selecciona este período en particular para evitar la superposición de patrones. No se despliegan aquellos barrios donde no se presentan casos. Algunas modalidades de hurtos parecen ser más

dispersas que otras. Además, se observa que los casos continúan concentrándose principalmente en el centro de la ciudad. En particular, se tiene un gran número de casos de cosquilleo en esta zona de la ciudad.

Con este análisis exploratorio es importante identificar algunos patrones que las autoridades pueden utilizar para la mitigación del crimen. En primer lugar, es evidente que la zona centro de la ciudad presenta el mayor número de casos en todo momento. Si bien esta característica se puede dar debido a factores socioeconómicos, es de vital importancia que las autoridades fortalezcan su presencia en estas zonas. En especial durante todas las mañanas, la noche temprana de lunes a jueves, y las noches de los fines de semana, ya que el crimen tiende a aumentar durante estos horarios.

Además, la distribución espacial del crimen no está necesariamente relacionada con la hora del día, pero la frecuencia de cada modalidad sí varía según el horario, por lo que se recomienda adaptar el control y la vigilancia en consecuencia por parte de las autoridades. Por ejemplo, se observa que engaño y rompimiento de vidrios son más frecuentes en la mañana y en la noche temprana, respectivamente. Adicionalmente, también se observa que cosquilleo es la modalidad con el mayor número de casos en la zona centro de la ciudad, por lo que se recomienda a las autoridades buscar maneras de mitigar esta problemática, que no es común en el resto de Pereira.



**Figura 4.** Mapas temáticos de los crímenes en la ciudad de Pereira por modalidades de hurto

## Redes

Una red es un sistema conformado por una colección de elementos conectados entre sí. Las redes se representan por medio de grafos. Los vértices o nodos representan los elementos de la red, mientras que las aristas o enlaces representan las conexiones entre ellos. Para más detalles ver (Newman, Barabási, & Watts, 2011); (Estrada, 2012);; (Barabási A.-L. , 2013); (Kolaczyk & Csárdi, 2014); (Latora, Nicosia, & Russo, 2017).

Una red  $G = (V, E)$  consiste de una colección de vértices  $V = \{v_1, \dots, v_N\}$  y una colección de parejas ordenadas  $E = \{(v_i, v_j) \in V \times V : i \neq j\}$ , donde cada pareja  $(v_i, v_j)$  representa una conexión dirigida del vértice  $v_i$  hacia el vértice  $v_j$ . Se dice que una red es no dirigida si los enlaces no gozan de direccionalidad, es decir,  $(v_i, v_j)$  es equivalente a  $(v_j, v_i)$  para todo  $i \neq j$ , de lo contrario, la red se denomina dirigida. Estas redes se conocen como redes binarias. Aquellas redes en las que los enlaces están asociados con cantidades numéricas para caracterizar las conexiones entre los vértices se denominan redes ponderadas. Al representar una red mediante un grafo, los pares  $(v_i, v_j)$  simbolizan las relaciones entre los nodos, y su definición dependerá de las características de dichas relaciones, como su reflexividad, peso, dirección o cualquier otra propiedad relevante según el contexto. La teoría de grafos proporciona los elementos fundamentales para la operabilidad del modelo de red, al facilitar la representación de entidades y sus vínculos mediante un grafo.

Una manera alternativa de representar una red es por medio de una matriz denominada matriz de adyacencia. Para una red con  $N$  vértices, la matriz de adyacencia  $A = [a_{i,j}]$  corresponde a la matriz binaria de  $N \times N$  tal que  $a_{i,j} = 1$  si  $(v_i, v_j) \in E$  y  $a_{i,j} = 0$  en caso contrario (si no se permiten enlaces reflexivos en la red, se tiene que  $a_{i,i} = 0$  para todo  $i$ ). Además, la matriz de adyacencia de una red no dirigida es necesariamente simétrica, mientras que la matriz de adyacencia de una red dirigida es posiblemente asimétrica. De manera análoga, la matriz de adyacencia de una red ponderada se define de tal manera que  $a_{i,j}$  es igual al peso correspondiente asociado con la relación del vértice  $v_i$  al vértice  $v_j$ , y es igual a cero en caso contrario.

Un vértice  $v_i$  es incidente en un enlace de  $E$  siempre y cuando  $v_i$  forme parte de la pareja  $(v_i, v_j)$  o  $(v_j, v_i)$  que constituye dicho enlace, dependiendo de si el grafo es dirigido o no dirigido. En un grafo dirigido, la incidencia distingue entre un vértice de origen (inicial) y un vértice de destino (final), mientras que en

un grafo no dirigido, ambos vértices tienen igual relación con el enlace. Así, el grado del vértice  $v_i$  se define como el número de enlaces de los que  $v_i$  es incidente. Esta cantidad permite identificar los vértices más conectados de la red. Además, una caminata del vértice  $v_i$  al vértice  $v_j$  corresponde a una secuencia sucesiva de vértices y enlaces que empieza en  $v_i$  y termina en  $v_j$ . De esta manera, la distancia entre dos vértices se define como la longitud de la caminata más corta que une tales vértices (esta distancia se define como infinito si no existen caminatas entre los vértices).

Las medidas de centralidad permiten caracterizar la importancia de los vértices de una red. La definición de estas medidas depende de la noción de "importancia". Por ejemplo, las medidas de centralidad de cercanía señalan que un vértice es importante si está cerca de muchos otros vértices, mientras que las medidas de centralidad de intermedio indican que un vértice es importante si se encuentra localizado entre muchos otros vértices Ver (Menczer, Fortunato, & Davis, 2020).

## Construcción de las redes de eventos espaciotemporales

Cuando se tienen datos espaciotemporales es preciso establecer un criterio bien definido para producir una red a partir de la información registrada tanto en el tiempo como el espacio. El enfoque considerado aquí es aquel en el que cada evento del patrón puntual es un nodo de la red. Así, la red de eventos corresponde al conjunto de nodos que representa los eventos de estudio en la ubicación espacial y en el momento del tiempo donde ocurrió el evento.

(Davies & Marchione, 2015) proponen un mecanismo general que permite construir lo que estos autores denominan una red de eventos, basándose en  $N$  eventos con ubicaciones espaciales y tiempos de ocurrencia  $s_1, \dots, s_N$  y  $t_1, \dots, t_N$ , respectivamente. Para especificar esta red, se calcula la distancia espacial  $d_{i,j} = d(s_i, s_j)$  y la separación temporal  $t_{i,j} = |t_i - t_j|$  de todo par de eventos  $i$  y  $j$ , y se especifica tanto un radio espacial  $d$  como un radio temporal  $t$  con el fin de establecer una noción de "cercanía" en cada dimensión. De esta manera, la red de eventos se define como aquella red dirigida en la que se tiene un enlace del evento  $v_i$  hacia el evento  $v_j$  siempre que  $d_{i,j} \leq d$  y  $t_{i,j} \leq t$ .

