

Una Herramienta de Análisis y Visualización de Redes Sociales para la Identificación de Bandas Delictivas

A Tool for the Analysis and Visualization of Social Networks towards the Identification of Criminal Gangs

Sebastián P. WAHLER¹², Martín L. LARREA³, and Diego C. MARTÍNEZ³

¹ Departamento de Informática, Facultad de Ingeniería, Universidad Nacional de la Patagonia San Juan Bosco, Mitre 655, U9100, Trelew, ARGENTINA.

<http://www.ing.unp.edu.ar/dpto-informatica.html>

² Departamento de Informática, Procuración General, Ministerio Público Fiscal, Poder Judicial de la Provincia del Chubut, Belgrano 521, U9102, Rawson, ARGENTINA.

<https://www.mpfchubut.gov.ar>

³ Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur, Av. Alem 1253, B8000CPB Bahía Blanca, ARGENTINA.

<https://cs.uns.edu.ar/>

(e-mail: spwahler@ing.unp.edu.ar, dcm@cs.uns.edu.ar, mll@cs.uns.edu.ar)

Resumen Se presenta un estudio de las técnicas y metodologías actuales de análisis inteligente de datos y visualización para la asistencia en la investigación criminal, a partir de los registros de actividades delictivas, sus autores y las relaciones de datos que puedan derivarse a partir de ellas. Como corolario de la investigación, se muestra el desarrollo de un módulo de software para la visualización. Es de especial interés la identificación de redes ilegales, tales como bandas delictivas o criminales para propender a una persecución penal inteligente.

Keywords: Investigación Criminal · Análisis Inteligente de Datos · Redes Sociales · Visualización.

Abstract A study of the current techniques and methodologies of intelligent data analysis and visualization for assistance in criminal investigation is presented, based on the records of criminal activities, their authors and the data relationships that can be derived from them. As a corollary of the investigation, the development of a software module for visualization is shown. The identification of illegal networks, such as criminal gangs, is of special interest in order to promote intelligent criminal prosecution.

Keywords: Criminal Investigation · Data Analysis · Social Networks · Visualization.

1. Introducción

En la actualidad las actividades criminales habituales en una ciudad o región van desde hurtos y robos de poca importancia, hasta otros de mayor gravedad como amenazas, cibercrimen, abusos sexuales y homicidios. Todos ellos son registrados de diferentes formas por las fuerzas de la ley, con datos de variada precisión que incluyen usualmente la tipificación del delito, los datos en tiempo y espacio, y en muchas ocasiones los autores correspondientes.

Toda esta información respalda los procesos de investigación judicial de cada caso, pero con el transcurso del tiempo constituyen una extensa base de conocimiento sobre la cual es posible extraer valiosa información para la prevención del delito y la búsqueda de la justicia. Por ejemplo, es posible identificar relaciones entre personas de acuerdo a un análisis transitivo de eventos criminales en tiempo y espacio que sugieren la conformación de bandas delictivas. Las relaciones de amistad o conveniencia entre diversos autores de actividades criminales también puede inferirse de los registros delictivos y es de extrema relevancia para la prevención del delito y la resolución de casos inconclusos.

Las organizaciones criminales son grupos que operan fuera de la ley, realizando actividades ilegales en beneficio propio y en detrimento de otros individuos o grupos sociales [15]. Pueden ser de diverso tamaño y cubrir áreas geográficas variadas, en muchos casos en conflicto con otras organizaciones similares. Una de las características particulares de este tipo de organizaciones es que, al estar enfocadas en actividades ilegales perseguidas por los organismos de seguridad pública, el anonimato y/o la discreción de sus miembros es de vital importancia. Esto requiere estudios de la información existente con el fin de identificar los criminales y realizar acciones apropiadas para la prevención del delito.

Los miembros de las organizaciones criminales tienen a su vez diversos grados de compromiso con cada una de ellas. En muchos casos los hechos criminales que son evidentes en la sociedad ocurren por individuos de baja jerarquía y responsabilidad en el grupo, motivados por la recompensa inmediata, las aspiraciones de ascenso y la reputación en su propio círculo de contactos. Por otro lado, existen otros individuos de mayor jerarquía y responsabilidad en la organización criminal, que ostentan cualidades de liderazgo, intereses a largo plazo, y la constante preocupación por la conservación del poder para el beneficio personal y de la organización. Con frecuencia, son los individuos del primer grupo los que cometen delitos percibidos y registrados por las fuerzas policiales, mientras que los miembros del segundo grupo se mantienen con mayor discreción. Adicionalmente, las estructuras jerárquicas, la forma de operar, y la cultura inherente de sus realidades socio-económicas imponen códigos propios que hacen difícil la identificación de la organización delictiva como un todo, con sus miembros y actividades relacionadas. Es aquí donde nuestro trabajo puede aportar un rol significativo.

En tal sentido, con el objetivo de ayudar en la identificación de las bandas delictivas, sus integrantes y el grado de importancia de cada uno dentro de ellas, son de interés dos áreas de las Ciencias de la Computación: el área de Visualización de Información, en particular la Visualización de Grandes Conjuntos

de Datos, que busca asistir a los usuarios en la adecuada comprensión de la información, y el Análisis de Redes Sociales, en donde se emplean técnicas y formalismos para la comprensión de las estructuras de las redes y sus nodos. La aplicación de técnicas visuales para la representación de este tipo de información es importante [43] [14] [29], así como el estudio de las tareas e interacciones que la visualización debe soportar [8], ya que son estas interacciones las que facilitan la exploración de la visualización de información.

En esta línea de investigación estudiamos la aplicación de estas técnicas y tecnologías, contribuyendo además al desarrollo de componentes de software para la visualización y análisis inteligente de los datos, incorporando nociones de analítica de grafos. En particular en este trabajo nos interesa la consideración del algoritmo de *PageRank*, aportando a la detección de delincuentes de relevancia entre las *comunidades de individuos*, de la misma forma que se puede discriminar la importancia de un conjunto de páginas web a través de sus vínculos. Para esto se cuenta con los registros de actividades criminales a través de la colaboración del Ministerio Público Fiscal de la provincia del Chubut, que provee la base de conocimientos para inducir el grafo de relaciones como se detalla en la siguiente sección.

