

Aplicando PageRank en Registros de Actividades Criminales: una Aproximación a la Detección de Bandas Delictivas

Sebastián P. WAHLER^{1,2}, 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. Es de especial interés la identificación de redes ilegales, tales como bandas delictivas o criminales para propender a una persecución penal inteligente. Como corolario de la investigación, se muestra el desarrollo de un módulo de software para la visualización, incorporando la utilización del algoritmo de PageRank y detección de comunidades.

Keywords: Investigación Criminal & Análisis Inteligente de Datos & Redes Sociales & Visualización & PageRank

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 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 inferir relaciones de amistad o conveniencia entre diversos autores de actividades criminales a partir 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 [5]. 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 son cometidos por individuos de baja jerarquía y responsabilidad en el grupo. Asimismo, existen otros individuos de mayor jerarquía y responsabilidad en la organización criminal, que ostentan cualidades de liderazgo. 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. En particular, la aplicación de técnicas visuales para la representación de este tipo de información no es nueva [15] [4] [11]. También es importante el estudio de las tareas e interacciones que la visualización debe soportar [2], ya que son estas interacciones las que facilitan la exploración de la visualización de información.

En este trabajo es de especial interés 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 nos interesa la consideración de algoritmos como *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. Marco de trabajo - MPF Chubut

Desde hace 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. 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 por las fuerzas policiales de la Provincia del Chubut y su Poder Judicial de la mano del Ministerio Público Fiscal (MPF [6]), registradas en el sistema *Coirón*. Ésta 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. También es una herramienta de administración de información, flujo de casos, planificación, organización, coordinación y control. Ha sido desarrollado a medida de las necesidades del MPF, adaptado al Código Procesal Penal vigente y a los lineamientos estratégicos definidos. 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. Para ello es central el concepto de "*Grupo de Pertenencia*", como se denomina 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. En esta línea de trabajo parte del objetivo es la creación de componentes de software que muestren gráficamente las relaciones entre las personas involucradas en los casos penales, enriquecida con información proveniente de la analítica de grafos. La idea es no sólo enfocarse en el grupo de pertenencia de una persona en particular, sino que mediante herramientas inteligentes, una visualización apropiada y con diversos filtros de búsqueda, se logre mostrar las relaciones entre un determinado grupo de personas y de esta manera poder inferir la conformación de posibles bandas delictivas.

En el grafo se representa a una persona como un nodo (con los roles ya mencionados) si está 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. Sin embargo, 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, mayor es la cantidad de casos penales en las que está involucrado. Los arcos 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, y esto no significa que no estén efectivamente 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).

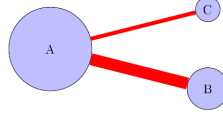


Figura 1. Ejemplo de relación entre tres personas.

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

3. Descripción general de los datos

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. Son 105586 casos penales, 183348 personas involucradas en casos, un universo de 132950 personas en total y 113010 delitos cargados. En relación al conjunto de datos para la visualización, constituyen: 33178 nodos, 16964 enlaces y 60513 relaciones Nodos/Enlaces.

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 denunciante, víctimas o damnificados. Una visualización de los datos puede verse en la Figura 2. Al analizar 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. En la Figura 2 se visualizan 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.

Para llevar a cabo la visualización del conjunto de datos obtenidos del análisis anteriormente descrito, se utilizó `Vis.js` [14], una biblioteca o librería de

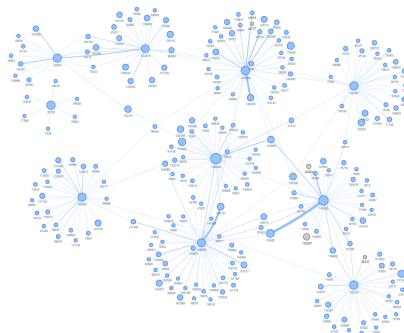


Figura 2. 10 personas con más casos en Coirón, con sus relaciones

visualización dinámica basada en lenguaje Javascript. Esta librería está diseñada para manejar grandes cantidades de datos dinámicos y permitir la manipulación y la interacción con los datos. El componente **Network** permite mostrar redes en grafos y manejar grandes cantidades de nodos. La librería además proporciona implementaciones de algoritmos de diseño forzados “force-directed graph drawing”, que 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.

4. Identificación de posibles bandas delictivas

Recordemos que en este trabajo nuestro principal interés es la identificación asistida de bandas delictivas. Las personas nos movemos habitualmente entre lugares conocidos o nodos (hogar, trabajo, supermercado, restaurante) y con frecuencia por las mismas calles o rutas. La teoría sugiere que muchos delitos ocurren cuando se cruzan delincuentes y víctimas dentro de algunas de estas zonas de actividad. Un delincuente tenderá a cometer un delito en algún lugar que se encuentre dentro o cerca del recorrido que realiza diariamente para trasladarse o su zona de movimiento habitual. 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), coincidencia en prisión, especialización (habilidades delictivas), la experiencia y otras capacidades y tipos de vínculo.

