UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIA FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

IDENTIFICACIÓN DE INDIVIDUOS E INTERACCIONES CLAVES CONSIDERANDO ATRIBUTOS INDIVIDUALES PARA EL ANÁLISIS DE AGRUPACIONES DELICTUALES BAJO EL CONTEXTO DE UNA RED SOCIAL

TESIS PARA OPTAR AL GRADO DE DOCTOR EN SISTEMAS DE INGENIERÍA

FREDY HUMBERTO TRONCOSO ESPINOSA

PROFESOR GUIA:

RICHARD WEBER HAAS

MIEMBROS DE LA COMISIÓN:

ANA MARTINS BOTTO DE BARROS

ROLANDO DE LA CRUZ MESÍA

FERNANADO ORDÓÑEZ PIZARRO

SANTIAGO DE CHILE

2015

Resumen

Una agrupación delictiva es una red ilegal en la que no necesariamente existe una estructura jerárquica, sino más bien una estructura horizontal, en la que se observa resiliencia y un reclutamiento y ubicación de los miembros basada en la naturaleza de la actividad, dada por las relaciones preexistentes de trabajo, familiares, étnicas, contactos etc. Un enfoque de red social ofrece un medio eficiente y eficaz para identificar, analizar y explicar el fenómeno de grupos delictuales, pues las relaciones humanas constituyen el mínimo común denominador.

En el análisis de agrupaciones delictivas bajo un enfoque de red social son dos los temas de gran interés: la identificación de individuos claves en la red y la búsqueda de asociación. La identificación de individuos claves permite identificar individuos con alta influencia en la red, lo que se relaciona a un rol clave en una agrupación delictiva. La búsqueda de asociación permite dejar al descubierto relaciones entre individuos, que a simple vista, parecieran no existir, facilitando la identificación de grupos delictivos. Los mecanismos tradicionales que permiten llevar a cabo estas tareas, consideran como información esencial aquella que permite establecer un vínculo entre individuos. Esta investigación plantea un nuevo enfoque, la incorporación de atributos que describen información inherente a cada individuo y que no es considerada para establecer una relación entre ellos. Esta información puede estar comprendida por antecedentes socioeconómicos, antecedentes delictivos, antecedentes educacionales, entre otros utilizados y que son fundamentales para el éxito de una investigación policial.

Mediante este enfoque, se propone un nuevo evaluador de la importancia de los nodos y un nuevo modelo para la búsqueda de asociación. En nuevo evaluador es llamado *Social Network Criminal Suspect Evaluato (SNCSE)* y para su elaboración se utilizó una novedosa perspectiva, basada en conceptos de la teoría de capital humano y social, en una estructura de análisis basada en una ego red y en una analogía entre la interacción social y la teoría de campos. El *SNCSE* fue aplicado y sus resultados comparado con evaluadores tradicionales, mostrando un mejor comportamiento y permitiendo concluir que la incorporación de atributos individuales permite identificar individuos sospechosos claves con más precisión. El nuevo modelo de búsqueda de asociación es llamado *Linear Rational Association Model (LiRAM)* e identifica la mejor asociación entre dos individuos, maximizando la función de utilidad de uno de ellos. El *LiRAM* es aplicado y sus resultados son comparados con otro método de asociación, dando cuenta de su eficacia y flexibilidad.

Dedicado a mi familia

Margarita, Esperanza, Paz, Consuelo y Roberth

Prefacio

Esta tesis ha sido desarrollada como requisito para la obtención del grado de Doctor en Sistemas de Ingeniería de la Universidad de Chile. Para su desarrollo se ha contado con el valioso apoyo de diversas instituciones entre las cuales se encuentra la Universidad del Bío-Bío, Conicyt a través de su programa de becas para doctorados nacionales, Instituto Sistemas Complejos de Ingeniería, Proyecto Anillo ACT087 “Quantitative methods in security”, dirigido por el Centro de Análisis y Modelamiento en Seguridad CEAMOS (www.ceamos.cl) y Policía de Investigaciones de Chile.

Agradecimientos

Agradezco a Dios por la oportunidad que me entregó para realizar estudios de postgrado. Agradezco a mi esposa Margarita y a mis hijas Esperanza, Paz y Consuelo por su amor, apoyo incondicional y todo el valioso tiempo que me han concedido. Agradezco a mis padres, hermanos, suegros y familiares en general por su continua preocupación.

Agradezco al profesor Richard Weber por su invaluable ayuda durante estos años en los que he permanecido en el programa de doctorado en sistemas de ingeniería, por creer en este trabajo y entregarle una parte valiosa de su tiempo.

Agradezco a la Universidad del Bío-Bío por el espacio de perfeccionamiento académico entregado. Agradezco el apoyo permanente del Instituto Sistemas Complejos de Ingeniería, del Centro de Análisis y Modelamiento en Seguridad CEAMOS y de CONICYT por la beca para estudios de doctorado otorgada.

F.T.

Tabla de Contenido

Índice de figuras

Capítulo 1

Introducción

En el siguiente capítulo se analizan los principales aspectos relacionados con la investigación de agrupaciones delictivas bajo el contexto de una red social y se describen las principales herramientas utilizadas para su investigación. Un nuevo enfoque de análisis es propuesto, entregando el marco general sobre el cual se basa esta investigación. Posteriormente se plantea el objetivo general y los objetivos específicos. Al final del capítulo se describen los principales aportes originales logrados y la organización del trabajo.

1.1. Redes sociales

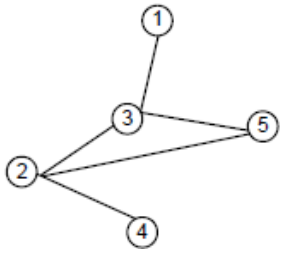
1.1.1. Definición de red

Una red es un grafo. En teoría de grafos, un grafo es definido como el conjunto $G = (N, A)$, donde N es el conjunto de vértices y A es el conjunto de aristas, $|N| = n$ y $|A| = m$ $m \leq n^2$. Los vértices pueden ser llamados nodos, puntos, objetos o individuos. Las aristas pueden ser llamadas arcos, vínculos o líneas. Un grafo puede ser dirigido o no dirigido dependiendo de la vinculación del origen y el destino. Un grafo puede también ser ponderado o no ponderado dependiendo si a cada vínculo está asociado una etiqueta numérica llamada ponderación.

Un grafo es representado tradicionalmente en forma gráfica y matricial. En la Fig.1.1a se muestra un grafo no dirigido consistente de cinco nodos. Los círculos representan los nodos y las líneas entre ellos los vínculos. Los nodos pueden ser etiquetados, por ejemplo, con números como en este caso.

Un grafo puede ser representado por una matriz cuadrada de $n \times n$, donde las filas y las

columnas representan el conjunto N de nodos. El valor de la celda (i, j) es la ponderación de un vínculo entre el nodo i y j. En la Fig.1.1b los valores 1 o 0 de las celdas indican la presencia o ausencia de un vínculo.



a

	1	2	3	4	5
1	0	0	1	0	0
2	0	0	1	1	1
3	1	1	0	0	1
4	0	1	0	0	0
5	0	1	1	0	0

b

Figura 1.1: Representación gráfica (a) y matricial (b) de una red.

1.1.2. Definición de red social

Es posible definir una red social como 'la estructura relacional de un grupo o sistema social más amplio que consiste en el patrón de las relaciones entre la colección de actores"[31]. Representar un sistema social mediante una red social atiende a una de las ideas más potentes en las ciencias sociales, la cual es la noción que los individuos están integrados a redes de relaciones e interacciones sociales [2].

Una forma adecuada para la representación de una red social es mediante un grafo el cual es llamado sociograma [21]. De esta forma una red social puede ser expresada en forma gráfica y matricial. Los nodos del grafo representan los individuos y los arcos la interacción social entre ellos.

1.2. Agrupaciones delictivas bajo el contexto de red social

Al hablar de grupo delictivo, se habla de una red ilegal, en la que no necesariamente existe una estructura jerárquica, sino más bien una estructura horizontal, en la que se observa resiliencia y un reclutamiento y ubicación de los miembros, basado en la naturaleza de la actividad, en las relaciones pre existentes de trabajo, de tipo familiar, étnicas, de contactos, entre otras [8].