2. Análisis de Redes Sociales (SNA)

Desde hace algunos pocos años, el Análisis de Redes Sociales (o SNA por sus siglas en inglés de Social Network Analysis) ha contribuido a las investigaciones criminales y a las actividades de inteligencia relacionadas. Una red social modela individuos como nodos, vinculados entre sí por arcos o aristas que representan las relaciones entre esos individuos. El estudio de estas redes es importante porque se enfoca en la abstracción de las relaciones humanas sobre uno o más aspectos particulares [26] [6]. De esta manera, las redes conforman estructuras de grafos en las cuales es posible identificar diversas propiedades, tales como la relevancia o la importancia relativa de los nodos individuales en función de las conexiones existentes o el flujo de información. Según Sage [38], existen cuatro pilares fundamentales del análisis de redes: el reconocimiento de la importancia de las relaciones sociales entre los individuos, la recolección y análisis de datos sobre estas relaciones, la importancia de la representación visual de estos datos y la necesidad de modelos matemáticos y computacionales que expliquen los patrones de conexión entre los individuos.

En particular, la vinculación entre el estudio de las redes sociales y la investigación criminal ha sido encarada por varios autores. A mediados de los 70 se utilizaban modelos básicos para establecer y cualificar las relaciones entre individuos o actores de un escenario particular, definiendo grafos de acuerdo a la información recolectada [20], pero el procesamiento era mayoritariamente manual y con varias etapas de refinamiento y valoración de datos. Esta es la que según Klerk [22] sería la primera generación de análisis de redes en criminalística. La segunda generación involucra el uso de herramientas computacionales que automatiza parte de la tarea de registro y estructuración de datos. Estas herra-

mientas además aumentaron notoriamente la cantidad de datos que se pueden analizar, haciendo mucho más ágil su registro y consulta. La tercera y actual generación establece la definición de modelos y técnicas matemáticas para la generación de nuevo conocimiento, como la identificación de posiciones de poder e influencia o la calidad de potenciales testigos o informantes. Métricas como la centralidad de un nodo en un grafo son especialmente útiles en este escenario.

Uno de los trabajos más importantes al respecto es el de Krebs [23], en donde se identifica una parte de la red de terroristas que fue responsable de los atentados del 11 de septiembre de 2001 en Nueva York. Aquí identifica agrupaciones de individuos que se conectan entre sí por los pilotos responsables del secuestro de las aeronaves. Otros estudios similares han sido efectivos en consecuencia [31] [36] [40]. Por otro lado, el análisis de redes sociales ha cobrado también interés en la investigación criminal tradicional como las estructuras de la mafia o el narcotráfico [3] [4] [18] [32] [33]. Estudios como el de Malm [28] han permitido identificar roles en la cadena de suministros para la fabricación de drogas ilícitas, lo que acarrea diferentes riesgos penales para cada uno de los colaboradores. Otros estudios se enfocan en el uso del análisis de las redes sociales para otras actividades criminales, como el tráfico ilícito de arte [2], el lavado de dinero [9] [39], corrupción policial [24] y bandas juveniles [30] [1]. Existen también líneas de investigación en la disciplina referente al cibercrimen [10] [11] [12]. Es claro entonces que el análisis de redes sociales puede ser aplicado a un amplio rango de actividades criminales y ha demostrado modelar apropiadamente características propias de las organizaciones ilegales, asistiendo a la prevención del delito y al diseño de políticas adecuadas para enfrentar estas actividades.

Existen sin embargo algunas dificultades que requieren aún estudios intensivos. La cantidad de información que debe manejarse es enorme, en muchos casos con información incompleta, contradictoria y no menos frecuentemente incorrecta. Además, las relaciones humanas tradicionales se mezclan naturalmente con las interacciones ilícitas entre los individuos por lo que es necesario identificar apropiadamente su naturaleza y consecuencias y determinar los límites sensatos de la red social analizada.

Actualmente los organismos estatales encargados de la Justicia y la prevención del delito cuentan con registros informatizados de las actividades criminales detectadas, así como de las etapas y eventos del subsecuente proceso penal. Esta información constituye en esencia, una forma de red social. Para este trabajo es de especial interés la información producida a tal efecto por las fuerzas policiales de la Provincia del Chubut y su Poder Judicial de la mano del Ministerio Público Fiscal (MPF [16]), registradas en el sistema *Coirón*. Existen decenas de miles de registros que son utilizados principalmente para la acción penal, pero que pueden ser empleados para modelar diferentes redes sociales sobre las cuales aplicar un análisis matemático y computacional en la búsqueda de nueva información. Esto permitirá conocer más sobre las actividades criminales y sus autores en la jurisdicción de esa provincia, con las particularidades propias de la información registrada digitalmente.

El análisis y exploración de estos grandes conjuntos de datos y sus relaciones debe ser asistido por técnicas y herramientas que faciliten este proceso y reduzcan la carga cognitiva que recae sobre los usuarios. En tal sentido, el área de Visualización de Información, en particular la Visualización de Grandes Conjuntos de Datos, busca asistir a los usuarios de tal manera. La aplicación de técnicas visuales para la representación de este tipo de información no es nueva [43] [14] [29]. También es importante el estudio de las tareas e interacciones que la visualización debe soportar [8], ya que son estas interacciones las que facilitan la exploración de la visualización de información.

3. Marco de Trabajo - Ministerio Público Fiscal del Chubut

Coirón es el sistema informático que colabora con la administración del flujo de casos ingresados al Ministerio Público Fiscal del Chubut. Es una herramienta que permite registrar, comunicar y gestionar las actividades, trámites y actuaciones que se realizan para un caso penal, desde la denuncia hasta su finalización. Como herramienta de registro construye una base de datos con el historial de cada caso, así como de las personas involucradas y de los responsables de la gestión en cada oficina. Como herramienta de comunicación, agrupa la información, entrecruza las relaciones, identifica pertenencias y vinculaciones entre casos, personas, sus antecedentes y sus lazos. Como herramienta de gestión administra el flujo de casos y el trabajo de los integrantes de las oficinas responsables de los mismos. Permite planificar, organizar, coordinar y controlar el flujo de trabajo afín a cada caso y la sumatoria de ellos. Ha sido desarrollado a medida de las necesidades del Ministerio Público Fiscal del Chubut, tomando como base el Código Procesal Penal vigente y adaptado a los lineamientos estratégicos de diseño y gestión de Oficinas Fiscales definidos por la Procuración General. Su progreso, mantenimiento y mejora continua está a cargo del Equipo de Desarrollo del Departamento de Informática del Área de Planificación y Control de Gestión de la Procuración General. Entre otras funcionalidades, es de interés la incorporación de herramientas de visualización de información, potenciando el análisis que realizarán luego los especialistas de análisis criminal. En este trabajo nos enfocamos en las bandas delictivas y la *importancia relativa* de sus integrantes.