Existen antecedentes reales que justifican la importancia de esta línea de trabajo. En el año 2019 fue necesaria una investigación criminal sobre reiterados robos de televisores LCD en domicilios, 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 perteneciente al MPF, 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. Esta necesidad ha activado la línea de trabajo actual, utilizando la información ya contenida en el sistema de gestión penal, buscando proveer de una forma más directa y visual la base para el apoyo a la toma de decisiones en las investigaciones de bandas delictivas. Esto ayuda a los especialistas a detectar triangulaciones, transitividades y por supuesto *centralidades* e importancias internas en la Red. Para esto es necesaria la consideración de técnicas que permitan distinguir la *importancia* de un nodo en un grafo particular, como explicamos en la siguiente sección.

5. 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 [10]. 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 [3].

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 [1]. 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 [13]. PageRank tiene una base intuitiva en el concepto de *random walks* sobre grafos [8]: 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 [7]. 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 [9].

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

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 generaba 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 paginas 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, reinterpretado 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 muestran a continuación el Cuadro 5 con los resultados de PageRank para 6 nodos, y la Figura 3 que refleja la visualización de la misma ejecución. 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 causas penales, pero es el más importante entre sus pares *de su propia red social* de contactos con antecedentes relacionados.

IdNodo	PageRank
145053	0,286845042094944
145262	0,199512347112651
116587	0,1910604814502
132310	0,109790233522352
123109	0,106270247927848
129137	0,106270247927848

Cuadro 1. Valores de PageRank por nodo luego de la ejecución del algoritmo

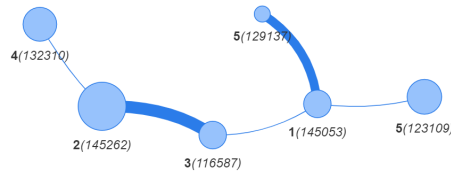


Figura 3. Resultado de ejecución de PageRank para 6 nodos.

La visualización del grafo que combina información penal con la relevancia basada en la topología del grupo de pertenencia provee mayor asistencia a una identificación primaria de individuos de importancia. Esto es compatible con el patrón de comportamiento que indica que, en varias organizaciones delictivas, los líderes no siempre evidencian la mayor visibilización judicial, contrario a lo que ocurre en los estratos inferiores que están mas expuestos al delito cotidiano. Por ejemplo, en las estructuras de mafia siciliana el *soldato* es quien realiza muchos de los actos delictivos simples de la Cosa Nostra, mientras que sus lideres los *capo-decina* y *capi-mandamento* se reservan mayor discreción.

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 que los nodos que están contenidos dentro de una misma comunidad compartan atributos, características comunes o relaciones funcionales [10]. 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. Mediante esa identificación es posible verificar para cada relación de 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 4 la misma visualización que se ha presentado con anterioridad en la Figura 2, 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.

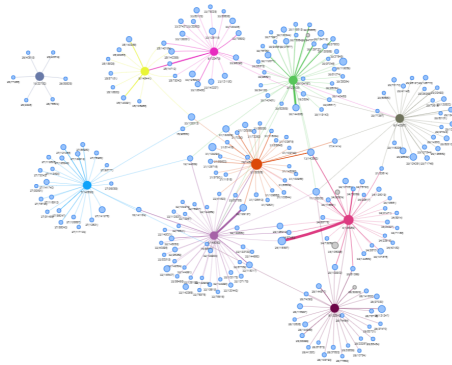


Figura 4. 10 personas con más casos en Coirón, con sus relaciones. Se agrega detección de comunidades por color.

6. 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 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. Como continuación de este trabajo, existen diversas líneas de investigación que quedan abiertas y en las que es posible continuar trabajando; algunas de ellas, están más directamente relacionadas con este trabajo y son el resultado de cuestiones que han ido surgiendo durante la realización del mismo.

Actualmente se continúa trabajando en el desarrollo de la aplicación de visualización. Se pretende incorporar a la fórmula original de PageRank, enriquecido 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 [12]. 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

1. 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)
2. 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)
3. Dubinsky, E.: Mathematical structures for computer science. by judith l. gersting. *The American Mathematical Monthly* **91**(6), 379–381 (1984)
4. 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)
5. Finckenaue, 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>
6. Fiscal, M.P.: Página web. <https://www.mpfchubut.gov.ar/>
7. Franceschet, M.: Pagerank: Standing on the shoulders of giants. *Communications of the ACM* **54**(6), 92–101 (2011)
8. Göbel, F., Jagers, A.: Random walks on graphs. *Stochastic processes and their applications* **2**(4), 311–336 (1974)
9. Lin, J., Dyer, C.: Data-intensive text processing with mapreduce. *Synthesis Lectures on Human Language Technologies* **3**(1), 1–177 (2010)
10. Ma, X., Wang, H., Li, H., Liu, J., Jiang, H.: Exploring sharing patterns for video recommendation on youtube-like social media. *Multimedia Systems* **20**(6), 675–691 (2014)
11. 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)
12. Newman, M.E.: A measure of betweenness centrality based on random walks. *Social networks* **27**(1), 39–54 (2005)
13. Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: Bringing order to the web. Tech. rep., Stanford InfoLab (1999)
14. visjs: Página web. <https://visjs.org/>
15. Xu, J., Chen, H.: Criminal network analysis and visualization. *Communications of the ACM* **48**(6), 100–107 (2005)