Las principales investigaciones se han desarrollado en la línea econométrica, microeconómica y bajo el contexto de una red social. En la línea econométrica, por ejemplo, se han es-

tablecido modelos de regresión que han permitido, obtener índices de propensión delictiva y se ha analizado su relación con otros índices como victimización [13] o violencia [20]. En la línea microeconómica, se han propuesto modelos que permiten entender el comportamiento de estas agrupaciones delictivas [11, 7] y la elección racional que explica la incorporación de un delincuente a una agrupación de este tipo [3].

Bajo el contexto de una red social, una agrupación delictiva se entiende como una red en la cual los nodos representan los integrantes y los arcos los vínculos existentes entre estos. Los vínculos entre individuos sirven como canales de transferencia o flujo de recursos materiales y/o inmateriales [19]. Bajo esta perspectiva, los integrantes de una agrupación delictiva y sus acciones, no son considerados como unidades autónomas, sino como unidades interdependientes.

Analizar y explicar el fenómeno de agrupaciones delictivas mediante un enfoque de red social, es un medio eficiente y eficaz, pues sus actores participan en el proceso de creación de redes sociales para la provisión de bienes y servicios ilícitos, donde se protege, regula y extorsiona a los participantes tanto del suministro como del consumo de estos bienes y servicios [19]. Ejemplo de esto ha sido su intensiva aplicación en el análisis de agrupaciones terroristas después del atentado del 11 de septiembre de 2001 [32, 33, 23, 38, 37, 25, 27, 26, 16].

1.3. Investigación policial de agrupaciones delictivas bajo el contexto de red social

En una investigación policial se quiere el despliegue de una gran cantidad de recursos humanos y técnicos de manera de encontrar el o los culpables de un determinado crimen. En la medida que los delitos adquieren más complejidad, la investigación requiere un mayor conocimiento, tecnología, experiencia y tiempo [35].

Una investigación policial orientada a la identificación de agrupaciones delictivas parte, por lo general, con un conjunto de individuos sospechosos, donde a mayor número, las alternativas de selección de un subconjunto de individuos para comenzar la investigación pueden ser variadas. Cada alternativa de investigación, requiere el despliegue de distintas cantidades y tipos de recursos, no garantizándose la consecución de resultados satisfactorios. Dado el carácter limitado de los recursos humanos, físicos y del tiempo destinado a la investigación policial, es

necesario contar con sistemas que apoyen la toma de decisiones y que permitan generar alternativas de investigación que permitan obtener resultados satisfactorios con alta probabilidad, mediante el uso eficiente de los recursos. La utilización de un contexto de red social para la investigación de agrupaciones delictivas, ha permitido la creación de este tipo de sistemas, como COPLINK Detect [14] creado para la elaboración de una red mediante el establecimiento de vínculos entre sospechosos, víctimas y otros elementos pertinentes.

El establecimiento de una red que represente las relaciones sociales existentes entre los individuos, es el primer desafío de una investigación policial bajo este contexto, lo que permite la posterior aplicación de técnicas de análisis de redes sociales para la explotación de la información contenida en ella.

El vínculo entre individuos por lo general no es explícito, por lo que es necesario un trabajo muy intensivo y demandante de gran parte del tiempo destinado a la investigación, para el procesamiento de la información disponible del conjunto de individuos considerados [35]. La información disponible puede ser extraída desde diferentes medios como: bases de datos, registros de declaraciones, movimientos de cuentas bancarias, grabaciones telefónicas, correos electrónicos, fotografías, filmaciones, llamadas desde celulares, entre otras. El trabajo de extracción y transformación de esta información en vínculos, ha sido tradicionalmente conocido como link análisis [28] [4].

Extraer y transformar los datos que describen la actividad humana para el establecimiento de redes sociales, ha sido uno de los principales problemas en la minería de la estructura de redes sociales [12]. Dentro de la minería de la estructura de redes existen dos áreas principales: minería de estructura estática y minería de estructura dinámica [36]. La minería de estructura estática estudia una “foto instantánea"de la red, esto es, nodos y vínculos son observados en un instante específico del tiempo. La minería de estructura dinámica analiza la estructura de la red basado en datos observados en múltiples puntos del tiempo. Un análisis estático es logrado descubriendo regularidades estructurales en una configuración específica de nodos y vínculos de una red en el momento de la observación. Un análisis dinámico es logrado al encontrar patrones de cambios en la red a lo largo del tiempo. El foco del análisis estático es en la estructura, mientras que el foco del análisis dinámico es en el proceso y en el mecanismo de evolución que sigue la estructura.

Es posible resumir en cuatro las principales técnicas [17] que permiten transformar la in-

formación para el establecimiento de un vínculo representativo entre individuos. Estas técnicas son:

- Self-Report: donde el vínculo se establece por el reporte de actores individuales.
- Comunicación: donde el vínculo se establece por la comunicación o transferencia de recursos entre dos individuos, es decir, la evidencia de comunicación puede indicar asociación.
- Similitud: donde el vínculo se establece sobre la base que los amigos tienden a ser similares en cuanto a su comportamiento social, por lo que un comportamiento similar entre los individuos, puede indicar asociación.
- Co-ocurrencia: donde el vínculo se establece bajo la idea que si muchos individuos comparten ocurrencias conjuntas, más que al azar, pueden estar relacionados.

Este trabajo se centra en al análisis de redes obtenidas mediante la minería de estructura estática, por lo que el establecimiento de los vínculos, representará el sistema social durante el periodo en que se registró la información del conjunto de individuos que la compondrá. Para la explotación de la información contenida en la red en forma efectiva, es necesario utilizar herramientas adecuadas de análisis estático. En la siguiente sección se describen las principales técnicas empleadas para este fin.

1.4. Técnicas para el análisis de redes sociales

En el contexto de una red social, las principales técnicas utilizadas para la investigación de agrupaciones delictivas son: la identificación de asociaciones y la identificación de individuos claves o con roles importantes dentro de la red. Dentro de ambas técnicas existen un conjunto de herramientas que resultan especialmente apropiadas, las cuales se describen a continuación.

1.4.1. Evaluación de la importancia de los nodos

La evaluación de la importancia de los nodos es utilizada para identificar individuos claves o con roles importantes dentro de la red. Esta evaluación permite obtener una medida de la *centralidad* de un nodo dentro de la red [10]. Así, un nodo será relativamente más importante

que otro, si su medida centralidad es mayor. Evaluadores eficiente y eficaces, tradicionalmente utilizados [19] son las medidas de centralidad del *Social Network Analysis (SNA)* y los algoritmos de evaluación de nodos.

Medidas de centralidad del Social Network Analysis

Las medidas de centralidad del SNA hacen referencia a aspectos estructurales diferentes de la red y pueden ser usadas en forma complementaria con gran eficacia [19]. Las medidas comúnmente utilizadas [18, 29] son:

- *Degree:* mide la proporción de nodos que están unidos a un nodo i. Esta medida está relacionada a la comunicación que tiene un nodo en la red con otros. Su valor es obtenido mediante la siguiente ecuación:

$$Degree(i) = \frac{deg(i)}{(n-1)} \tag{1.1}$$

  Donde la $deg(i)$, representa el número de nodos conectados directamente al nodo i, y n representa el total de nodos de la red.

- *Betweenness:* mide la proporción de rutas más cortas entre dos nodos que pasan por un nodo i. Esta medida de centralidad se relaciona al potencial de un nodo para el control de la comunicación. Su valor es obtenido mediante la siguiente ecuación:

$$Betweenness(i) = \sum_{s \neq i \neq t \in N} \frac{l_{st}(i)}{l_{st}} \tag{1.2}$$

  Donde $l_{st}(i)$, es el número de rutas más cortas entre el nodo s y t que pasan por un nodo i y l_{st} es el número total de rutas más cortas entre los nodos s y t.

- *Closeness:* mide la cercanía de un nodo i a otros nodos de la red. Esta medida está relacionada a la independencia de un nodo o a su eficiencia en la comunicación. Su valor es obtenido mediante la siguiente ecuación:

$$Closeness(i) = \sum_{j \in N/i} d_{G(i,j)} \tag{1.3}$$

Donde $d_G(i,j)$, es la distancia geodésica o distancia más corta entre un nodo i y un nodo j.