Una vez registrada toda la información relacionada a un hecho, y con ayuda de herramientas y vinculaciones con otros sistemas, se pueden obtener salidas que permiten llevar adelante la investigación de un caso o de un conjunto de hechos con características comunes.

Actualmente nos encontramos trabajando en la incorporación de herramientas de visualización de información que permitirán ver en modo gráfico lo que hoy se muestra en grillas, y listados, potenciando el análisis que realizarán luego los especialistas. Desde el análisis criminal, se puede realizar un perfilamiento relacional para la persecución penal, a través de la vinculación de "compañeros" de delitos y de redes sociales; a fin de identificar si forman parte de una banda o alguna organización criminal mayor [37]. En este sentido los analistas de redes

sociales utilizan dos tipos de herramientas matemáticas para representar información sobre los patrones de relaciones entre actores sociales: matrices y grafos, de gran ayuda visual cuando se trabaja con una gran cantidad de registros.

Existen muchas variaciones en los grafos, pero todos ellos comparten la característica común del uso de un círculo etiquetado para cada actor en la población que describimos y segmentos de línea entre pares de actores para representar el hecho que existe un vínculo entre ellos. Se denomina "*Grupo de Pertenencia*" en el Sistema Coirón a la relación directa que existe entre un individuo dentro del universo de personas cargadas como actores de delitos (roles: denunciado, sospechoso o imputado) y otros individuos del mismo universo, con los cuales existan uno o más casos penales en común.

Crear un módulo de software "Red de Grupos de Pertenencia" donde se muestre gráficamente las relaciones entre las personas involucradas en los casos penales es el objetivo principal de esta investigación. No sólo enfocarse en el grupo de pertenencia de una persona en particular, sino que mediante una visualización y con diversos filtros de búsqueda, se logre mostrar gráficamente las relaciones entre un determinado grupo de personas y de esta manera poder inferir la conformación de posibles bandas delictivas.

La idea central es reflejar de manera gráfica, mediante un Grafo, los grupos de pertenencia. Dentro del mismo se le llamará nodo a cada círculo, y representa a una persona (con los roles ya mencionados: imputado, sospechoso o denunciado) involucrada en dos o más casos penales. Existe un gran cúmulo de personas en el sistema con sólo un caso con rol de *denunciado*, por esa razón se los excluye del universo a analizar, no obstante podrían ser parte del dataset a visualizar si alguno/s de ellos se encuentran relacionados con otros nodos del primer grupo. El tamaño del nodo posee una relación directa con la cantidad de casos penales en los que se encuentre involucrada la persona. Cuanto mayor sea el tamaño del nodo en más cantidad de casos penales estará involucrado.

Los segmentos de líneas entre pares de nodos, vinculan a las personas entre sí y representan el o los casos que tienen en común. El grosor de la vinculación será directamente proporcional a la cantidad de casos en común entre un par de personas. Hay nodos que se encontrarán aislados en el grafo, esto no significa que no estén involucrados en casos, sino que quizás no existan relaciones para el filtro de búsqueda que se utilice en esa vista en particular.

Supongamos que una persona "A" se encuentra asociada a 8 casos penales, una persona "B" a 4 y una persona "C" a 2 casos. Agreguemos que las personas "A" y "B" se encuentran relacionadas entre sí, por estar en 3 casos en común (casos 1, 2 y 3). Por otro lado las personas "A" y "C" también se encuentran relacionadas, por tener un caso en común (caso 4). Una representación gráfica de dicha situación se muestra en la Figura 1, y puede observarse el doble de tamaño entre el nodo "A" y el nodo "B", representando justamente la diferencia de casos entre ambos nodos (8 y 4 casos). También se ve a simple vista el grosor del enlace entre "A" y "B" tres veces más grande que el enlace entre "A" y "C" (3 casos en común entre el primer par de nodos, y sólo un caso para el último par de nodos mencionado).

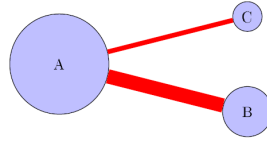


Figura 1. Ejemplo de relación entre tres personas.

Esta es, en primera instancia, una caracterización de la importancia de los individuos en la red. Sin embargo, analíticas mas complejas podrían aplicarse.

4. Descripción general de los Datos

En esta sección, describimos nuestro conjunto de datos de casos penales y la red de personas asociada, así como algunas características interesantes que se han de mencionar.

Dataset de Delitos Nuestro conjunto de datos consta de casos penales, actuaciones (bitácora de eventos del proceso penal), delitos, personas, elementos (denunciados y secuestrados), todos ellos relacionados; registrados entre octubre de 2006 y mayo de 2022 en la Circunscripción Judicial de Trelew - Chubut. Este conjunto de datos incluye lugares (relativos a personas y a hechos delictivos), fechas, estados procesales de los casos y las personas, como así también los vínculos entre todos los conjuntos mencionados. En el Cuadro 4, resumimos algunas de las características más importantes del conjunto de datos.

Característica	Cantidad total
Casos	105586
Personas	132950
Personas en Casos	183348
Delitos	113010
Nodos	33178
Enlaces	16964
Relaciones Nodos/Enlaces	60513

Cuadro 1. Totalizadores del conjunto de datos generales

Propiedades de la red A partir de los datos de los casos penales, pudimos construir la red de *Grupos de Pertenencia*. En esta red, se eliminan los nodos de aquellas personas cuyos roles no sean referidos a actores delictivos, como ser: denunciantes, víctimas, damnificados, etc. En la Figura 2 se muestra una visualización de la red. En la misma se puede observar un grafo compuesto de más de 30000 personas con más Casos Penales registrados en el Sistema de Gestión

Coirón (con los siguientes criterios: involucradas en más de un caso con rol de imputado, sospechoso o denunciado; se incluyen personas fallecidas, menores y personas jurídicas). Se visualizan además en la figura todas las relaciones que existen entre esas personas y sus grupos de pertenencia. A los sentidos prácticos