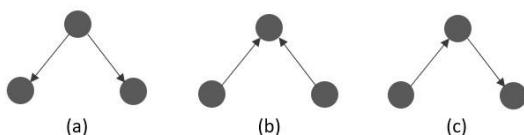
Para hacer posible la construcción de estas redes y para poder utilizar todo el histórico de los datos en un fenómeno de interés (Sanabria,

Bohorquez, Rentería, & Mateu, 2022) construyen series de tiempo de redes para una sucesión de radios espaciales dentro de ventanas específicas de tiempo, esto es, no para cada tiempo sino para franjas, por ejemplo, mañana, tarde y noche. En esta red dos eventos se relacionan entre si están en un rango de distancia  $d_{i,j} \leq d$  y dentro de una ventana temporal  $t$ . La red resultante corresponderá a una red dirigida según el orden de ocurrencia de los eventos. Esto es, dentro de cada ventana de tiempo un evento pasado  $v_i$  que ocurre en  $t$  se conecta con un evento  $v_j$  en un  $t'$  posterior mediante un enlace dirigido. Así, se obtienen series de tiempo de redes de eventos para cada  $d$  considerada. Por lo tanto, criterios que permitan una elección adecuada de los parámetros  $d$  y  $t$  es fundamental para el uso de esta metodología.

### Motifs de interés

La idea consiste en caracterizar estructuralmente una red por medio de la enumeración completa de *motifs* especiales. En particular, (Sanabria, Bohorquez, Rentería, & Mateu, 2022), realizan el estudio identificando tres tipos de *motifs* en la red de eventos: *motifs out-2-star*, *motifs in-2-star* y *motifs 2-path* (ver **Figura 5**).

De esta manera, resulta imperante enumerar completamente los *motifs* específicos de la **Figura 5** para caracterizar estructuralmente las redes de eventos. Aunque el conteo exacto es deseable, esta alternativa no es viable dado que para redes de tamaño considerable la complejidad computacional es inmanejable. .



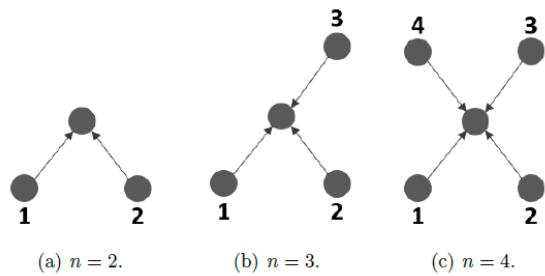
**Figura 5** (a) Motif out - 2 - star, (b) Motif in - 2 - star y (c) Motif 2 - path

En la literatura hay disponibles algunas alternativas para aliviar tal complejidad. Entre ellas, el generador empírico de eventos de red aleatorios (ERGEN, *empirical random network event generator*) propuesto por (Sanabria, Bohorquez, Rentería, & Mateu, 2022), que permite simular bajo aleatoriedad, redes con cada una de las configuraciones de la **Figura 5**. Con base en los eventos observados, es necesario saber si los conteos de cada tipo de *motif* son inusualmente altos o inusualmente bajos, y así determinar patrones en la configuración espaciotemporal de los eventos. Usando ERGEN se generan redes aleatorias y por tanto una distribución empírica de

cualquier medida de ellas. ERGEN realiza la permutación del orden temporal de los eventos. En particular es de interés, el conteo de cada uno de los tipos de *motif* aquí considerados.

En este trabajo, se utiliza una versión modificada de ERGEN. Se encuentra una expresión cerrada para el conteo de las combinaciones para dos de las tres configuraciones presentadas en la **Figura 5** por medio de la fórmula de Stirling (Dutka, 1991), considerando un radio espacial predefinido y todas las combinaciones posibles dentro de cada ventana de tiempo.

Utilizar la fórmula de Stirling es de vital importancia porque proporciona una aproximación más cercana al número exacto de *motifs out-2-star* y *in-2-star*, al mismo tiempo que se incrementa la eficiencia computacional del método regular basado en muestras que postula (Sanabria, Bohorquez, Rentería, & Mateu, 2022). La idea principal consiste en expresar el número total de *motifs in-2-star* y *out-2-star* por nodo por medio de combinatorias específicas. Por ejemplo, el número de *motifs in-2-star* es  $nC2$ , donde  $n$  es el número de vértices que están conectados con el vértice central y cumplen con las condiciones para formar *motifs in-2-stars* (la **Figura 6** presenta una ilustración del conteo de *motifs in-2-star* para  $n = 2,3,4$ ).



**Figura 6.** Ilustración del conteo de motifs in - 2 - star

Aplicando logaritmos seguidos de la fórmula de Stirling se tiene que

$$\begin{aligned} \ln \binom{n}{2} &= \ln \frac{n!}{(n-2)!2!} \\ &= \ln n! - \ln(n-2)! - \ln 2 \\ &\approx (n \ln(n) - n) - ((n-2) \ln(n-2) - (n-2)) - \ln 2 \\ &= n \ln(n) - (n-2) \ln(n-2) - (2 + \ln 2), \end{aligned}$$

de donde se obtiene una aproximación para el número de *motifs in-2-star* luego de exponentiar la mano derecha de la ecuación. El procedimiento para *motifs out-2-stars* es totalmente análogo. Sin embargo, para *motifs 2-path* no es posible llevar a cabo una

aproximación de este calibre debido a la forma secuencial de los subredes (en este caso se utiliza la versión regular de ERGEN basada en muestras junto con la restricción espacial definida previamente). En todos los casos, se realiza un escalamiento de la matriz de conteos en cada instante de tiempo para evitar el enmascaramiento en el conteo de los *motifs*.

### Distribución del grado de la red para la definición de la distancia

(Sanabria, Bohorquez, Rentería, & Mateu, 2022) utilizan una combinación del método de punto de cambio y del *scan statistics* espacial para la detección de un salto significativo en la cantidad de conexiones de las redes generadas con este algoritmo de un punto a otro del tiempo. Además, se hace un escalamiento de la matriz de conteos en cada momento del tiempo con el fin de evitar problemas de enmascaramiento del conteo de conexiones. Para cada distancia  $d$  y cada tiempo  $t$  considerados se construye el vector  $Y_{td}$  que contiene los conteos de cada uno de los 3 motivos buscados, esto es,

$$Y_{td} = (\#in - 2 - star, \#out - 2 - star, \#2 - path)$$

Y posteriormente se escalan calculando  $S_{td}^{-1/2}(Y_{td} - \bar{Y}_{td})$ , donde  $S_{td}$  es la matriz de covarianza del vector  $Y_{td}$  y  $\bar{Y}_{td}$  es el vector de medias correspondiente.

En este trabajo, nuestra propuesta metodológica considera una mejora a esta metodología utilizando como criterio para definir la distancia, uno que tenga en cuenta las propiedades de la topología de redes. Dado que es posible construir tantas redes de eventos como combinaciones de tiempos y radios espaciales sea posible para cada conjunto de datos, se pueden elegir aquellas que presenten la mejor capacidad predictiva de la distribución del grado de acuerdo con la familia Log-Normal de distribuciones, que es lo suficientemente flexible (y parsimoniosa!) para producir un sin fin de distribuciones probabilísticas que caractericen cualquier distribución de grado realista. De esta manera, en cada uno de los casos, se prueba el sistema de hipótesis

$H_0$ : "La distribución del grado de la red de eventos tiene distribución Log-Normal"

frente a

$H_1$ : "La distribución del grado de la red de eventos no tiene distribución Log-Normal"

por medio de la prueba de Kolmogorov-Smirnov (ver por ejemplo (D'agostino, 2017) y

(Sanabria, Bohorquez, Rentería, & Mateu, 2022)).

Este resultado es primordial, ya que se tiene evidencia empírica de que las redes se comportan de manera similar cuando corresponden a un mismo fenómeno. Para esto, se comparan los resultados con el trabajo realizado por (Broido & y Clauset, 2019). En ese trabajo, los autores ponen a prueba la idea de que las redes que se forman por eventos del mundo real siguen una topología libre de escala. Para esto, utilizan alrededor de 1000 redes sociales, biológicas, tecnológicas, de transporte e información. En el caso de las redes sociales, se encontró que en la mayoría de los casos la distribución del grado de las redes se puede ajustar a una distribución Log-Normal igual o incluso mejor que una ley de potencias.