- *Eigenvector:* Es una medida de centralidad que considera el hecho que la importancia de las conexiones entre un nodo i y otros nodos no son iguales (a diferencia de degree) [24]. Esta medida de centralidad considera la influencia del vecindario de un nodo en su evaluación, de esta forma, un nodo que tiene un alto valor para esta medida de centralidad, será aquel cuyos nodos adyacentes también tienen un alto valor. Esta medida es obtenida mediante la siguiente ecuación:

$$x_i = \frac{1}{\lambda} \sum_{j \neq i \in N} A_{ij} x_{ij} \tag{1.4}$$

Donde x_i denota la centralidad del nodo i, λ es una constante y

$$A_{ij} = \begin{cases} 1 & \text{si i está conectado con j} \\ 0 & \text{si } no \end{cases} \tag{1.5}$$

Al definir el vector de centralidades o Eigenvector como $X = \{x_1, x_2 ... x_n\}$ la Ec.(1.4) puede ser escrita en forma matricial como:

$$\lambda X = A \cdot X \tag{1.6}$$

Algoritmos de evaluación de la importancia de los nodos

Dentro de los algoritmos de evaluación de la importancia de los nodos se encuentran: Page Rank [22], Hits [15] y Topological Potential [30]. Page Rank e Hits, son algoritmos creados como motores de búsqueda de páginas web y Topological Potential ha sido desarrollado y utilizado durante la última década.

- *PageRank:* es un algoritmo creado como motor de búsqueda de páginas web y es utilizado también para obtener el valor de la importancia de un nodo i [22]. La medida de centralidad entregada por este algoritmo se realiza mediante la siguiente función:

$$PR_i = \frac{1-d}{n} + d \sum_{j \in M(i)} \frac{PR_j}{L_j} \tag{1.7}$$

Donde $M(i)$ es el conjunto de páginas o nodos unidos a la página o nodo i. L_j es el número de hipervínculos o arcos salientes desde la página o nodo j. El valor de PR_i representa la probabilidad de estar en la página o nodo i y d indica la probabilidad de seguir un hipervínculo o arco de la actual página o nodo i. PageRank modela la navegación web como un paseo aleatorio, donde un navegante aleatorio selecciona y sigue hipervínculos saltando ocasionalmente a nuevas páginas web para comenzar otro recorrido en una estructura de hipervínculos. El ranking entregado, bajo este contexto, es la fracción de tiempo que el navegante podría gastar en la página si el proceso aleatorio iterara al infinito. Este puede ser determinado mediante el cálculo de la distribución de estado estable del proceso aleatorio[12].

- *Hits:* es también un algoritmo creado como motor de búsqueda de páginas web, utilizado para obtener el valor de la importancia de un nodo i [15]. En este algoritmo, las páginas o nodos referenciados por algún hipervínculo o arco son llamadas Authorities y la página web o nodos que apunta a esta página son llamadas Hub. La medida de centralidad entregada por este algoritmo se realiza mediante la siguiente función:

$$a(i) = \sum_{j \in B(i)} h(j) \tag{1.8}$$

$$h(i) = \sum_{j \in I(i)} a(j) \tag{1.9}$$

Donde $a(i)$ representa el nivel de Authority y $h(i)$ representa el nivel de Hub. $B(i)$ representa el set de páginas o nodos referidos de la página o nodo i e $I(i)$ representa el set de páginas o nodos de referencia de la página o nodo i. Los niveles de Authority y Hub son determinados mediante iteraciones que actualizan los niveles de Authority y Hub de una página, basado en los niveles de Authority y Hub de las páginas de su vecindario inmediato [12].

- *Topological Potential:* Este algoritmo utiliza una analogía de la interacción social entre nodos, con la interacción entre partículas descrita por la Teoría de Campos [30]. Cada nodo en la red es visto como una partícula que crea un campo a su alrededor, interactuando con todos los otros nodos, formando un campo topológico sobre la red. La medida de

centralidad de un nodo i es entregada mediante una función del tipo gaussiana, la cual representa el potencial de una partícula, como se muestra a continuación:

$$\varphi(V_i) = \frac{1}{n} \sum_{j=1}^{n} e^{-\left(\frac{d_{ij}}{\sigma}\right)^2} \tag{1.10}$$

Donde d_{ij} es la distancia topológica entre los nodos *i,j* y σ es un parámetro que controla la región de influencia de un nodo. Con el propósito de lograr el mejor resultado del potencial topológico, se debe realizar una selección adecuada del parámetro σ. Una técnica basada en la minimización de la entropía permite determinar un valor un valor óptimo de σ para una red en particular.

1.4.2. Búsqueda de patrones de asociación

Los patrones de asociación bajo el contexto de una red social pueden ser encontrados entre individuos y subgrupos de individuos. La identificación de patrones de interacción permite dejar al descubierto relaciones entre individuos o subgrupos de individuos, que no son identificables a simple vista y que pueden resultar esenciales para la obtención de buenos resultados en una investigación policial. A continuación se mencionan tres técnicas estudiadas:

Patrones de asociación entre grupos

- *Blockmodel analysis:* dada una red particionada, blockmodel analysis determina la presencia o ausencia de una asociación entre un par de subgrupos mediante la comparación de la densidad de los vínculos entre ellos bajo un valor predefinido de umbral [32]. De esta forma, Blockmodel analysis introduce un resumen de los detalles de la interacción individual entre grupos de manera de que la estructura completa de la red llega a ser más evidente.

  Un blockmodel consiste en un mapeo de actores aproximadamente equivalentes dentro de blocks o posiciones o una declaración referente a las relaciones entre las posiciones o blocks. Los datos estudiados mediante un análisis relacional consiste de R variables relacionales, las cuales son medidas en g actores donde:

$$X_{ijr} = \begin{cases} c & \text{si } \textit{el actor i se relaciona con j al nivel c de la variable r} \mid r \in R, i \neq j \\ 0 & \text{si } no \end{cases} \tag{1.11}$$

Estas variables son agrupadas en R grupos de matrices de $g \times g$, donde $X_1, X_2, ...X_R$ representan estas matrices. El enfoque estándar de un analisis blockmodel, busca patrones con respecto a sus accesos, mediante permutaciones simultáneas de filas y columnas en estas matrices. Se buscan las particiones de los g actores en B posiciones tales que los actores que son aproximadamente equivalentes (los actores que muestran los mismos patrones de entradas en las filas y columnas en las matrices) se asignan a la misma posición [1].

Patrones de asociación entre individuos

- *Algoritmo modificado de la ruta más corta:* este método identifica el patrón de interacción entre dos individuos dentro de una red mediante la identificación de ruta más corta de máximo vínculo [34]. Para esto se asume la magnitud del vínculo como una medida de probabilidad que indica la probabilidad de que dos individuos estén relacionados, asumiendo también que la vinculación entre individuos son eventos independientes. Así, dos individuos aparentemente no vinculados podrían estarlo, mediante una ruta consistente de individuos intermedios vinculados. La probabilidad de existencia de esta ruta o asociación estará representada por el producto de las probabilidades o vínculos pertenecientes a la ruta. La mejor ruta o asociación entre dos individuos será aquella con mayor probabilidad (la ruta más probable). En la Fig. 1.2 se muestra un ejemplo donde existen dos rutas o asociaciones probables entre el nodo origen A y el nodo destino D, $(A-B-C-D)$ y $(A-E-D)$. El valor de cada arco representa la probabilidad de que dos individuos estén relacionados. La probabilidad de existencia de la asociación $(A-B-C-D)$ es 0.28 (0.5 $\times$0.8$\times$0.7), y la probabilidad de existencia de la asociación $(A-E-D)$ es 0.24 (0.8 $\times$0.3), por lo que la asociación más probable está dada por la ruta $(A-B-C-D)$.

  Este enfoque, permite identificar la asociación más probable entre dos individuos utilizando un algoritmo de la ruta más corta. Para esto es necesario realizar una transformación logarítmica del vínculo, como se muestra en la siguiente ecuación:

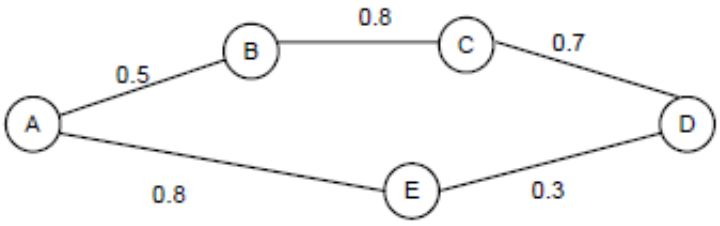


Figura 1.2: Conexión indirecta de los nodos A y D.

$$l = -ln\, w \quad 0 < w \leq 1 \tag{1.12}$$

Donde w representa el vínculo entre dos individuos.