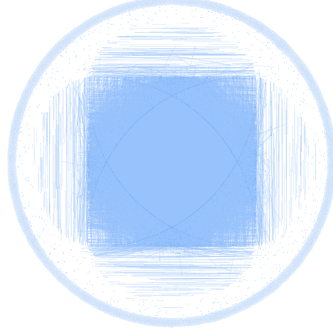


Figura 2. Más de 30000 personas con más casos y sus relaciones.

de la investigación penal, una visualización con tantos nodos y relaciones no es representativa ni conduce a ningún tipo de detección de bandas delictivas, pero es un claro ejemplo del universo de datos que se disponen en el dataset utilizado, como así también la potencia de la herramienta de visualización. En la Figura 3 se pueden observar ejemplos en los que se han tomado en cuenta los grupos de pertenencia de cada nodo a mostrar, es decir que se visualizan las personas con sus grupos de pertenencias particulares (según parámetros de búsqueda seleccionados). En la imagen (a) se muestran 10000 personas, en (b) 1000 y en (c) 100 personas con más casos y sus grupos de pertenencia relacionados. Al analizar

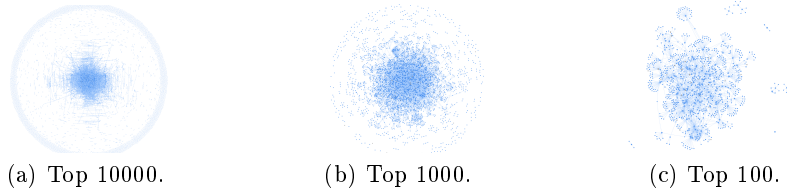


Figura 3. Personas en más casos, con la inclusión de sus grupos de pertenencia particulares.

la composición de la red obtenida podemos observar las relaciones que existen entre los nodos y como se "equilibra" el grafo, haciendo que aquellos nodos con pocas o nulas relaciones queden en la periferia de la gráfica. Sumado a ello también es apreciable la medida de centralidad de aquellos nodos que son rodeados por sus relacionados. Una aproximación más clara para denotar la medida de centralidad puede verse reflejada en la Figura 4, en donde se visualiza sólo las 10 personas con más Casos y sus grupos de pertenencia. Claramente esos 10 nodos

principales quedan rodeados de sus grupos de pertenencia y se pueden observar transitividades entre ellos a través de nodos que conforman parte del grupo de pertenencia de más de un nodo principal.

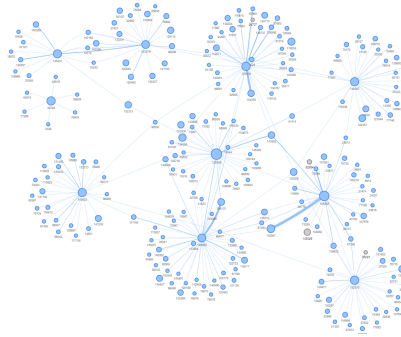


Figura 4. 10 personas con más casos en Coirón, con sus relaciones

Centralidad de Grado La centralidad de grado es una de las medidas más simples de centralidad. En esta se mide el número de enlaces o conexiones que tiene un nodo con los demás nodos pertenecientes a un grafo. Cuando se aplica un análisis de este tipo pueden determinarse diferentes medidas. Por ejemplo, en redes sociales podemos medir el grado de entrada de un nodo como la popularidad o preferencia que posea y la salida definirla como un indicador de sociabilidad. En nuestro caso de estudio, los miembros de las bandas delictivas modifican dinámicamente sus relaciones con otros miembros de la red, lo que resulta en un cambio de su rol e importancia. Una serie de medidas de centralidad de grado pueden ayudar a identificar estos cambios. Estas estadísticas se pueden utilizar para filtrar la vista de la red en función del valor de un nodo específico y resaltar su posición dentro de la red. El grado de centralidad en nuestro grafo se definirá entonces como el número de enlaces directos que tiene un delincuente. Un nodo con un alto grado puede verse como un "centro", un nodo activo e importante en la red [7].

Transitividad El coeficiente de agrupamiento (transitividad) de un gráfico mide el grado de conexión de una red. Altos coeficientes de agrupamiento significan la presencia de un alto número de triángulos en la red. Es bien conocido en la bibliografía [42] que las redes sociales muestran valores altos del coeficiente de agrupamiento cuando reflejan la estructura social subyacente de los contactos entre amigos/conocidos. Además, los valores altos del coeficiente de agrupamiento local se consideran un indicador confiable de los nodos cuyos vecinos están muy bien conectados y entre los cuales puede fluir una cantidad sustancial de información.

Desarrollo de la Visualización Para llevar a cabo la visualización del conjunto de datos obtenidos del análisis inteligente anteriormente descrito, se uti-

lizó `Vis.js` [41], una biblioteca o librería de visualización dinámica basada en lenguaje Javascript. La misma está diseñada para que sea fácil de usar, para manejar grandes cantidades de datos dinámicos y para permitir la manipulación y la interacción con los datos. La biblioteca consta de los componentes `DataSet`, `Timeline`, `Network`, `Graph2d` y `Graph3d`.

En nuestro caso particular utilizamos el componente "Network", que permite mostrar redes en grafos. La visualización es fácil de usar y admite formas, estilos, colores, tamaños, imágenes, etc. Funciona sin problemas en cualquier navegador moderno para hasta unos pocos miles de nodos y bordes. Para manejar una mayor cantidad de nodos, Network tiene soporte de agrupamiento. La red utiliza canvas HTML para la renderización.

Vis.js proporciona implementaciones de algoritmos de diseño forzados "Force-directed graph drawing". Estos algoritmos dirigidos por fuerza intentan posicionar los nodos considerando las fuerzas entre dos nodos (atractivos si están conectados, repulsivos de lo contrario). Generalmente son iterativos y mueven los nodos uno por uno hasta que ya no es posible mejorar o se alcanza el número máximo de iteraciones. Los enlaces tienen más o menos la misma longitud y el menor número posible de enlaces cruzados. Los nodos conectados se juntan más mientras que los nodos aislados se alejan hacia los lados.

5. Identificación de posibles Bandas Delictivas

En esta sección, describimos nuestro problema, algunos de los enfoques prácticos existentes utilizados por las fuerzas de la ley y nuestro enfoque basado en la teoría de grafos con características generadas principalmente por la distribución de los datos anteriormente descrita.