Por todo lo anterior, para la construcción de una red de eventos, nuestra propuesta es seleccionar la distancia de acuerdo al grado de la red de eventos que mejor se ajuste a la distribución del grado de una red con topología conocida. En principio se podría considerar hacer el ajuste a una red libre de escala ya que han sido consideradas las redes que mejor gobiernan la dinámica de distintos eventos del mundo real. No obstante, se pone a prueba si lo establecido por (Broido & y Clauset, 2019), se cumple en este tipo de redes sociales, es decir, se ajustan más comúnmente a redes donde su distribución del grado sigue una distribución log-normal, y no a redes libres de escala. Por este motivo, se realizan pruebas de bondad de ajuste sobre redes de eventos construidas para diferentes valores de distancia y así se valida estadísticamente con cuál distancia el comportamiento de la red corresponde con lo reportado en la literatura y además permite hacer inferencia para garantizar un nivel de confianza en los resultados obtenidos.

### Algoritmos de agrupamiento

Se utilizan dos algoritmos para agrupar los datos y tener un punto de partida en la construcción de las redes: *k*-medias y DBSCAN (por sus siglas en inglés, *Density-Based Spatial Clustering of Applications with Noise*). Mientras *k*-medias divide los datos en *k* grupos basándose en la cercanía a los denominados centroides, requiriendo que se especifique el número de grupos de antemano, DBSCAN agrupa los datos según la densidad de los mismos, identificando grupos de formas arbitrarias sin necesidad de predefinir el número de grupos. Además, *k*-medias asume grupos de forma esférica y puede ser sensible a datos atípicos, mientras que DBSCAN es más

flexible en cuanto a la forma de los grupos y robusto frente a los datos atípicos, aunque puede ser más exigente computacionalmente, especialmente para conjuntos de datos grande. El algoritmo DBSCAN depende de dos cantidades fundamentales, la distancia máxima  $\epsilon$  entre dos puntos para que se consideren vecinos y el número mínimo de puntos a una distancia  $\epsilon$  para formar un grupo (minPts). Tanto  $\epsilon$  como minPts definen la naturaleza y la forma de los grupos. Ver (Wang, Yang, Liu, Samarasinghalage, & Zang, 2022), para una comparación exhaustiva de estos algoritmos.

### Cálculo de la distancia espacial

Las distancias espaciales entre dos puntos se calculan en este trabajo utilizando la distancia de Haversine. Esta distancia se define como  $\text{Hav}(\theta) = \sin^2(\theta/2)$ , con  $\theta = d/r$ , donde  $d$  es la distancia entre dos puntos de una esfera de radio  $r$ . Así, resolviendo para  $d$ , se tiene la inversa de Haversine dada por

$$d = 2r \arcsin(\sin^2((\phi_2 - \phi_1)/2) + \cos \phi_1 \cos \phi_2 \sin^2((\lambda_2 - \lambda_1)/2))^{1/2}$$

donde  $\phi_i$  y  $\lambda_i$  son la latitud y la longitud del punto  $i$ , respectivamente, para  $i = 1, 2$ .

La distancia de Haversine es particularmente útil para aplicaciones en las que se requieren cálculos precisos de distancias cortas y en las que las diferencias de altitud son mínimas. Sin embargo, esta distancia no es adecuada para situaciones en las que se requieren cálculos de distancias más largas ya que la tierra típicamente se asume como un esferoide oblat (ver (Banerjee, Carlin, & Gelfand, 2014) para más detalles).

Finalmente, hay una gran variedad de trabajos que tratan detalladamente el modelamiento de patrones puntuales espacio-temporales, como (Higdon D. , 2006); (Diggle, 2013); (Banerjee, Carlin, & Gelfand, 2014); (Wikle, Zammit-Mangion, & Cressie, 2019); (Oyana, 2020); (Haining, 2020) y (Kent, 2022).

### Paso a paso de la metodología propuesta

A continuación, se enumeran detalladamente los pasos de la metodología propuesta:

1. Análisis exploratorio de los datos con el fin de determinar atributos que permitan caracterizar los eventos.
2. Aplicación de los algoritmos de agrupamiento.
3. Definición de los subredes de interés según el contexto.

4. Selección del radio espacial entre eventos utilizando la prueba de bondad de ajuste, a partir de una sucesión de distancias generada con los métodos de agrupamiento.
5. Construcción de las de redes de eventos.
6. Conteo de los subredes de interés dentro de cada ventana de tiempo.
7. Estandarización de los conteos e inferencia para determinar motifs cuya frecuencia es significativa.

Todo el código de Python para implementar la versión modificada de ERGEN se puede consultar en

[https://github.com/miquelng/2027283-un\\_tf\\_maestria](https://github.com/miquelng/2027283-un_tf_maestria).

### ILUSTRACION: CRIMINALIDAD EN PEREIRA, COLOMBIA

Ahora se presenta la agrupación de los datos, la construcción de redes de eventos y el conteo de subredes relevantes, para los datos de criminalidad en Pereira.

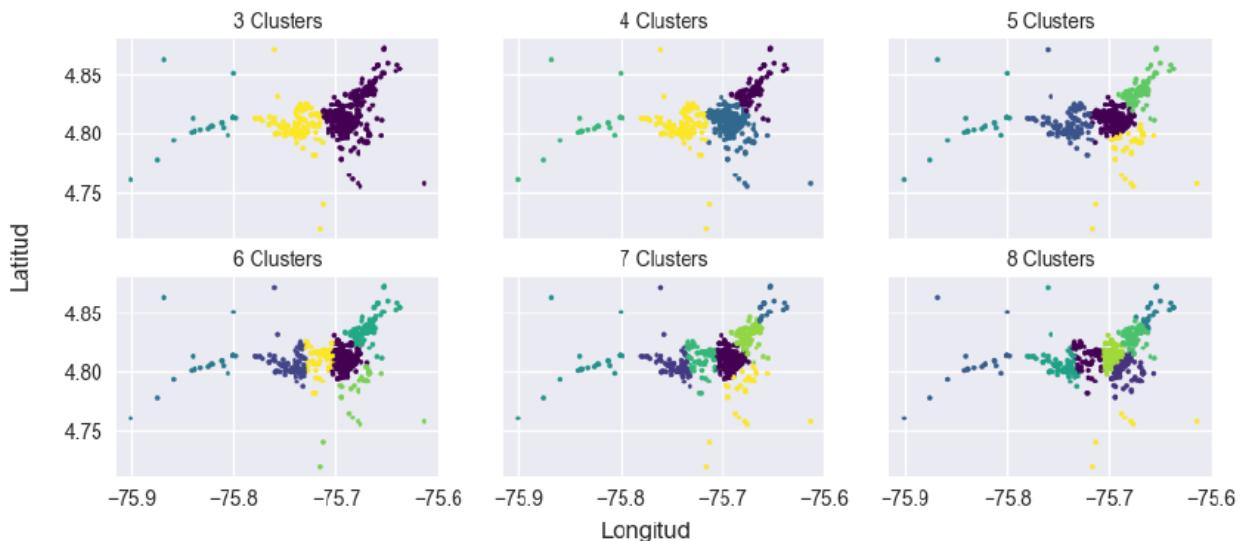
### Agrupación de datos

Dada la fenomenología propia del hurto (por ejemplo, los barrios presentan diferentes características sociodemográficas), se recomienda que dos eventos solo se puedan conectar en la red de eventos si ocurren dentro de un mismo “grupo” de acuerdo con sus características espaciales. (Felson & Boba, 2012) afirman que los patrones puntuales del crimen tienen a agruparse. Por tal motivo, a continuación, se realiza una agrupación de los datos teniendo en cuenta la georreferenciación de los hurtos, con el fin de identificar aquellas zonas de mayor incidencia delictiva. En particular, se comparan dos métodos de agrupamientos muy populares en la literatura que involucran distancias:  $k$ -medias (Macqueen, 1967); (Lloyd, 1982); (Arthur, 2007) (Celebi, Kingravi, & Vela, 2013)) y DBSCAN (Schubert, Sander, Ester, Kriegel, & XU, 2017).