La transformación logarítmica de la Ec.(1.12) permite obtener las siguientes propiedades [34]:

- Asegurar que el nuevo grafo no contenga valores negativos y poder aplicar cualquier algoritmo de la ruta más corta sin problemas [9].
- Que un bajo vínculo en la nueva red corresponda a un alto vínculo en la red original.
- Que la ruta más corta (mediante la suma de los vínculos) entre un par de nodos i y j, genere la ruta de máximo producto o más probable de entre las rutas posibles.

- *Edge-dual graph and k-connectivity concepts:* este método permite medir la fuerza de la asociación entre dos individuos [5] [6]. Para esto se transforma la red original en una red llamada *red edge-dual*, en la cual las relaciones únicas entre dos nodos, son reemplazadas por un nodo llamado nodo relación, como se muestra en la Fig. 1.3. Una vez transformada la red, la medición de la fuerza de la relación entre dos nodos es realizada utilizando conceptos de k-conectivity, los cuales permiten medir la cohesión de un red. Se dice que una red está k conectada si es posible remover no menos de k nodos para dividir el grafo en dos o más partes. El concepto que aquí se aplica para medir la fuerza de la asociación entre dos individuos, es el de conectividad local a nodos. Dados dos nodos i y j, la fuerza de la conexión entre ellos, estará determinada por el número k de nodos relación mínimo que deben ser removidos para desconectar i de j. Esto implica que mientras mayor sea el valor de k mayor será la fuerza de la asociación entre dos nodos. Este problema de conec-

tividad es resuelto mediante una modificación al algoritmo tradicional de flujo máximo. El algoritmo tradicional de flujo máximo es usado para calcular la máxima conectividad de los arcos desde un nodo fuente a un nodo destino y la modificación es realizada para calcular la máxima conectividad de nodos relación (cuántos nodos relación es requerido remover para desconectar el nodo fuente del nodo destino).

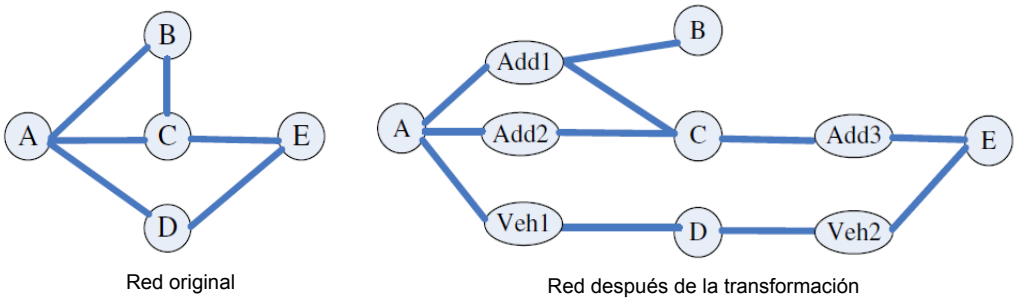


Figura 1.3: Transformación de una red en una red edge-dual.

1.5. Un nuevo enfoque para el análisis de agrupaciones delictivas bajo el contexto de red

En esta sección se propone un nuevo enfoque para abordar el análisis de redes delictivas. En base a este nuevo enfoque se plantea el objetivo general, los objetivos específicos de la investigación y los aportes originales alcanzados.

1.5.1. El nuevo enfoque

Las técnicas de análisis de redes sociales investigadas, consideran el vínculo entre individuos como la principal fuente de información, sin dar lugar a la incorporación de información complementaria que permitan mejorar la efectividad en la identificación de individuos claves y de asociaciones. El enfoque que plantea este trabajo es considerar en el análisis de agrupaciones delictivas información complementaria existente en los antecedentes personales de los individuos de la red, para lo cual se propone un nuevo evaluador de la importancia de los nodos y un nuevo modelo para la búsqueda de asociación.

Los antecedentes personales a considerar en el análisis de agrupaciones delictivas, son aquellos que dan cuenta de la adquisición de conocimientos y habilidades para cometer cierto tipo de delito grupal. Estos antecedentes personales son expresados mediante un conjunto de atributos relevantes que caracteriza a cada individuo y que son resumidos en un indicador de propensión individual para cometer un delito grupal (Pcg), definido por:

$$Pcg_i = r(S_i) \tag{1.13}$$

Donde S_i es el conjunto de atributos relevantes de un individuo i y r es una función, elegida bajo cierto contexto, que transforma este conjunto en un valor de propensión delictiva.

1.5.2. Objetivos de la investigación

Objetivo general

La presente investigación tiene como objetivo principal aportar al desarrollo de técnicas para el análisis de agrupaciones delictivas en el contexto de una red social, las que incorporarán además del vínculo entre individuos, información complementaria existente en sus antecedentes personales, resumida mediante un indicador de propensión individual para cometer un delito grupal. El desarrollo de esta investigación aborda dos técnicas investigadas: evaluación de la importancia de los nodos, para la identificación de individuos claves dentro de la red y la búsqueda de patrones de asociación, para identificar interacción entre individuos aparentemente no relacionados.

Objetivos específicos

Los objetivos específicos de la investigación son:

- Establecer una base teórica que permita estructurar y modelar la información disponible para el análisis de agrupaciones delictivas bajo el contexto de una red social considerando antecedentes personales de los individuos de la red.
- Proponer un evaluador de la importancia de los nodos que permita la identificación de individuos claves o con roles importantes dentro de la red, considerando antecedentes

personales mediante un indicador de propensión individual para cometer un delito grupal.

- Proponer un modelo para la búsqueda de patrones de asociación entre individuos dentro de una red que permita identificar la interacción entre individuos no relacionados, considerando antecedentes personales mediante un indicador de propensión individual para cometer un delito grupal.

- Aplicar el evaluador de la importancia de los nodos y el modelo para la búsqueda de patrones de asociación propuesto, utilizando información real para el establecimiento de vínculos, para la creación de la red y para la obtención de la propensión delictual.

1.6. Resultados de la Investigación

Los principales resultados de esta investigación, son los siguientes trabajos enviados a revistas científicas:

- *A Social Network Approach to Identifying Key Police Suspects*, enviado a la revista Journal of Quantitative Criminology.

- *A Decision Support Model for Detecting Criminal Network Associations Using Personal Attributes*, enviado a la revista Decision Support Systems.

Adicionalmente, se presentaron los siguientes trabajos en distintos congresos:

- Troncoso, F., Weber, R.: Modelo de optimización para detectar grupos delictuales en redes sociales. Workshop de la Red Chileno-Argentina de Gestión de Operaciones (ChAGO), Santiago, Chile, Marzo 2015.

- Troncoso, F., Weber, R.: A social network approach to identifying key police suspects using data mining. 20th Conference of the International Federation of Operational Research Societies. Barcelona, España, 2014.

- Troncoso, F., Weber, R.: A Social network approach to identifying key police suspects. Workshop on Analysis and Modeling of Security (WAMOS), Santiago, Chile, Abril 2014.

- Troncoso, F: Identificación de individuos sospechosos en búsqueda de un patrón de agrupación delictiva bajo un enfoque de red social. X Optima VI RED-M, Universidad de Concepción, Chile, Octubre 2013.

1.7. Aportes Originales de la Investigación

El resto de esta Tesis se encuentra dividido en los siguientes capítulos, donde se describen los aportes originales de la investigación:

Capítulo 2: A Social Network Approach to Identifying Key Police Suspects

Dentro de los métodos utilizados para el análisis de agrupaciones delictivas, bajo el contexto de una red social, se encuentra la evaluación de la importancia de los nodos, cuya aplicación ha resultado apropiada para la identificación de individuos claves. Bajo el planteamiento de un nuevo enfoque, donde además de la información entregada por los vínculos se considera la información incorporada en los antecedentes personales de los individuos, es necesario un nuevo evaluador de la importancia de los nodos. Este trabajo propone un nuevo evaluador de la importancia de los nodos que considere estos antecedentes personales mediante un índice de propensión delictiva de manera de mejorar la efectividad de una investigación policial de agrupaciones delictivas. El evaluador propuesto es llamado *Social Network Criminal Suspect Evaluato (SNCSE)* y para su elaboración se utilizó una novedosa perspectiva, basada en conceptos de la teoría de capital humano y social, en una estructura de análisis basada en una ego red y en una analogía entre la interacción social y la teoría de campos. El *SNCSE* es aplicado a dos casos reales, considerando uno de ellos para medir su efectividad y comparar su desempeño con el de los evaluadores tradicionales del Social Network Analysis y con el de dos algoritmos de evaluación. En general el *SNCSE*, presentó un mejor desempeño que los evaluadores considerados. El análisis de los resultados revela que la integración de la propensión delictiva y los vínculos permite determina en forma más exacta individuos claves de la red.