Métodos existentes Recordemos que en este trabajo nuestro principal interés es la identificación asistida de bandas delictivas y sus cualidades.

Las personas nos movemos habitualmente entre lugares conocidos o nodos (hogar, trabajo, supermercado, restaurante) y por las mismas calles o rutas. La teoría sugiere que cuando ocurre un delito es porque se cruzan delincuentes y víctimas dentro de algunas de estas zonas de actividad (nodo, ruta). A partir del análisis del lugar del delito se pueden determinar distintos tipos de víctimas y delincuentes que lo frecuentan, entender por qué concurren a ese lugar y qué hace que se encuentre la dupla delincuente-víctima. Es una manera estructurada de conocer e investigar patrones de comportamiento.

Por otro lado se puede deducir que los delincuentes se comportan igual que el resto de las personas, realizan actividades diariamente, se mueven por rutas conocidas para ir de la casa al trabajo, o a algún otro lugar que frecuenten. Es decir, mantienen una cierta rutina en sus vidas. Un delincuente tenderá a cometer un delito en algún lugar que se encuentre dentro o cerca del recorrido que realiza diariamente para trasladarse desde la casa al trabajo, del trabajo a algún lugar de recreación u otro lugar habitual.

De ambos enfoques se busca encontrar la mayor cantidad de patrones de ocurrencia entre diversos hechos de similar criminalidad y patrones horarios, como así también las zonas geográficas en donde se producen.

La naturaleza de los vínculos de los integrantes de una banda delictiva es una variable que aporta información sobre las características y similitudes de los miembros del grupo, atendiendo a criterios concretos: vínculo familiar, cultural, de proximidad (proviene del mismo barrio), han compartido prisión, de especialización (habilidades delictivas), la experiencia u otras capacidades, y otros tipos de vínculo.

Enfoque propio Ante los enfoques teóricos y prácticos estudiados anteriormente, nuestro desarrollo de software propio, que permite mostrar de manera gráfica las relaciones entre actores delictuales en el Sistema Penal de la Provincia del Chubut, se potencia como una herramienta vital de apoyo en la toma de decisiones de la investigación penal de bandas delictivas.

Poder visualizar relaciones entre las personas involucradas en casos penales ayuda a los especialistas a detectar triangulaciones, transitividades y por supuesto centralidades en la Red. Todo ello, sumado a los indicios de investigación y la propia expertís en la temática completan una herramienta de análisis para determinar ciertas bandas o grupos altamente relacionados.

En el año 2019 existieron investigaciones vinculadas a reiterados robos de televisores LCD en domicilios [21], como así también una serie de hechos consecutivos vinculados al robo de cajas fuertes en empresas del parque industrial de la ciudad de Trelew.

La UAC (Unidad de Análisis Criminal), organismo auxiliar de la Procuración General perteneciente al Ministerio Público Fiscal del Chubut, sirvió como equipo de apoyo en la investigación de ambos *modus operandi*, haciendo uso de toda la información de los legajos fiscales, consultas generales y específicas contenidas en el Sistema Coirón. Fue de vital uso la información referida a los *grupos de pertenencia* de cada persona, pero devino en un arduo trabajo entrecruzando información de personas, para dar con las supuestas bandas delictivas detrás de estos hechos.

Dichas investigaciones sirvieron como puntapié inicial para realizar este trabajo y poder facilitar la información ya contenida en el sistema de gestión penal, de otra manera, de una forma más directa y visual a la hora de investigar, que sirva directamente como apoyo a la toma de decisiones en las investigaciones de bandas delictivas.

A continuación se puede observar una visualización extraída de este trabajo, utilizando como filtros de búsqueda dos personas (nodos 116587 y 145262) con muchos casos y relaciones en el sistema, a fin de encontrar si existe algún tipo de relación directa entre ambos, y a su vez si existen nodos que produzcan transitividades o sean a su vez centrales de otros grupos. Desde la visualización se agregó gracias a la vinculación con el sistema de la oficina de identificación de personas, fotografías para colocar en los nodos y hacer de este trabajo una herramienta aún más potente. Aquellas personas que no hayan sido identificadas en

sede judicial no tendrán fotografía. Por cuestiones judiciales se han desenfocado las fotografías y se han colocado identificadores en vez de los nombres reales de las personas intervinientes.

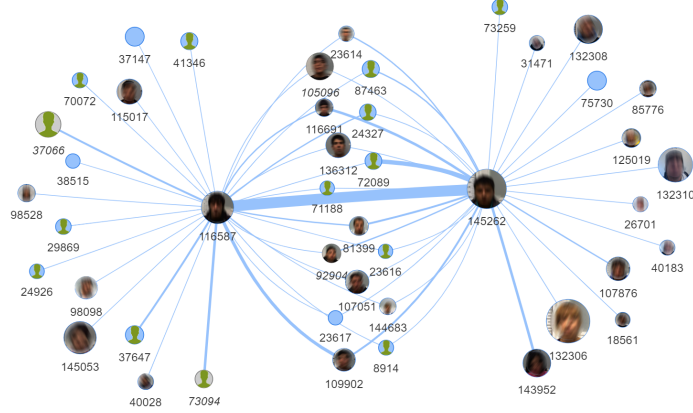


Figura 5. Relación entre dos personas. Se agregan fotografías.

Como puede verse, existen en la parte central de la imagen muchas personas que se encuentran relacionadas delictualmente con ambos nodos en cuestión. De esta manera se pueden tomar acciones con respecto a estas personas en pos de encontrar patrones de ocurrencia que los vinculen ante la posibilidad de identificarlos como una supuesta banda delictiva.

6. PageRank y detecciones comunitarias

Como parte simplificada de la estructura de una comunidad de nodos en una red social, cada uno representa a un individuo y la red tiene una segmentación multitudinaria [27]. Algunas personas son centrales en la comunidad, algunas están al margen, establecen menos relaciones con otros y, por lo tanto, tienen una influencia menor. En esta sección, presentamos un nuevo enfoque de descubrimiento comunitario basado en el algoritmo PageRank para encontrar a estos delincuentes “importantes” ó con “mayor influencia” en nuestro grafo, con el fin de analizar supuestas bandas delictivas. Recordemos que un grafo es un par $G = (N, A, g)$ donde N es un conjunto finito no vacío de elementos denominados *nodos* (vértices), A es un conjunto de arcos y g es una función que asocia a cada arco a perteneciente a A con un par no ordenado (x, y) , siendo x e y nodos pertenecientes a N . Se dice que a es un arco con vértices extremos x e y [13].