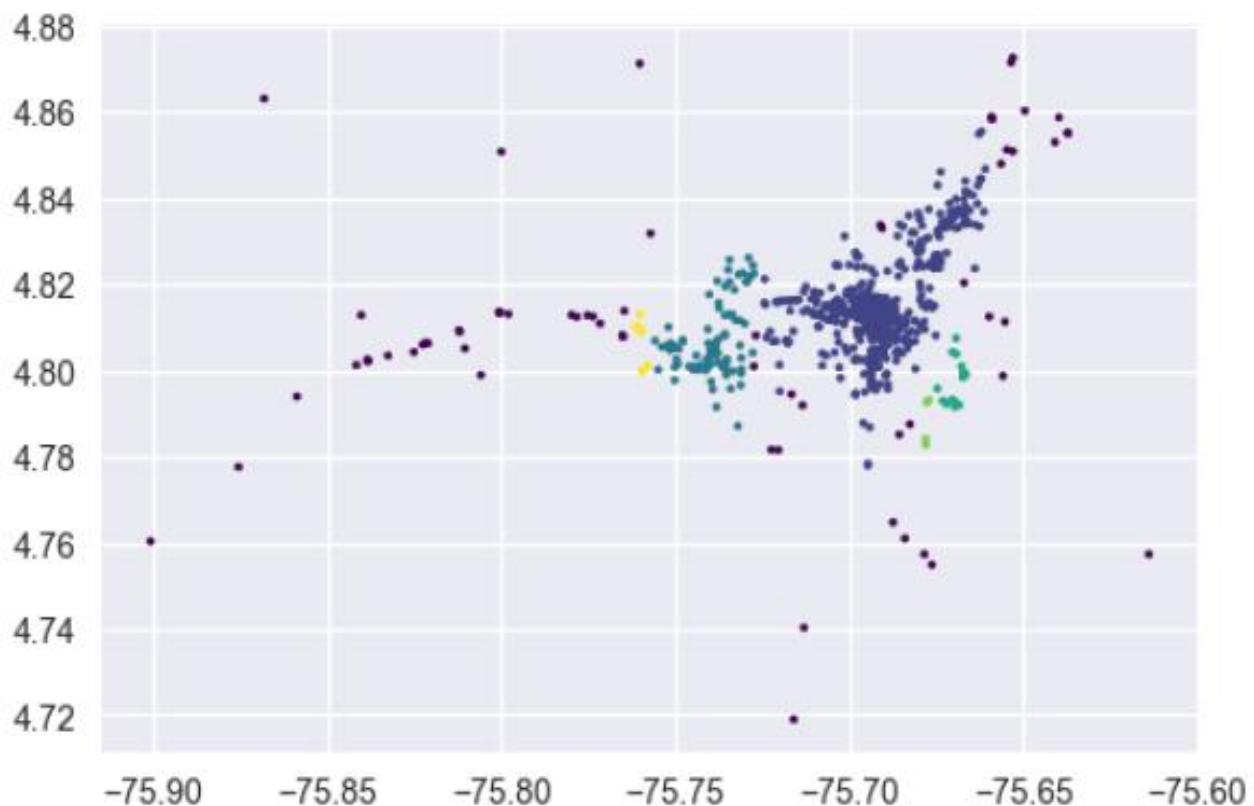
En la **Figura 7** se presenta la agrupación espacial de los hurtos cometidos en mayo, junio y julio de 2019 utilizando el algoritmo de  $k$ -medias con la distancia Euclidiana. De acuerdo con el “método del codo” (James, Witten, Hastie, Tibshirani, & Taylor, 2023), el número óptimo de grupos para segmentar los eventos es seis. Se observa que el centro de la

ciudad corresponde a uno de los grupos, y los barrios aledaños, incluyendo el municipio de Dosquebradas, corresponden a otros grupos. Aunque estos hallazgos tienen sentido en términos prácticos, se debe tener en cuenta que el algoritmo de *k-medias* tiende a dividir

los datos de manera esférica y puede ser sensible a datos atípicos, por lo que se recomienda contrastar los resultados con el algoritmo de DBSCAN.



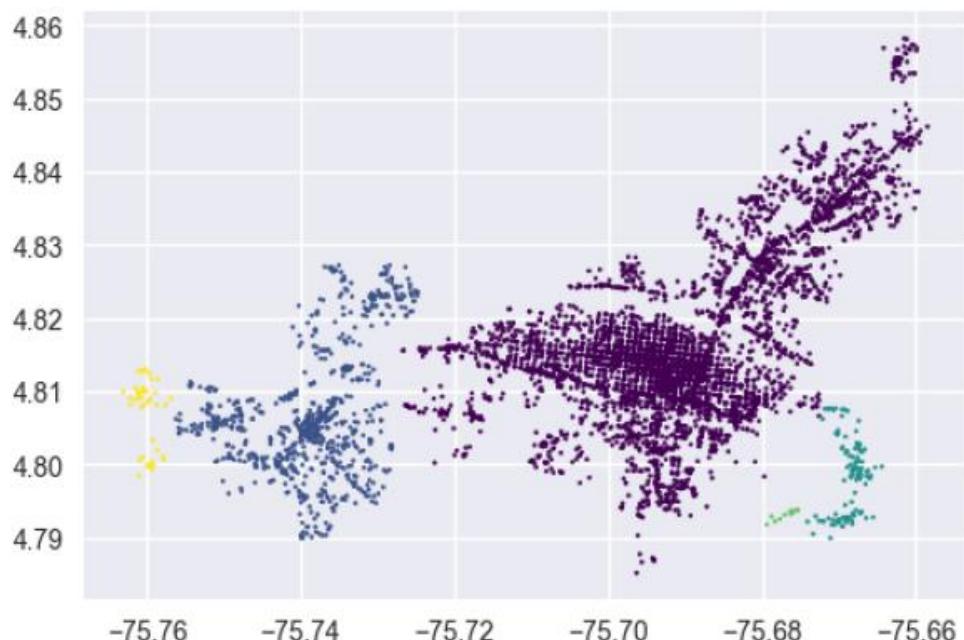
**Figura 7.** Agrupación espacial de los hurtos cometidos en mayo, junio y julio de 2019 utilizando el algoritmo de *k-medias* con la distancia Euclidiana.