Capítulo 3: A Decision Support Model for Detecting Criminal Network Associations Using Personal Attributes

Otra técnica utilizada para el análisis de agrupaciones delictivas, bajo el contexto de una red social, es la identificación de patrones de asociación entre individuos, la cual permite dejar en evidencia relaciones entre individuos no identificables a simple vista. Los individuos incorporados en la asociación pueden ser considerados claves en un análisis policial de agrupaciones delictivas. Para esto, el vínculo entre individuos es la única fuente de información, quedando fuera de este proceso información relevante como los antecedentes personales de los individuos de la red. Este trabajo propone un modelo para la identificación de patrones de asociación entre individuos considerando la incorporación de estos antecedentes personales. Este nuevo modelo es llamado *Linear Rational Association Model (LiRAM)* e identifica la asociación entre dos individuos maximizando la función de utilidad de uno de ellos. El LiRAM es aplicado mediante el uso de información real para la creación de la red de aplicación y para la obtención de la propensión delictual. Sus resultados son comparados con los resultados obtenidos mediante otro método de búsqueda de asociación. En análisis de los resultados da cuenta de la eficacia del LiRAM y del aporte de flexibilidad al análisis de agrupaciones delictivas.

1.8. Conclusiones Generales y Trabajos Futuros

La investigación policial de agrupaciones criminales requiere de una gran despliegue de recursos que deben ser enfocados en la investigación de individuos claves, de manera de detectar la existencia de agrupaciones criminales o prevenir su formación. Mediante la definición de una agrupación criminal como una red social es posible identificar estos individuos claves utilizando dos técnicas principales: la evaluación de la importancia de los nodos y la búsqueda de asociación.

Esta investigación propone dos modelos que fortalecen y enriquecen las estas técnicas existentes de manera de mejorar la efectividad de la aplicación de este tipo de técnicas en la identificación de individuos sospechosos de participar en agrupaciones delictuales. El primer modelo es un evaluador de la importancia de los nodos llamado *Social Network Criminal Suspect Evaluator (SNCSE)* y el segundo un modelo para la búsqueda de asociación llamado *Linear Rational Association Model (LiRAM)*.

El aporte original de estos modelos radica en su desarrollo y en la incorporación de información policial referente a atributos de los individuos de la red, representados como un índice de propensión individual de pertenecer a una agrupación criminal. Este índice de propensión se considera en forma complementaria al vínculo entre los individuos, considerado en los modelos tradicionales.

El buen desempeño de ambos modelos se debe precisamente a la incorporación de los atributos individuales. Estos buenos resultados pueden tener un impacto significativo en los resultados de corto plazo de una investigación policial lo que permitiría una mejora en la eficiencia del uso de los recursos utilizados en la investigación.

La evaluación entregada por el *SNCSE* es generada mediante dos fuentes: la información de las relaciones entre los individuos en la red y los atributos de cada individuo en la red. El hecho que el *SNCSE* incorpora en la evaluación de un nodo la información de su vecindario, sugiere que los individuos relacionados directamente con aquellos mejor evaluados también puedan estar relacionados a una agrupación criminal.

La asociación entre dos individuos entregada por el *LiRAM* es el resultado de un proceso de decisión en el cual uno de ellos, llamado planificador criminal, escoge de entre los individuos de la red a quienes le permiten maximizar su utilidad en base a la capacidad delictual representada por el índice de propensión y en base a la confianza representada por el vínculo social entre individuos. Este modelo ofrece flexibilidad al análisis de agrupaciones criminales al permitir generar diferentes alternativas de asociación entre dos individuos.

Como trabajo futuro se consideran las siguientes líneas:

- El desarrollo de una metodología para la detección de comunidades en sub redes compuestas por las ego redes de los individuos mejor evaluados mediante el SNCSE.
- Considerar la magnitud del vínculo entre dos individuos como medida de probabilidad de la existencia de su relación, asumiendo que estas probabilidades son independientes entre los individuos [34]. Este supuesto permitiría generar una función de utilidad no lineal que podría mejorar los resultados obtenidos con el LiRAM.
- La incorporación conjunta de propensiones a crímenes grupales complementarios que provean diferentes escenarios para el análisis policial.

Se observan las siguientes restricciones en ambos modelos, las cuales pueden ser usadas como punto de partida para futuras investigaciones en esta área:

- La falta de datos de buena calidad para en el estudio cuantitativo de la delincuencia. Los modelos propuestos asumen la disponibilidad de datos que no necesariamente se tendrán para en una aplicación real. Por esta razón se hace necesario investigar cómo modelar la información disponible, de manera que ambos modelos puedan ser implementados y cómo considerar la incertidumbre en aplicaciones en donde los datos no pueden ser conocidos o modelados con exactitud.

- Ambos modelos parten de supuestos, los cuales son necesarios de analizar en tipos de crimen específicos. También es necesario analizar la posibilidad de establecer una generalización para abordar el fenómeno de agrupaciones delictivas de manera más realista.

Capítulo 2

A Social Network Approach to Identifying Key Police Suspects

A Social Network Approach to Identifying Key Police Suspects

Fredy Troncoso · Richard Weber

Abstract One of the most common methods used with social network analysis of criminal groups is node importance evaluation, which focuses on the links between network members to identify likely crime suspects. This study introduces a new approach that incorporates members' individual criminal propensities. Since testing of traditional node evaluators found that they tended not to take full advantage of this additional information, a more efficient evaluator called the Social Network Criminal Suspect Evaluator (SNCSE) is proposed. It employs a novel perspective based on concepts of human and social capital, a structure built on ego networks and an analogy between social interaction and field theory. The SNCSE is applied to two real-world cases, one of which serves to test the effectiveness of the proposed formulation by comparing its performance with that of traditional evaluators. The results show that the SNCSE performs better than the others and that the integration of the criminal propensity factor into a criminal network enables the evaluator to identify key criminal suspects more accurately.

Keywords Crime analytics · Criminal groups · Social Networks · Node Evaluation · Human and Social Capital · Field Theory

1 Introduction

Police investigation of individual and group criminality requires large quantities of resources and demands ever greater amounts of expert knowledge, technology, experience and time as criminal behavior becomes more sophisticated. One way of boosting the efficiency and effectiveness of investigative work would be to improve the identification of individual suspects for any given crime. This would enable authorities responsible for public safety and crime prevention to better focus their resources on the most likely candidates and drop pursuit of the least likely ones. The benefits could be particularly significant in cases where the initial population of possible suspects is large.

Criminal groups can be understood as social networks, implying that social network analysis (SNA) can be successfully used to analyze and explain them. In traditional social network methods, the links joining network members are the main analytical element. Crime analytics extracts information from social networks using techniques such as node importance evaluation to identify key individuals. As well as the information used by police to establish the links between individuals, additional records often exist on their personal backgrounds that reveal each one's propensity to commit certain types of offences. These data have the potential to complement those on the links between them and their inclusion in crime analytics could therefore significantly enhance the quality of police investigative work.

The present study proposes a new approach based on the incorporation of the criminal propensities of individual network members into criminal group SNA using the full set of personal data available on each

F. Troncoso
Departamento de Ingeniería Industrial , Facultad de Ingeniería, Universidad del Bío- Bío, Concepción, Chile.
Tel.: +56-041-3111381
E-mail: ftroncos@ubiobio.cl

R. Weber
Departamento de Ingeniería Industrial, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Santiago, Chile.
E-mail: rweber@dii.uchile.cl

one. With the integration of this information, however, methods such as node importance evaluation may no longer be appropriate given that they consider the link as the only magnitude of interest. The approach proposed here offers a novel node importance evaluator that takes account of criminal propensities as well as links, generating results that provide greater support for police work, particularly in criminal group analysis.