PageRank (PR) es un método que fue implementado a través de un algoritmo originalmente utilizado por Google que asigna a cada página web de un conjunto

dado, un puntaje que refleja su importancia dentro del conjunto. A este puntaje se lo denomina *valor de PageRank*. Ante una consulta, el buscador utiliza estos puntajes para determinar el nivel de relevancia de las páginas, y retorna en primer lugar aquellas con un puntaje más alto. Para calcular los puntajes, PageRank utiliza la estructura de enlaces de la web [5]. Una página web tiene un valor de PageRank alto si es apuntada por muchas otras páginas, o bien si es apuntada por páginas con puntajes altos [35]. PageRank tiene una base intuitiva en el concepto de *random walks* sobre grafos [19]: supongamos que un navegante aleatorio empieza a navegar la web desde una página cualquiera. El navegante puede hacer clic en forma aleatoria sobre alguno de los enlaces presentes en la página en la que se encuentra actualmente con una probabilidad d a la que se denomina *damping factor*, o bien con probabilidad $1 - d$ accede aleatoriamente a cualquier otra página web. Este proceso se repite indefinidamente. Luego, el valor de PageRank de una página P puede ser interpretado como la probabilidad de que el navegante aleatorio se encuentre en P al finalizar el proceso. PageRank es definido formalmente de la siguiente manera [17]. Sean q_i el número de enlaces salientes que posee la página i , n el número total de páginas web, d el *damping factor* que por lo general adquiere el valor 0.85, π un vector columna denominado *vector PageRank*, y $H = (h_{ij})$ una matriz cuadrada de tamaño n tal que $h_{ij} = 1/q_i$ si existe un enlace desde la página i a la página j , y $h_{ij} = 0$ en caso contrario. El valor h_{ij} corresponde a la probabilidad de acceder a la página j desde la página i en un paso, a partir de hacer clic en alguno de los enlaces que aparecen en esta última. El valor de PageRank correspondiente a la página j es π_j , y se define recursivamente como se muestra en la ecuación 1 [25].

$$\pi_j = \frac{1 - d}{n} + d \sum_{i=1}^n \pi_i h_{ij} \quad (1)$$

Aplicación de PageRank para bandas delictivas Nuestro dataset descrito anteriormente se obtiene a partir de consultas SQL a la Base de Datos de Coirón. Para hacer uso del algoritmo de PageRank se decidió incorporarlo dentro de esas consultas SQL de modo de obtener un resultado que pueda ser utilizado para la visualización. Dentro de la consulta original se genera una tabla para los nodos y otra tabla para las relaciones, de esta manera el software realizado para la visualización obtiene dichos datasets y renderiza el grafo. Para incorporar el cálculo de PageRank, inicialmente se adecuaron ambas tablas para la utilización de la fórmula, y se necesitó de ciertas tablas temporales para el cálculo. Por un lado se computó el grado de salida de cada nodo (*Out Degree*), es decir el número de enlaces que lo conectan con otros nodos. Luego se declara el *damping factor*, en nuestro caso 0.85, luego el conteo total de nodos, y se calcula el *PageRank inicial* de cada nodo, para después comenzar la iteración buscando cumplir con la sumatoria de la fórmula. El *damping factor* corresponde a un valor probabilístico que, aplicado al escenario de páginas web, pretende capturar la posibilidad de que un usuario continúe haciendo click en los links de una página en una sesión de navegación continua. Aquí este factor tienen un significado diferente, reintre-

pretado como el factor en el que se diluye la importancia de un individuo entre sus pares a través de una cadena de arcos. Actualmente estamos estudiando un valor apropiado en función de los datos existentes, puesto que el damping factor es esencialmente un valor empírico. En este trabajo optamos por usar el valor propio de la propuesta original del PageRank.

```

INSERT INTO #OutDegree
SELECT #Node.id, COUNT(#Edge.src)
FROM #Node
LEFT OUTER JOIN #Edge ON #Node.id = #Edge.src
GROUP BY #Node.id
DECLARE @dampingFactor float = 0.85
DECLARE @Node_Num int
SELECT @Node_Num = COUNT(*) FROM #Node
INSERT INTO #PageRank
SELECT #Node.id, rank = ((1 - @dampingFactor) / @Node_Num)
FROM #Node
INNER JOIN #OutDegree ON #Node.id = #OutDegree.id
DECLARE @Iteration int = 0

WHILE @Iteration < 50
BEGIN
--Iteration Style
SET @Iteration = @Iteration + 1
INSERT INTO #TmpRank
SELECT #Edge.dst, rank = ((1 - @dampingFactor) / @Node_Num)
+ (@dampingFactor * SUM(#PageRank.rank / #OutDegree.degree))
FROM #PageRank
INNER JOIN #Edge ON #PageRank.id = #Edge.src
INNER JOIN #OutDegree ON #PageRank.id = #OutDegree.id
GROUP BY #Edge.dst
END

```

Una vez finalizado el desarrollo de la fórmula, se procedió a realizar pruebas que corroboren el buen funcionamiento del código. Se realizaron ejemplos para pocos nodos con pocas relaciones, de manera tal que sea sencilla la verificación. Se muestra a continuación la Figura 6 que refleja la visualización de la ejecución de PageRank para 6 nodos. En cada nodo se muestra su posición de PageRank, y entre paréntesis el identificador de cada nodo. Como se puede observar el nodo central por el PageRank calculado es el referido al ID *145053*, cuyas relaciones con 3 nodos pesan sobre las relaciones que poseen el resto de los nodos visualizados en el grafo. Es interesante ver que éste individuo no es el que posee necesariamente la mayor cantidad de casos penales, pero es el más importante entre sus pares *de su propia red social* de contactos relacionados.



Figura 6. Resultado de ejecución de PageRank para 6 nodos.

Se ejecutó el algoritmo también para casos de estudio real resueltos en el MPF, donde las bandas y sus líderes han sido identificados, y de esta manera validar la relevancia de la implementación. Como ejemplo a continuación se muestra la Figura 7, sobre el caso real de robo de LCDs mencionado en el capítulo anterior. Pudo validarse que todos los integrantes de la banda se encuentran en el centro del subgrafo, altamente relacionados con los nodos de colores, que reflejan a los de más valor de PageRank.