**Figura 8.** Clusterización espacial de los hurtos en mayo, junio y julio de 2019 utilizando el algoritmo de DBSCAN con la distancia Haversine,  $\text{minPts} = 5$  y  $\epsilon = 0.000005$

En la **Figura 8** se presenta la clusterización espacial de los hurtos cometidos en mayo, junio y julio de 2019, pero esta vez utilizando el algoritmo de DBSCAN con la distancia Haversine y  $\text{minPts} = 5$  (este valor de  $\text{minPts}$  evita segmentaciones extremas con pocos o numerosos grupos). Nuevamente, de acuerdo con el “método del codo”, el valor óptimo de la distancia  $\epsilon$  para segmentar los datos es 0,00005. Se observa un gran grupo correspondiente al centro de la ciudad de Pereira y al municipio de Dosquebradas. Además, se identifican los datos atípicos de la segmentación (en color morado) que se localizan en su mayoría en zonas periféricas y carreteras alejadas la ciudad. Estos resultados están en concordancia con el comportamiento observado del crimen en la ciudad y siguen los principios de criminología ambiental correspondientes a la realidad del municipio.

Finalmente, en la **Figura 9** se presenta la clusterización espacial de los hurtos cometidos en toda la ventana de observación utilizando el algoritmo de DBSCAN siguiendo la metodología anterior y descartando los datos atípicos. Se observan cinco grupos bien definidos. Si bien la zona centro presenta el mayor número de casos, también es evidente una concentración importante de hurtos en el oeste de la ciudad. Como la zona centro y el municipio de Dosquebradas forman un único grupo, se concluye que ambas regiones están altamente relacionadas (posiblemente existe un efecto de contagio). Como medida de mitigación, se recomienda a las autoridades intervenir la frontera de Dosquebradas con el centro de la ciudad con el fin de lograr un impacto positivo en la reducción de la delincuencia en el resto del municipio.



**Figura 9.** Clusterización espacial de los hurtos en toda la ventana de observación utilizando el algoritmo de DBSCAN con la distancia Haversine,  $\text{minPts} = 5$  y  $\epsilon = 0.00005$ .

### Red de eventos

Se procede con la construcción de la red de eventos. Se especifican varios radios espaciales. Posterior a esta segmentación generada por los radios espaciales, y al interior de cada uno de ellos, se generan las conexiones dirigidas según el orden temporal de ocurrencia. Posteriormente se elige el mejor radio espacial, de acuerdo con la distribución del grado de la red.

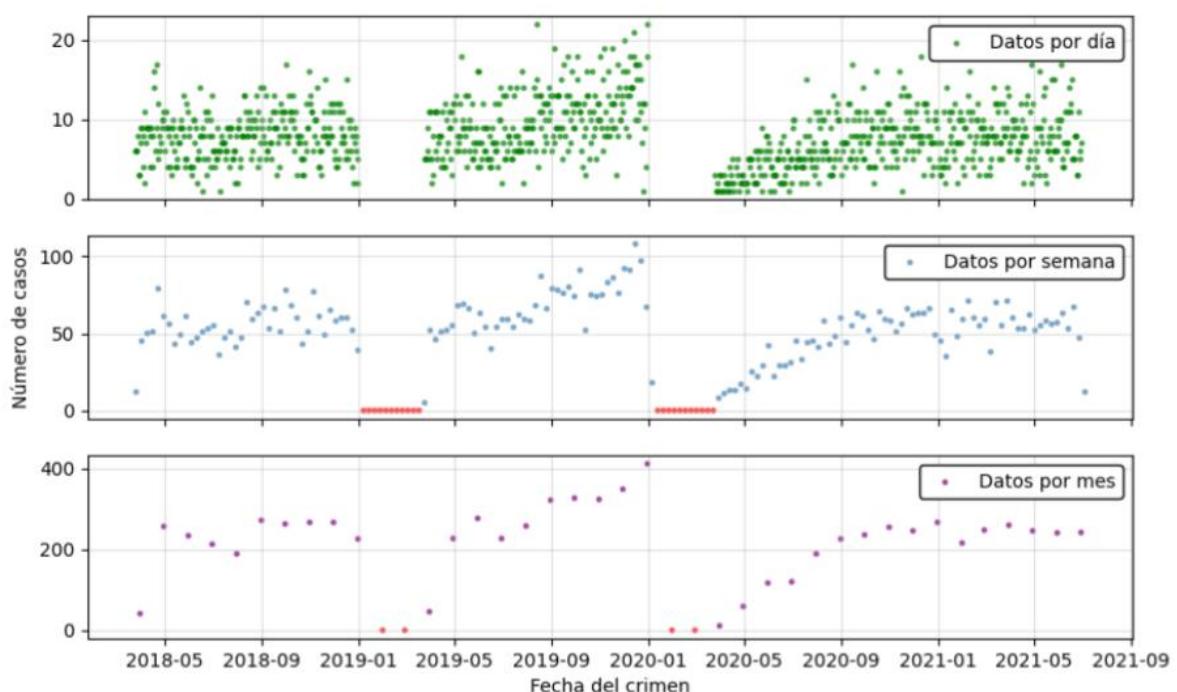
### Selección del radio espacial y del radio temporal

El radio espacial  $d$  y el radio temporal  $t$  determinan directamente el grado de “separación” de los hurtos con el fin de establecer el sistema relacional de los crímenes. Para la dimensión temporal, se hace  $t = 1$  para caracterizar cada uno de los 36 meses de observación, esto es, pueden estar conectados aquellos hurtos perpetrados dentro de una ventana de tiempo de un mes a partir del día específico de registro. Esta elección es

atractiva dado que abre un amplio rango de posibilidades a nivel temporal, lo que permite estudiar fácilmente la faceta espacial de las posibles conexiones entre los hurtos (Sanabria, Bohorquez, Rentería, & Mateu, 2022).

Con respecto a esta selección de la temporalidad de los datos, en la **Figura 10** se pueden observar diferentes agregaciones de los números de casos reportados para las distintas fechas. Se resaltan en rojo las semanas y meses para las fechas entre la primera semana de enero del 2019 y la segunda semana del mes de marzo del 2019, junto con el periodo de la segunda semana de enero de 2020 y la última semana de marzo de 2020. Para estos dos períodos no se tienen registros de crímenes, lo cual corresponde a procesos de principio de año en

los que se hacen reestructuraciones y también con el principio de la pandemia por Covid-19 en 2020. Para el desarrollo del caso de estudio no se tendrán en cuenta estos meses para no cometer errores en los análisis realizados. A pesar de ser de interés comparar diferentes ventanas de tiempo, no fue posible en este caso, la construcción de las redes con ventanas de tiempo más finas, debido a la calidad de los datos en general, pero además en particular por los cambios en los patrones durante la pandemia. Se espera que la calidad de los datos mejore en un futuro cercano, dada la actualización de los formularios y técnicas para los registros de las denuncias, para que sea posible hacer análisis incluso por horas o por franjas horarias.



**Figura 10.** Número de crímenes por: día, semana, mes. En rojo las semanas y meses en donde no se tienen datos.

Ahora, con el fin de establecer el radio espacial  $d$ , se considera la secuencia de distancias 250, 290, 300, 325, 350, 400, 450 y 500 metros (no se consideran distancias mayores que 500 ya que el área geográfica de Pereira es de tan solo 702 km<sup>2</sup> y un radio de 500 metros corresponde a un área de 0.785 km<sup>2</sup> por evento), y de estas alternativas, se escoge aquella cuya distribución del grado que se "ajuste mejor" a la distribución Log-Normal, por ejemplo, (Forbes, Evans, Hastings, & Peacock, 2011).