The remainder of this article is organized into four sections. Section 2 provides background information on the application of the social network approach to criminal group analysis, the tools this analysis traditionally uses and the need for a new node importance evaluator that incorporates the criminal propensity of network members. Section 3 develops the proposed new evaluator, discusses its theoretical basis and lays out the conditions for its application. Section 4 describes the application of the evaluator to two real-world cases and formally tests its effectiveness in one of them. Finally, Section 5 sets out the conclusions of this study and some suggestions for future research.

2 Background

A social network may be defined as *the relational structure of a group or larger social system, consisting of the pattern of relationships among the collection of actors* (Wasserman, 1994). Representing a social system as a social network gives concrete expression to one of the most potent ideas in the social sciences, which is that individuals are embedded in webs of social relations and interactions (Borgatti et al., 2009).

A criminal group can be interpreted as a social network in which the nodes represent criminals and the arcs represent the links between them. Links between criminals serve as channels for the transfer or flow of material and/or non-material resources (McIllwain, 1999). On this perspective, criminal group members and their actions are seen not as autonomous units but as interdependent ones.

Using the social network approach to analyze and explain the criminal group phenomenon is efficient and effective because the group members are involved in a process of social networking, both for the provision of illicit goods and services and the protection, regulation and extortion of those involved in their provision and consumption (McIllwain, 1999).

The analysis of criminal groups in a social network context generally aims to identify criminal structures, key individuals and other members in important roles based on the links found to exist between them. To this end, information from databases both public and private is utilized. Published databases on terrorist groups have been widely employed in criminal analysis, especially since the attacks of September 11, 2001 (Xu and Chen, 2003; Xu et al., 2004; Qin et al., 2005; Yang et al., 2006; Yang, 2008; Shaikh and Jiaxin, 2006; Shaikh et al., 2007; Shaikh and Jiaxin, 2008; Lauchs et al., 2012).

In police investigative work using the social network approach, a key step is to establish a representative link between network members. This requires that information describing human behavior garnered from diverse sources be properly modelled and transformed, which is one of the main problems arising in spatio-temporal mining of social networks. Lauw et al. (2005) specify four techniques for transforming information in order to establish a representative link between individuals:

- Self-Report: establishes a link based on the reports of the individual actors.
- Communication: establishes a link based on the communication or transfer of resources between individuals. Evidence of communication thus indicates association.
- Similarity: establishes a link on the basis that friends tend to be similar in their social conduct. Similar behavior among individuals may therefore indicate association.
- Co-occurrence: establishes a link on the notion that if references to individuals occur together more than mere chance would explain, they may be related.

A social network created through this process of establishing links will represent the social system as it was during the period in which the information on the set of individuals comprising the network was recorded. To ensure the analysis is effective, therefore, the network must be updated as new information on the individuals becomes available. This dynamic will be reflected in changes in the network's membership and social relations.

Two tools commonly used to extract information from social networks are the establishment of associations and the evaluation of node importance. The former is undertaken as part of crime investigation efforts aimed at discovering relationships between one or more criminals who are not directly related. Xu and Chen (2004) propose a method for identifying associations using modified shortest-path algorithms

in which the association between two individuals in a network is represented by the path that is the shortest and has the highest link weight.

In a similar vein, Ding and Dixon (2008); Ding et al. (2011) propose an edge-dual graph to transform a traditional social network graph into a relation context oriented graph, using modified k-connectivity concepts to evaluate the robustness of the connection between two nodes. The connectivity number k of a graph is the minimum number of nodes whose removal will divide the graph into two pieces. This means that given two nodes A and B, the robustness of the connection between them is determined by the number of relation nodes that must be removed to disconnect A from B.

As regards the technique of node importance evaluation, this is utilized to identify social structures, key individuals and other important members in a social network. Traditionally used evaluators (McIllwain, 1999) are centrality measures from *Social Network Analysis (SNA)* and node evaluation algorithms. The most common centrality measures(McGloin and Kirk, 2010; van der Hulst, 2009) are degree, closeness, betweenness, density and eigenvector, while algorithms include Page Rank and Hits, both frequently employed, and the recently developed Topological Potential (Wang and Pan, 2012; Getoor and Diehl, 2005). Both approaches are useful in a social network context where the significant data relate to ties between members.

In addition to links, this study posits that information on network members' criminal propensity can also be included. Criminal propensity in the present case will be based on an individual's membership in a criminal group and is thus referred to as the propensity to belong to a criminal group (Pcg). It will be considered as an indicator of an individual's status as a suspect, and social networks used to analyze criminal groups that integrate Pcg will be called suspect networks.

Under this new approach incorporating the Pcg indicator, the evaluation of node importance must assess the degree of suspicion that an individual belongs to a criminal group on the basis of propensity as well as links. It thus takes to heart Robin's warning (Robins (2009)) that representing social systems as social networks without considering the attributes of the individuals at the nodes runs the risk of producing an incomplete and perhaps even spurious analysis. Since the propensity to belong to a criminal group is a node attribute, any available information on it should be exploited to achieve a fuller analysis of the system. Traditional node evaluators, which focus strictly on links, may therefore be inadequate for the task.

In light of the foregoing, and with a view to achieving a more complete and effective analysis of criminal groups, the present study proposes a new node importance evaluator that bases its evaluations on not only the links between individuals but also their respective propensities to belong to a criminal group. This approach delivers an efficient identification of the most important individuals in a suspect network. The evaluator itself and the theoretical notions that underpin it are formally introduced in the following section.

3 A new evaluator under a new approach

The proposed new evaluator emerges from a novel perspective based on concepts borrowed from the theories of human and social capital, an analytic structure built around an ego network, and an analogy between the social interaction of individuals and the interactions of particles in field theory. It is this perspective that will enable the new evaluator to integrate links between individuals with their propensities to belong to a criminal group.

3.1 Criminal group human capital: a suspect evaluator

The ability of an individual to carry out a given economic activity is defined by a set of attributes that reflect his or her acquisition of skills and knowledge over time. These attributes constitute the person's human capital, and the greater is the quality of this capital, the better able he or she will be to identify and take advantage of economic opportunities (Schultz, 1961). Human capital has been the subject of studies from a variety of approaches and perspectives, much of them conducted from a social angle.

An individual's ability to engage in a given criminal economic activity will depend on their human capital for committing certain types of offences. This criminal human capital is defined by knowledge and skills acquired over time that are part of the individual's personal attributes. If these attributes

and criminal human capital can be somehow determined, an individual possessing a high level of human capital for a particular criminal activity can be readily classified as a strong suspect for past or present involvement in it.

Dnes and Garoupa (2010) build a microeconomic model of gang formation and propose the concept of group human capital that measures the contribution of an individual to a criminal group, and more specifically to a gang. According to the model, a criminal group demands a certain minimum level of human capital from each of its members based on minimum required skills and a basic level of commitment to the group. This formulation is corroborated by Chang et al. (2005), whose study indicates that under the assumption criminal organizations pay members as a function of their criminal skills, individuals with medium and high abilities will always tend to join them.

By determining individuals' criminal group human capital, those with the highest levels of such capital can be classed as the individuals most likely to belong to a criminal group. In the context of a suspect network, a metric for measuring criminal group human capital will enable those with the greatest amount of such capital to be identified as the important or most suspect individuals in the network. The proposed evaluator is just such a metric and is set out in what follows.

3.2 An expression for criminal group human capital

Coleman (1988) posits that the determination of an individual's human capital must take his or her social capital into account because it is through the latter that human capital is transferred to the individual from those he or she maintains relations with. An individual's social capital is determined by the complete set of these social relations, which in the context of a social network are given by the network links. The criminal group human capital of an individual in a suspect network is thus determined by the links he or she maintains and the criminal group human capital of those he or she is linked to. The expression we propose for determining an individual's criminal group human capital is therefore a suspect network node importance evaluator.