Detección de comunidades Una comunidad puede ser definida como un conjunto de nodos que están más densamente conectados entre ellos que con el resto de la red. La importancia de este planteamiento radica en que se espera

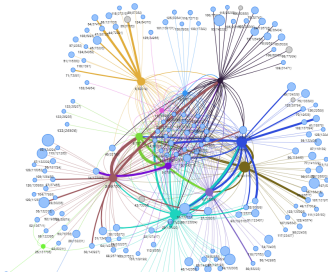


Figura 7. Resultado de ejecución de PageRank para el caso real de robo de LCDs.

que los nodos que están contenidos dentro de una misma comunidad compartan atributos, características comunes o relaciones funcionales [27]. En este trabajo se aprovecha el algoritmo SQL descrito con anterioridad en el cual se dividen nodos y relaciones para ser luego visualizadas por la herramienta *Vis.JS*. En este sentido, nuestro enfoque se basa en la búsqueda de posibles personas que participen en bandas delictivas. Del origen de datos surge que para cada nodo pueden conocerse todas sus relaciones, de modo que podemos asignar a cada uno de esos nodos referentes un identificador de grupo ó cluster. Es posible entonces verificar para cada par de nodos, si pertenecen a un grupo en particular (uno de ellos será "referente" y podremos identificarlo), y de esta manera asignar a cada relación también un grupo determinado. Con ambas tablas (nodos y relaciones) actualizadas, es posible desde la herramienta de visualización, asignar colores a cada cluster, y así generar un grafo aún más práctico a la vista.

Si bien para la visualización se utiliza lenguaje JavaScript de la mano de la librería anteriormente mencionada *Vis.JS*, y los dataset se obtienen desde la Base de Datos del Sistema Coirón a través de consultas SQL, el software de estudio que toma los datos del dataset y los procesa para luego llamar a la librería de visualización, se encuentra desarrollado en lenguaje C Sharp de .Net Framework. A continuación se exhibe en la Figura 8 la misma visualización que se ha presentado con anterioridad en la Figura 4, pero ahora con la detección de grupos por color. Se puede observar que cada grupo o cluster de nodos comparte el mismo color para los enlaces internos.

7. Conclusiones y trabajos futuros

Se ha presentado una propuesta de estudio de las técnicas y metodologías actuales de análisis inteligente de datos y visualización para la asistencia en la investigación criminal. Todo ello a partir de los registros de actividades delictivas, sus autores y las relaciones de datos que pueden derivarse a partir de ellas. Fue de especial interés la identificación de redes ilegales, tales como bandas delictivas o criminales para propender a una persecución penal inteligente.

Del estudio propuesto, como se explicó en los capítulos anteriores, se desarrolló un módulo de software como herramienta gráfica para visualizar la red de

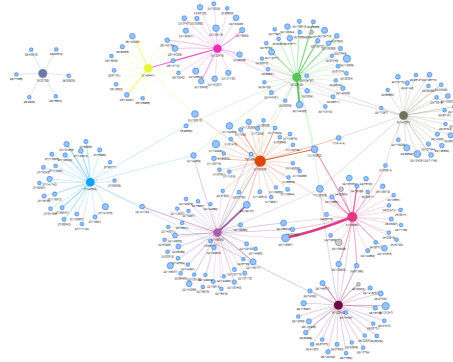


Figura 8. 10 personas con más casos en Coirón, con sus relaciones. Se agrega detección de comunidades por color.

grupos de pertenencia de los actores delictuales, incorporando un algoritmo de PageRank para reflejar aquellas personas importantes y Detección de comunidades asignándole colores a cada grupo/cluster de nodos. El desarrollo de software se implementó en el mismo Ministerio Público Fiscal, del cual se tomaron los datos para generar los datasets de pruebas. De esta manera se ha logrado no sólo estudiar las técnicas y metodologías expuestas, sino también alcanzar la puesta en producción y uso de la herramienta, por los propios actores de la investigación.

Las primeras impresiones de aquellos especialistas de investigaciones penales han sido muy satisfactorias y permiten evaluar a este trabajo como el inicio de futuros desarrollos visuales para el apoyo a la toma de decisiones en la investigación penal. Actualmente se continúa trabajando en el desarrollo de la aplicación de visualización. Se pretende modificar la fórmula original de PageRank, enriqueciéndola con información adicional según los registros judiciales. Por ejemplo, la posibilidad de darle mayor ranking inicial a aquellos nodos que tengan un peso mayor que otros según los registros. También sería interesante hacer lo mismo con el peso que poseen los enlaces entre nodos, ya que no es lo mismo una relación de 2 casos penales en común entre dos personas, que una de 18 casos en común. De esta manera lograríamos darle más ranking también a aquellos nodos que estén relacionados con otros, en mayor cantidad de casos penales. Esto es sin duda relevante para la investigación criminal basada en antecedentes penales. También se buscará profundizar sobre diversos algoritmos de centralidad. Existen algunas nociones que son de relevancia para la identificación de la importancia de una persona en la red inducida por las causas penales. Por ejemplo, *betweenness centrality*, que modela la medida en que un nodo en particular se encuentra entre otros nodos en una red, o *closeness centrality*, que es la inversa de la suma de los caminos más cortos (geodésicas) que conectan un nodo particular con todos los demás nodos de una red [34]. De manera similar, *eigenvector centrality*, es otra forma de asignar la centralidad a un actor de la

red basado en la idea de que si un nodo tiene muchos vecinos centrales, también debería ser central.