Así, se generan redes de eventos para las combinaciones de 36 y para los radios espaciales predefinidos en la secuencia de las 8 distancias, de las cuales se eligieron las de

mejor capacidad predictiva de la distribución del grado. En cada uno de los  $36 \times 8 = 288$  casos, se lleva a cabo la prueba de hipótesis de Kolmogorov-Smirnov para determinar si se puede suponer que la distribución del grado de la red sigue una distribución Log-Normal. Los resultados se resumen en la

**Tabla 2** donde se presentan los valores  $p$  promedio de las respectivas pruebas de Kolmogorov-Smirnov para cada una de las distancias de la secuencia. Utilizando un nivel de significancia del 5%, se observa que para todas las distancias existe suficiente evidencia para rechazar la hipótesis nula, exceptuando para las distancias 400 y 500 (la distancia de

500 metros es la que tiene asociada el mayor valor  $p$ ). De acuerdo a estos resultados se encuentra evidencia de que las redes de eventos criminales generadas para cada distancia, no siempre se comportan de manera similar.

**Tabla 2.** Valores  $p$  promedio de la prueba de Kolmogorov-Smirnov para cada uno de los radios espaciales predefinidos.

Distancia	Valor $p$
250	0.0003
290	0.0024
300	0.0018
325	0.0212
350	0.0221
400	0.0531
450	0.0486
500	0.0660

Finalmente, la

**Tabla 3** presenta el valor de la assortatividad para cada una de las redes de eventos resultantes para cada fecha. La assortatividad es una medida que indica cuándo los nodos tienden a unirse a otros que tengan un valor similar a su grado. Se puede observar que los valores obtenidos para todos los meses son similares. En general se tiene que el valor de la assortatividad de las redes está entre 0.15 y 0.35. Si bien esto significa que las redes en general pueden ser no assortativas, se tiene una tendencia a valores positivos. Este valor de la assortatividad se puede deber a que la mayoría de los crímenes ocurren dentro de la zona centro de la ciudad, en donde los nodos con un grado muy alto se conectarán con otros nodos con grado muy alto, y en los demás clusters de la ciudad se tendrán que nodos con grado bajo se conectan con nodos de grado similar. Desde el punto de vista de las autoridades, esta medida refuerza la idea de que es importante centrarse en las zonas con una alta densidad de crímenes.

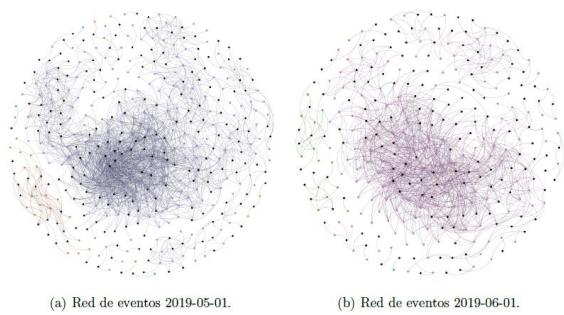
**Tabla 3.** Assortatividad de las redes de eventos generadas para una distancia de  $D=500$ .

Fecha	Asortatividad
2018-03-01	0.362
2018-04-01	0.337
...	...
2019-09-01	0.291
2019-10-01	0.292
...	...
2020-07-01	0.343
2020-08-01	0.281
...	...
2021-05-01	0.192
2021-06-01	0.241

### Conteo de motifs

Teniendo en cuenta los resultados de la Sección anterior, se construyen 36 redes de eventos utilizando un radio espacial  $d = 500$  metros y un radio temporal  $t = 1$  mes. En la **Figura 11** se ilustran dos redes para dos fechas específicas, a saber, 2019-05-01 y 2019-06-01. El color de los vértices indica el grupo al cual pertenecen los eventos. En ambos casos hay un grupo predominante, el cual corresponde a la zona centro de la ciudad. La red del 2020-06-01 es menos densa que la red del 2020-05-01, debido a la reducción sustancial de reportes por la pandemia de Covid-19.

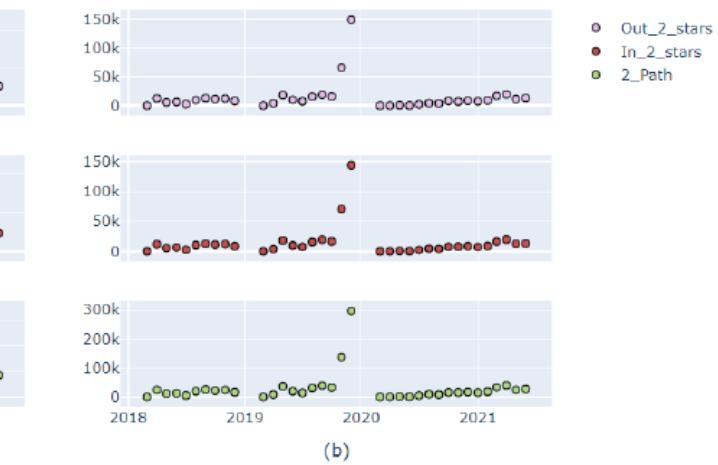
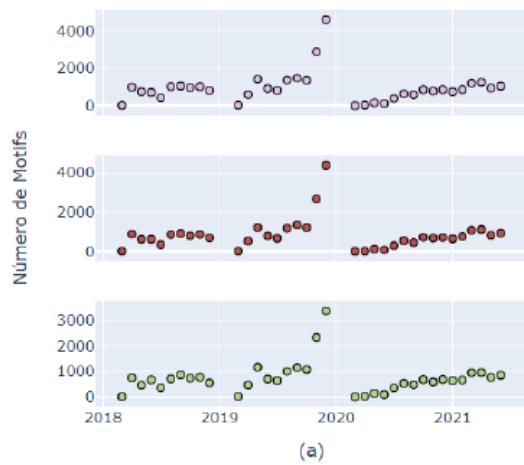
Una vez construidas las redes de eventos, se lleva a cabo el conteo de *motifs* *out-2-stars*, *in-2-stars* y *2-path* (ver **Figura 5**) en cada red utilizando tanto el algoritmo ERGEN regular basado en muestras, ERGEN-M, ver (Sanabria, Bohorquez, Rentería, & Mateu, 2022) como nuestro algoritmo ERGEN modificado basado en la fórmula de Stirling, ERGEN-S. La **Figura 12** presenta el conteo de los *motifs* usando cada uno de estos algoritmos. Se observa un comportamiento similar en todos los conteos, con un pico en diciembre de 2019. No obstante, el conteo usando ERGEN-S es mucho mayor dado que emplea una aproximación al conteo exacto de *motifs*. En ambas alternativas, las diferencias de los conteos a lo largo del tiempo se deben principalmente al número de eventos y conexiones de las redes. Ciertamente, no todos los meses tienen el mismo número de denuncias (en especial los primeros meses del año). Por tal motivo, se recomienda cautela cuando se comparan mensualmente los resultados, ya que el conteo de *motifs* depende en gran medida de la totalidad de registros del mes: un mayor número de denuncias conlleva a un mayor número de *motifs*.



**Figura 11.** Redes de eventos para dos fechas específicas usando un radio espacial  $d = 500$  metros y un radio temporal  $t = 1$  mes

En general, se tiene que motifs 2 – *path* ocurren con una mayor frecuencia. Por tal motivo, se

recomienda que, ante la ocurrencia de hurtos conectados de manera consecutiva, las autoridades inspeccionen 500 metros alrededor de la zona del segundo delito, ya que es bastante probable que el próximo esté anclado sucesivamente al anterior. En efecto, es mucho menos frecuente que dos crímenes desencadenen uno nuevo o que un crimen desencadene otros dos. Así, los hallazgos señalan firmemente que la “criminalidad se está moviendo” constante e individualmente de acuerdo con los datos observados. Por lo tanto, se recomienda que el gobierno local preste especial atención a bandas criminales que transitan constantemente por los distintos grupos obtenidos.