To write the expression, we first define $G(N, A)$ as a graph representing a social network composed of a set N of nodes or individuals and a set A of arcs or links between the individuals. Then,

$$Hcg_i \;=\; Pcg_i \;+\; HCcg_i \;\; \forall i \in N \tag{1}$$

where

- Pcg_i is the propensity of individual i to belong to a criminal group. This reflects i's initial criminal group human capital upon joining the network, thus excluding the contribution to that capital made by the other suspect network members the individual subsequently has relations with. To obtain this value we consider a set of attributes that measure the acquisition of knowledge and skills for some type of group crime. This set is given by

  $$Pcg_i = r(S_i) \;\; \forall i \in N \tag{2}$$

  in which S_i is the set of relevant attributes of individual i and $r(.)$ is a function that transforms this set into a propensity value.

  The use of attributes has been used in the context of a social network, but with the aim to inferring the existence of links in a cover social network using a bayesian approach Rhodes and Keefe (2007).
- $HCcg_i$ is the contribution to individual i's criminal group human capital of the individuals he or she has relations with.

  The contribution of others to an individual's human capital is a function of his or her social capital represented by the set of links to those he or she has relations with. The function expressing this contribution is defined using an ego network structure. In general, the ego network of an individual i is the network formed with i at the center, known as *Ego*; the set of individuals with whom i is in direct relation, called *Alters*; and the relationships between all of these individuals (Borgatti et al., 1998) as shown in Fig. 1.

  In this ego structure, the social capital of i or *Ego* is a function of the number of *Alters* who possess the characteristics desired by i and of the link i maintains with each of them. Assuming that criminal group human capital is the desired characteristic and that this capital is transferred to i from the

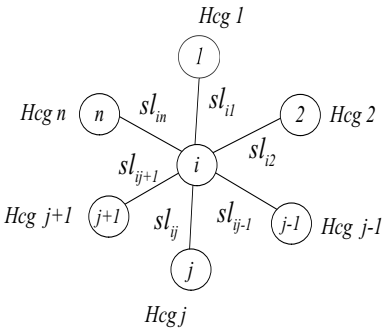


Fig. 1 Base ego network of individual i for obtaining HCcg

Alters through these links, the general function representing the contribution to i's criminal group human capital is given by

$$HCcg_i \;=\; f(Hcg, sl)_V \;\; \forall i \in N \tag{3}$$

where f is a function that expresses the value of the criminal group human capital transferred to i from the set of individuals $V \subset N$ with whom i maintains a social link *sl*. It is extremely important that an appropriate form is adopted for this function so that a truly representative value for this transfer is obtained. In the next subsection we present a key relationship for determining this form.

3.3 Social interaction and field theory: a key relationship

An appropriate form for the criminal group human capital contribution function must reflect the fact that the transfer of human capital from one individual to another is a function of the link between them. It is not well established how much human capital can be transferred between individuals through a link, but what is clear is that the amount will depend directly on the link's magnitude. In other words, the greater the link, the greater the amount of human capital that can be transferred (Coleman, 1988).

The form finally chosen for the function was inspired by an analogy between social interaction among individuals and the interaction between particles described by field theory (Landau and Lifshits, 1975). According to this theory, every particle generates a field around itself that exerts a force or influence on every other particle located within its radius of action. Borrowing this idea, we assume that an individual's human capital exerts an influence on other individuals within his or her radius of action. A similar analogy underlies certain evaluation algorithms, which measure the topological potential of a network node based on structural aspects of the network (Nan et al., 2007; Wang et al., 2010; Le and Hewei, 2010; Cheng et al., 2011).

The potential of a particle is expressed by the following Gaussian function:

$$\varphi(N_i) \;=\; \sum_{j=1}^{n} m_j e^{-\left(\frac{d_{ij}}{\sigma}\right)^2} \tag{4}$$

where n is the number of particles, m_j represents the mass of a particle j, d_{ij} is the topological distance between particles i and j, and σ is a parameter that controls the particles' region of influence.

Using a Gaussian function such as in Eq.(4) to represent the transfer of human capital allows us to capture the fact that the interaction between individuals has local characteristics and that human capital's influence decays as the link weakens.

Before applying Eq.(4) we make the following assumption:

Assumption 1 *Assume that in transferring human capital between individuals, what is transferred from individual j to individual i is only what is useful to i.*

This assumption states that if at instant t individual i contributes part of his or her human capital to individual j and at the following instant $t+1$ individual j contributes part of his or her human capital

to individual i, the human capital useful to i is a quantity that takes no account of the quantity that was transferred by i to j at the preceding instant t.

Using Eq.(4) and Assumption 1, we propose the following function to represent an individual's contribution of criminal group human capital:

$$HCcg_i = \sum_{j \in V} (Hcg_j - Hcg_i) e^{-(\frac{sl_{ij}}{\sigma})^2}, \quad \forall i \in N \tag{5}$$

where σ governs the region of influence over which a network member can contribute criminal group human capital to another member. Given the properties of the Gaussian function, the region of influence of each node is approximately $3\sigma/\sqrt{2}$. Thus, when $0 < \sigma < \sqrt{2}/3$ there is no interaction between nodes, and when $\sqrt{2}/3 < \sigma < 2\sqrt{2}/3$ each node is influenced only by nodes it is directly linked to (Nan et al., 2007; Wang et al., 2010). Since we have adopted the ego network of Fig.1 as our basic structure, the individuals directly linked to i attain their maximum influence when parameter σ is $2\sqrt{2/3}$, which will therefore be the value the parameter will be set to.

It should be noted that when individual j transfers human capital to individual i, j is transferring part of the human capital previously contributed to him or her by individuals he or she is directly linked to, but who are not necessarily directly linked to i. Thus, the influence of individuals directly linked to i includes the influence of individuals not directly linked to i. In other words, the measurement of an individual's contribution to criminal group human capital indirectly takes into account the influence of all the individuals in the network.

3.4 Social Network Criminal Suspect Evaluator: The New Node Evaluator

We now formally introduce our proposed new node importance evaluator, which we will call the Social Network Criminal Suspect Evaluator (SNCSE). Substituting Eq.(5) into Eq.(1), we get

$$Hcg_i = r(S_i) + \sum_{j \in V} (Hcg_j - Hcg_i) e^{-(\frac{sl_{ij}}{\sigma})^2} \quad \forall i \in N \tag{6}$$

Letting

$$e_{ij} = e^{-(\frac{sl_{ij}}{\sigma})^2} \tag{7}$$

and

$$\alpha = \frac{1}{1 + \sum_{j \in V} e_{ij}} \quad \forall i \in N \tag{8}$$

we solve for Hcg_i to get the general form of the *SNCSE*:

$$Hcg_i = \alpha r(S_i) + \alpha \sum_{j \in V} Hcg_j e_{ij} \forall i \in N \tag{9}$$

This can be written in matrix form as

$$(I - E)Hcg = R \tag{10}$$

where I is the identity matrix, E is a square matrix with elements $\alpha e_{ij} \geq 0$ and $\alpha e_{ii} = 0$ $\forall i, j \in N$, Hcg is the column vector of elements $Hcg_i \forall i \in N$ and H is the column vector of elements $\alpha r(S_i) \geq 0$. To determine how the evaluator can be applied to a suspect network, observe first of all that it has the following form:

$$x_i = c_i + \sum_{j \in V} a_{ij} x_j \tag{11}$$

This form is a system of linear equations similar to the one used for solving Leontief's input-output model (Leontief, 1986)[1]. In that model, a_{ij} is interpreted as the input of product i per unit of output of

[1] The Leontief input-output model is a quantitative economic technique that represents the interdependencies between different areas of a national economy or different regional economies.

j, x_i as the output of the emphith industry and c_i as the amount of the c_ith product in the bill of goods. The matrix form is

$$(I - A)X = C \tag{12}$$

where X and C are column vectors containing n components and A is the square matrix containing elements a_{ij}.

The input-output model assumes that the a_{ij} values are non-negative, as can be deduced from the definition of a product input and the postulate that each primary production process has only one output. The same is also true of the e_{ij} values in Eq.(9), in which a single type of human capital (i.e., criminal group) is transferred in order to obtain a different level of the same type, and the minimum transferrable amount from one individual to another is zero.

The main formal question in the input-output model is the existence of a static solution to the system of linear equations (Eq (11)) that produces a bill of goods without negative outputs. Such a solution can be found if the corresponding dynamic system of product transfer is stable (Solow, 1952).