Referencias Bibliográficas

1. Bichler, G., Malm, A., Enriquez, J.: Magnetic facilities: identifying key juvenile convergence places with social network analysis. *Crime Delinq* **60**(7), 971–998 (2014)
2. Bichler, G., Malm, A.: Small arms, big guns: a dynamic model of illicit market opportunity. *Global Crime* **14**(2-3), 261–286 (2013)
3. Bouchard, M., Amirault, J.: Advances in research on illicit networks. *Global crime* **14**(2-3), 119–122 (2013)
4. Bright, D.A., Greenhill, C., Reynolds, M., Ritter, A., Morselli, C.: The use of actor-level attributes and centrality measures to identify key actors: A case study of an australian drug trafficking network. *Journal of contemporary criminal justice* **31**(3), 262–278 (2015)
5. Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems* **30**(1-7), 107–117 (1998)
6. Burcher, M.: *Social network analysis and law enforcement: Applications for intelligence analysis*. Springer (2020)
7. Carley, K.M.: Destabilization of covert networks. *Computational & Mathematical Organization Theory* **12**(1), 51–66 (2006)
8. Chen, H., Atabakhsh, H., Tseng, C., Marshall, B., Kaza, S., Eggers, S., Gowda, H., Shah, A., Petersen, T., Violette, C.: Visualization in law enforcement. In: *CHI'05 extended abstracts on Human factors in computing systems*. pp. 1268–1271 (2005)
9. Colladon, A.F., Remondi, E.: Using social network analysis to prevent money laundering. *Expert Systems with Applications* **67**, 49–58 (2017)
10. Décary-Héту, D.: Information exchange paths in irc hacking chat rooms. *Crime and Networks*. New York: Routledge pp. 218–230 (2014)
11. Décary-Héту, D., Dupont, B.: The social network of hackers. *Global Crime* **13**(3), 160–175 (2012)
12. Décary-Héту, D., Dupont, B.: Reputation in a dark network of online criminals. *Global Crime* **14**(2-3), 175–196 (2013)
13. Dubinsky, E.: Mathematical structures for computer science. by judith l. gersting. *The American Mathematical Monthly* **91**(6), 379–381 (1984)
14. Feng, M., Zheng, J., Ren, J., Hussain, A., Li, X., Xi, Y., Liu, Q.: Big data analytics and mining for effective visualization and trends forecasting of crime data. *IEEE Access* **7**, 106111–106123 (2019)
15. Finckenauer, J.O.: Problems of definition: what is organized crime? *Trends in organized crime* **8**(3), 63–83 (2005), <https://doi.org/10.1007/s12117-005-1038-4>
16. Fiscal, M.P.: Página web. <https://www.mpfchubut.gov.ar/>
17. Franceschet, M.: Pagerank: Standing on the shoulders of giants. *Communications of the ACM* **54**(6), 92–101 (2011)
18. Giommoni, L., Aziani, A., Berlusconi, G.: How do illicit drugs move across countries? a network analysis of the heroin supply to europe. *Journal of Drug Issues* **47**(2), 217–240 (2017)
19. Göbel, F., Jagers, A.: Random walks on graphs. *Stochastic processes and their applications* **2**(4), 311–336 (1974)

20. Harper, W.R., Harris, D.H.: The application of link analysis to police intelligence. *Human Factors* **17**(2), 157–164 (1975)
21. Jornada, D.: Caso de estudio real. https://www.diariojornada.com.ar/57375/policiales/Como_era_el_trabajo_de_la_banda_de_los_LCD_que_fue_desbaratada_esta_semana_en_Trelew
22. Klerks, P.: The network paradigm applied to criminal organisations: Theoretical nitpicking or a relevant doctrine for investigators? recent developments in the netherlands. *Connections* **24**(3), 53–65 (1999)
23. Krebs, V.E.: Mapping networks of terrorist cells. *Connections* **24**(3), 43–52 (2002)
24. Lauchs, M., Keast, R., Yousefpour, N.: Corrupt police networks: uncovering hidden relationship patterns, functions and roles. *Policing & society* **21**(1), 110–127 (2011)
25. Lin, J., Dyer, C.: Data-intensive text processing with mapreduce. *Synthesis Lectures on Human Language Technologies* **3**(1), 1–177 (2010)
26. M., K.R.P., Mohan, A., Srinivasa, K.G.: Practical social network analysis with python. *Computer Communications and Networks*, Springer International Publishing, Basel, Switzerland, 1 edn. (aug 2018)
27. Ma, X., et al.: Exploring sharing patterns for video recommendation on youtube-like social media. *Multimedia Systems* **20**(6), 675–691 (2014)
28. Malm, A., Bichler, G.: Networks of collaborating criminals: Assessing the structural vulnerability of drug markets. *Journal of research in crime and Delinquency* **48**(2), 271–297 (2011)
29. Mathew, A., Mary Jose, A., Sabu, C., Raj, A., et al.: Criminal networks mining and visualization for crime investigation. Caniya and P, Mufeed and Raj, Asha, *Criminal Networks Mining and Visualization for Crime Investigation* (July 8, 2021) (2021)
30. McGloin, J.M.: Policy and intervention considerations of a network analysis of street gangs. *Criminology & Public Policy* **4**(3), 607–635 (2005)
31. Medina, R.M.: Social network analysis: a case study of the islamist terrorist network. *Security Journal* **27**(1), 97–121 (2014)
32. Morselli, C.: Hells angels in springtime. *Trends in organized crime* **12**(2), 145–158 (2009)
33. Morselli, C.: Assessing vulnerable and strategic positions in a criminal network. *Journal of Contemporary Criminal Justice* **26**(4), 382–392 (2010)
34. Newman, M.E.: A measure of betweenness centrality based on random walks. *Social networks* **27**(1), 39–54 (2005)
35. Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: Bringing order to the web. Tech. rep., Stanford InfoLab (1999)
36. Qin, J., Xu, J.J., Hu, D., Sageman, M., Chen, H.: Analyzing terrorist networks: A case study of the global salafi jihad network. In: *International Conference on Intelligence and Security Informatics*. pp. 287–304. Springer (2005)
37. Rua, G., González, L.: Sistemas judiciales: Una perspectiva integral sobre la administración de justicia. *Análisis criminal en América Latina* **23** (2020)
38. Scott, J., Carrington, P.J.: *The SAGE handbook of social network analysis*. SAGE publications (2011)
39. Soudijn, M.R.: Using strangers for money: A discussion on money-launderers in organized crime. *Trends in organized crime* **17**(3), 199–217 (2014)
40. Stollenwerk, E., Dörfler, T., Schibberges, J.: Taking a new perspective: mapping the al qaeda network through the eyes of the un security council. *Terrorism and Political Violence* **28**(5), 950–970 (2016)
41. visjs: Página web. <https://visjs.org/>

42. Wasserman, S., Faust, K., et al.: Social network analysis: Methods and applications (1994)
43. Xu, J., Chen, H.: Criminal network analysis and visualization. *Communications of the ACM* **48**(6), 100–107 (2005)