**Figura 12.** Conteo de motifs usando (a) ERGEN-M y (b) ERGEN-S

Desde el punto de vista computacional, el cálculo usando ERGEN-S es claramente menos costoso. De hecho, ERGEN-M tiene una duración de  $8.78 \pm 0.219$  segundos, mientras que ERGEN-S tiene una duración de  $1.56 \pm 0.016$  segundos. Estos algoritmos se implementaron en un computador comercial con recursos estándar (memoria RAM 16GB, disco duro 516GB SSD, procesador AMD Ryzen 7 4800HS, tarjeta gráfica GTX 1650 4GB, 8 núcleos). Adicionalmente, ERGEN-S permite aprovechar de manera más eficiente los recursos computacionales (ver el repositorio de GitHub).

## DISCUSION

Este trabajo provee avances sustanciales en varias direcciones. Primero, se expone una manera simple y efectiva de construir redes de eventos a partir de patrones puntuales espaciotemporales. Esto permite estudiar relationalmente los patrones puntuales de manera directa utilizando todas las herramientas asociadas con las redes. Segundo, se propone una metodología altamente eficiente para identificar *motifs* de interés. Esta propuesta nace como una modificación al algoritmo ERGEN de (Sanabria, Bohorquez, Rentería, & Mateu, 2022) y emplea la fórmula de Stirling para lograr su cometido. Finalmente, proporcionando toda la implementación computacional correspondiente en un repositorio de código

abierto, se hace un estudio detallado de la criminalidad del hurto en la ciudad de Pereira, Colombia, entre el 25 de marzo de 2018 y el 30 de junio de 2021. Allí, se lleva a cabo un análisis exhaustivo que incluye la descripción y clusterización de los datos, la construcción de redes de eventos y el conteo de subredes relevantes, para hacer recomendaciones realistas a las autoridades locales. Estos aportes hacen de esta investigación un punto de referencia importante para estudiar la dinámica de patrones relacionales espaciotemporales desde una perspectiva metodológica. Desde un punto de vista técnico, hay muchas alternativas de investigación que quedan abiertas para llevar a cabo en el futuro.

Desde modificaciones metodológicas (por ejemplo, métodos de clusterización, tipos de

distancias, selección de radios espaciotemporales, distribuciones del grado, entre muchas otras), hasta la implementación de modelos probabilísticos, por ejemplo, (Sosa & Buitrago, 2021) para investigar relationalmente los eventos (por ejemplo, clusterización y predicción) y cuantificar la incertidumbre correspondiente, tanto para redes de eventos simples como multicapa. También resulta muy llamativo llevar a cabo estudios exhaustivos de simulación para evaluar empíricamente la metodología propuesta. Algunas de estas propuestas serán publicadas posteriormente.

Aunque la metodología propuesta se ha ilustrado aquí con datos de crimen, se puede aplicar en general para la detección de patrones de eventos espaciotemporales. Se usa el potencial que tienen las redes conjuntamente con métodos estadísticos para identificar diferentes tipos de configuraciones. Así, el método puede utilizarse para estudios de salud pública, epidemiología, redes sociales o medioambiente, tales como la propagación de enfermedades, mensajes, noticias, comportamientos, entre muchos otros.

## REFERENCIAS

- Arthur, D. &.** (2007). k-means++: The advantages of careful seeding. *Soda*, 7, 1027-1035.
- Atluri G, K. A. (2018).** Spatio-temporal data mining: a survey of problems and methods. *ACM Comput Surv*, 51.(4) 1-41. <https://doi.org/10.3390/analytics2020027>
- Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2014).** Hierarchical modeling and analysis for spatial data. <https://doi.org/10.1201/b17115>. Chapman and Hall/CRC.
- Barabási, A.-L. &.** (1999). Emergence of scaling in random networks. *Science*, 286(5439), 509-512.
- Barabási, A.-L. (2013).** Network science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371. DOI: [10.1126/science.286.5439.509](https://doi.org/10.1126/science.286.5439.509).
- Betancourt, B., Rodriguez, A., & BOYD, N. (2020).** Modelling and prediction of financial trading networks: an application to the New York Mercantile Exchange natural gas futures market. *Applied Journal of the Royal Statistical Society Series C*, <https://doi.org/10.1111/rssc.12387>. doi:<https://doi.org/10.1111/rssc.12387>
- Boba, R. L. (2016).** Crime analysis with crime mapping. *Sage publications*.
- Boccaletti, S., Bianconi, G., Criado, R., Del Genio, C., Gómez-Gardeñes, J., Romance, M., . . . W. Z. (2014).** The structure and dynamics of multilayer networks. *Physics reports*, 544 (1), 1-122. doi: [10.1016/j.physrep.2014.07.001](https://doi.org/10.1016/j.physrep.2014.07.001)
- Borgatti SP, M. A. (2009).** Network analysis in the social sciences. *Science*, 323(5916), 892-895. DOI: [10.1126/science.116582](https://doi.org/10.1126/science.116582)
- Broido, A. D., & y Clauset, A. (2019).** Scale-free networks are rare. *Nature communications*, 10(1), 1017. <https://doi.org/10.1038/s41467-019-08746-5>
- Celebi, M. E., Kingravi, H. A., & Vela, P. A. (2013).** A comparative study of efficient initialization methods for the k-means clustering algorithm. *Expert systems with applications*, 40(1), 200-210. <https://doi.org/10.1016/j.eswa.2012.07.021>
- Cozens, P., Love, T., & Davern, B. (2019).** Geographical juxtaposition: A new direction in CPTED. En S. Sciences.
- da Cunha BR, M. P. (2020).** Assessing police topological efficiency in a major sting operation on the dark web. *Sci Rep*, 10, 1-10. doi:<https://doi.org/10.1038/s41598-019-56704-4>
- D'agostino, R. B. (2017).** *Goodness-of-fit-techniques*. Routledge. <https://doi.org/10.1201/9780203753064>
- Dang TA, C. J. (2018).** A comparative study of urban mobility patterns using large-scale spatio-temporal data. *IEEE international conference on data mining workshops (ICDMW)*. (págs. 572-579). DOI: [10.1109/ICDMW.2018.00089](https://doi.org/10.1109/ICDMW.2018.00089).
- Davies, T., & Marchionne, E. (2015).** Event networks and the identification of crime pattern motifs. *PLoS one*, 10(11). <https://doi.org/10.1371/journal.pone.0143638>
- Diggle, P. J. (2013).** *Statistical analysis of spatial and spatio-temporal point patterns*. CRC press. <https://doi.org/10.1201/b15326>
- Dorogovtsev, S. N., Goltsev, A. V., & Mendes, J. F. (2008).** Critical phenomena in complex networks. *Reviews of Modern Physics*, 80(4), 1275. <https://doi.org/10.1103/RevModPhys.80.1275>
- Durante, D. &.** (2014). Nonparametric Bayes dynamic modelling of relational data. *Biometrika*, 101(4), 883-898. <https://doi.org/10.1093/biomet/asu040>
- Dutka, J. (1991).** The early history of the factorial function. *Archive for history of exact*

- sciences, 225-249.**  
<https://doi.org/10.1007/BF00389433>
- Estrada, E. (2012).** The structure of complex networks: theory and applications. *American Chemical Society*, <https://doi.org/10.1093/acprof:oso/9780199591756.001.0001>
- Felson, M., & Boba, R. L. (2012).** Crime and everyday life. *Sage*. <https://doi.org/10.4135/9781483349299>
- Forbes, C., Evans, M., Hastings, N., & Peacock, B. (2011).** *Statistical distributions*. John Wiley & Sons. DOI:10.1002/9780470627242
- Grubesic TH, M. E. (2008).** Spatio-temporal interaction of urban crime. *Journal of Quantitative Criminology*, 24, 285-306.
- Gupta, S., Sharma, G., & Dukkipati, A. (2019).** A generative model for dynamic networks with applications. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33, 7842-7849. DOI: <https://doi.org/10.1609/aaai.v33i01.33017842>
- Haining, R. P. (2020).** *Regression Modelling With Spatial and Spatial-Temporal Data: A Bayesian Approach*. CRC Press. <https://doi.org/10.1201/9780429088933>
- Higdon, D. (2006).** A primer on space-time modeling from a Bayesian perspective. *Monographs on Statistics and Applied Probability*, 107, 217.
- Higdon, P. D. (2015).** Multilinear tensor regression for longitudinal relational data. *The annals of applied statistics*, 9(3), 1169. <https://doi.org/10.1214/15-AOAS839>
- James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023).** *An introduction to statistical learning: With applications in python*. Springer Nature. <https://doi.org/10.1007/978-3-031-38747-0>
- Jazayeri A, Y. C. (2020).** Motif discovery algorithms in static and temporal networks: a survey. <https://doi.org/10.48550/arXiv.2005.09721>
- Jiménez-García W., R.-R. R.-S. (2023).** Space-time analysis of theft from persons in Pereira (2019-2021). An approach to the theory of environmental munificence for crime. *Revista Criminalidad*, 65(1), 121-137. <https://doi.org/10.47741/17943108.405>
- Jiménez-García, W. G. (2014).** Hacia una tipología de lugares peligrosos. Caso de estudio de la comuna 11 de Dosquebradas, Colombia. *Revista Criminalidad*, 56(1), 133-156.
- Kent, J. T. (2022).** *Spatial analysis*. John Wiley & Sons. DOI:10.1002/9781118763551
- Kim, B., Lee, K. H., Xue, L., & Niu, X. e. (2018).** A review of dynamic network models with latent variables. *Statistics surveys*, 12.
- Kolaczyk, E. D., & Csárdi, G. (2014).** *Statistical analysis of network data with R*. New York: Springer. <https://doi.org/10.1007/978-3-030-44129-6>
- Lang JC, D. S. (2018).** Analytic models for sir disease spread on random spatial networks. *J Complex Netw.* <https://doi.org/10.1093/comnet/cny004>, 948-970
- Latora, V., Nicosia, V., & Russo, G. (2017).** *Complex networks: principles, methods and applications*. Cambridge University Press. <https://doi.org/10.1080/00107514.2018.1450296>
- Leong, K., & Sung, A. (2015).** A review of spatio-temporal pattern analysis approaches on crime analysis. *International E-Journal of Criminal Sciences*, 9, 9.
- Lloyd, S. (1982).** Least squares quantization in PCM. *IEEE transactions on information theory*, 28(2), 129-137. DOI: 10.1109/TIT.1982.1056489
- Lotero, L., Hurtado, R. G., Floría, L. M., & Gómez-Gardeñes, J. (2016).** Rich do not rise early: spatio-temporal patterns in the mobility networks of different socio-economic classes. *Royal Society open science*, 3(10). <https://doi.org/10.1098/rsos.150654>
- Macqueen, J. (1967).** Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 1, págs. 281-297.
- Menczer, F., Fortunato, S., & Davis, C. A. (2020).** *A first course in network science*. Cambridge University Press.
- Newman, M. (2018).** *Networks*. Oxford University Press. <https://doi.org/10.1093/oso/9780198805090.001.0001>
- Newman, M., Barabási, A.-L., & Watts, D. J. (2011).** *The structure and dynamics of networks*. Princeton university press.
- Oberoi KS, D. M. (2021).** Graph-based pattern detection in spatio-temporal phenomena. . *16th Spatial analysis and geomatics conference (SAGEO)*.
- Oyana, T. J. (2020).** *Spatial analysis with R: statistics, visualization, and computational methods*. CRC press. <https://doi.org/10.1201/9781003021643>

- Pasquaretta C, D. T.-M. (2021).** Analysis of temporal patterns in animal movement networks. *Methods Ecol Evol*, 12(1), 101-113. <https://doi.org/10.1111/2041-210X.13364>
- Salje H, C. D. (2016).** Estimating infectious disease transmission distances using the overall distribution of cases. *Epidemics*, 17, 10-18. <https://doi.org/10.1016/j.epidem.2016.10.001>
- Sanabria, A. M., Bohorquez, M. P., Rentería, R. R., & Mateu, J. (2022).** Identification of patterns for space-time event networks. *Applied Network Science*, 7(1). <https://doi.org/10.1007/s41109-021-00442-y>
- Schubert, E., Sander, J., Ester, M., Kriegel, H.-P., & XU, X. (2017).** DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems (TODS)*, 42(3), 1-21. <https://doi.org/10.1145/3068335>
- Scott J, C. P. (2011).** En *The SAGE handbook of social network analysis*. SAGE. <https://doi.org/10.48550/arXiv.2005.09721>.
- Sewell, D. K., & Chen, Y. (2017).** Latent space approaches to community detection in dynamic networks. *Bayesian Analysis*, 12(2), 351-377. <https://doi.org/10.1214/16-BA1000>
- Sosa, J., & Buitrago, L. (2021).** A Review of Latent Space Models for Social Networks. *Revista Colombiana de Estadística*, 44(1), 171-200. <https://doi.org/10.15446/rce.v44n1.89369>, 171-200
- Sznajd-Weron, K. y. (2020).** Opinion evolution in closed community. *International Journal of Modern Physics C*, 11(6), 1157-1165. <https://doi.org/10.1142/S0129183100000936>
- U. M. Butt, S. L. (2020).** Spatio-Temporal Crime HotSpot Detection and Prediction: A Systematic Literature Review. *IEEE Access*, 8. DOI: 10.1109/ACCESS.2020.3022808
- Wang, K., Yang, R., Liu, C., Samarasinha, T., & Zang, Y. (2022).** Extracting Electricity Patterns from High-dimensional Data: A comparison of K-Means and DBSCAN algorithms. *IOP Conference Series: Earth and Environmental Science*. IOP Publishing.,
- Wernicke, S. &. (2006).** FANMOD: a tool for fast network motif detection. *Bioinformatics*, 22(9), DOI: 10.1093/bioinformatics/btl038.
- Wikle, C. K., Zammit-Mangion, A., & Cressie, N. (2019).** *Spatio-temporal statistics with R*. Chapman and Hall/CRC.
- Zhang, Y. &. (2020).** Graph deep learning model for network-based predictive hotspot mapping of sparse spatio-temporal events. *Computers, Environment and Urban Systems*, 79. <https://doi.org/10.1016/j.compenvurbsys.2019.101403>
- Zhao, X. &. (2017).** Modeling temporal-spatial correlations for crime prediction. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, (págs. 497-506).

**Remitido:** 14-10-2024**Corregido:** 03-02-2025**Aceptado:** 07-04-2025