In a suspect network, a negative value for criminal group human capital has no meaning as it would imply that upon transferring part of his or her human capital an individual becomes in some sense the opposite of suspect. In reality, of course, such a transfer can only make one individual less suspect than another. To ensure the system of equations in Eq.(9) produces non-negative outputs, the dynamic system that transfers criminal group human capital between individuals must be stable. The level of human capital in this dynamic system is obtained via the following system of difference equations:

$$IHcg_{t+1} - EHcg_t = R \tag{13}$$

for which a stable solution is reached when $Hcg_{(t+1)} = Hcg_t = Hcg$. This solution will represent the maximum criminal group human capital levels that can be attained in a sufficiently large amount of time. A solution for the system in Eq.(13) arrived at through iteration can be expressed as

$$Hcg_t = E^t Hcg_0 + (I + E + E^2 + ... + E^{t-1})R = E^t Hcg_0 + R\sum_{i=0}^{t-1} E^n \tag{14}$$

where criminal group human capital converges to a stable value if and only if the absolute values of the eigenvalues of matrix E are less than 1, that is, $|\lambda| < 1 \ \forall i \in E$ (Solow, 1952), in which case $I+E+E^2+...$ converges to $(I - E)^{-1}$ and $E^t Hcg_0$ to the null matrix.

There are thus two cases:

- if $|\lambda| < 1$ is satisfied for matrix E, then the SNCSE is applied by solving the system of equations given in Eq.(10), whose solution is given by $Hcg = (I - E)^{-1}R$.
- if $|\lambda| < 1$ is not satisfied for matrix E, the difference equation system in Eq.(13) can be solved until a minimum difference in the criminal group human capital value of $\epsilon = \|Hcg_{t+1} - Hcg_t\|$ for two consecutive periods is reached using Algorithm 1.

Algorithm 1 Application of the SNCSE if $|\lambda_i| < 1$ is not satisfied

Require: $Hcg_1 = R, \epsilon$
1: **while** $\delta \geq \epsilon$ **do**
2: $\quad Hcg_{(t+1)} = (I - E)^{-1} Hcg_t$
3: $\quad \delta = \|Hcg_{t+1} - Hcg_t\|$
4: **end while**
5: **return** Hcg_{t+1}

In view of all the above, the proposed methodology for applying the SNCSE to obtain the social network and the propensity to belong to a criminal group is as follows:

- Identify the total set of individuals G to be included in the evaluation.
- Transform the available data for determining the links between individuals $sl_{ij} \, \forall i, j \in N = \{1...m] \subset G$, which must represent their social relations.

- Establish the social network given by a set $N \subset G$ of nodes (individuals) in terms of a set A of arcs (social links).
- Identify the set of attributes that determine an individual's propensity to belong to a criminal group (or to participate in a group crime) and define a representative indicator.
- Obtain the $m \times m$ square matrix E and check whether its eigenvalues λ satisfy the condition that $|\lambda| < 1$.
- If $|\lambda| < 1$ is satisfied for matrix E, apply the SNCSE by solving the linear system of equations $Hcg = (I - E)^{-1}R$ and generate a ranking of the evaluated individuals.
- If $|\lambda| < 1$ is not satisfied for matrix E, apply the SNCSE through Algorithm 1 to generate a ranking of the evaluated individuals.

4 Applications of the Social Network Criminal Suspect Evaluator

In the previous section we introduced the *SNCSE* node evaluator and analyzed the conditions for its application. In this section we present two real-world applications of the evaluator. In the first one, the SNCSE was applied to a set of five social networks of students based on data drawn from an evaluation survey of an American program for youth gang prevention known as Gang Resistance Education and Training (G.R.E.A.T.). In the second application, the network consisted of individuals charged with petty drug offences and was based on data supplied by the Chilean Criminal Investigation Police.

4.1 Application: Evaluation of the Gang Resistance Education and Training (G.R.E.A.T.) Program

The information for this application is contained in a study of the G.R.E.A.T. Program carried out mainly at the University of Nebraska at Omaha in the United States during 1995-1999 (Esbensen, 2003). The study conducted surveys to gather data on the social and gang-related behaviors as well as the personal and family characteristics of students at secondary schools where the program was implemented. We used this information to construct five social networks, evaluate the importance of the network nodes in terms of the value of gang human capital measured by the SNCSE, and test the evaluator's effectiveness.

4.1.1 Construction of networks and application of the SNCSE

The selected data related to the set of students who responded fully every year to the survey questions. A recent paper (Decker et al., 2014) found that such kind of self-nomination was strongly related to embeddedness in gangs in a similar study. In our case it was used to build a social network for each city and determine the propensity of each network member to belong to a criminal group, or in this case, a youth gang. The links between students of a single city were represented by the social distance separating them, defined as the standardized Euclidean distance between the numeric vector of each student's responses and a set of 29 social behavior questions. Among these questions were the following:

- If you had to choose between studying to get a good grade on a test or going out with your friends, which would you do?
- How many of your current friends have been thought of as good students?
- How many of your current friends have purposely damaged or destroyed property that did not belong to them?
- How many of your current friends have stolen or tried to steal a motor vehicle?
- How many of your current friends have attacked someone with a weapon?

The distance derived from the responses was standardized using the following equation:

$$sd_{ij} = \frac{D_{ij}}{maxD_{ij}} \tag{15}$$

where

$$D_{ij} = \sqrt{\sum_{p \in P} (As_{pi} - As_{pj})^2} \ \ \forall i, j \in G \tag{16}$$

In Eq. 15, dividing by $maxD_{ij}$ standardizes the social distance between 0 and 1 while in Eq. (16), As_{pi} is the magnitude of the response of student i to question $p \in P$ and G is the set of students who responded to the survey in each period.

To establish the links between students of a single city, it was assumed that the existence of a link between individuals would be indicated by a similar social behavior (Lauw et al., 2005). If the social distance between any pair of students was less than the mean of the distance between all pairs of students in each application of the survey, the pair's social behavior was deemed to be similar and a link between them was established. In formal terms,

$$sl_{ij} = \begin{cases} 1 \text{ if } sd_{ijq} < \beta_q \quad \forall i, j \in G, \forall q \in Q \\ 0 \text{ if } no \end{cases} \tag{17}$$

where β_q is the mean of the distance between students of a single city q for all application periods.

The social networks thus constructed for the social interactions of the students in each city are shown in Fig.2.

To represent each student's propensity to belong to a gang (Pcg), we used the score generated by a classification model (Liu et al., 2011) chosen on the basis of performance. The classifiers considered were: *Support Vector Machine, Naive Bayes, Neural Net Auto MPL, Logistic Regression and K-Nearest Neighbor* [2]. The attributes included to training the models were those used by Gibson et al. (2009) and Melde and Esbensen (2012) to obtain an index of the propensity to belong to a youth gang and are listed in the Table 1. Note that the social attribute was not included as it was already reflected in the network links.

Attributes Included
Sex
Race
Age
Parental Monitoring
Self Control
School Commitments
Delinquency
Violent Victimization

Table 1 Attributes included to obtain the index of the propensity to belong to a youth gang

The classifier that performed best was K-Nearest Neighbor with a parameter value of $K = 5$. In the training stage, its accuracy was 79.55 % +/- 4.34 % and its precision was 79.74 % while in the validation stage, its maximum accuracy was 82.22 % and its precision was 75 %. Cross-validation was performed in the training stage with 10 validations and a balanced and independent database of 300 students, 150 of which stated they belonged to a youth gang. The testing stage was conducted on an independent database composed of 90 students, 20 of which declared they were youth gang members. After testing with a set of threshold values, maximum accuracy attained in the validation stage was obtained with a threshold of 0.5.

Having obtained each student's social network and propensity to belong to a gang (Pcg), we then obtained the E matrices for each network. Since the eigenvalues were found to be less than 1 in every case, the SNCSE was applied to each network by solving the system of linear equations Eq.(10), as determined by the methodology described in Section 3.4

The results of this evaluation could then be used to focus investigative resources on the prevention of gang formation among students with the highest levels of gang human capital.

4.1.2 SNCSE Effectiveness Test

In broad terms, the effectiveness of an evaluator in the present context is its ability to concentrate the strongest suspects in the highest positions of the ranking generated by its node evaluations. This definition flows directly from the central purpose of the SNCSE, which is to identify a small number of individuals among a large population of possible suspects as the best candidates for police investigation.

[2] These five classifiers are among the ones that come with the software platform Rapid Miner 5.0.

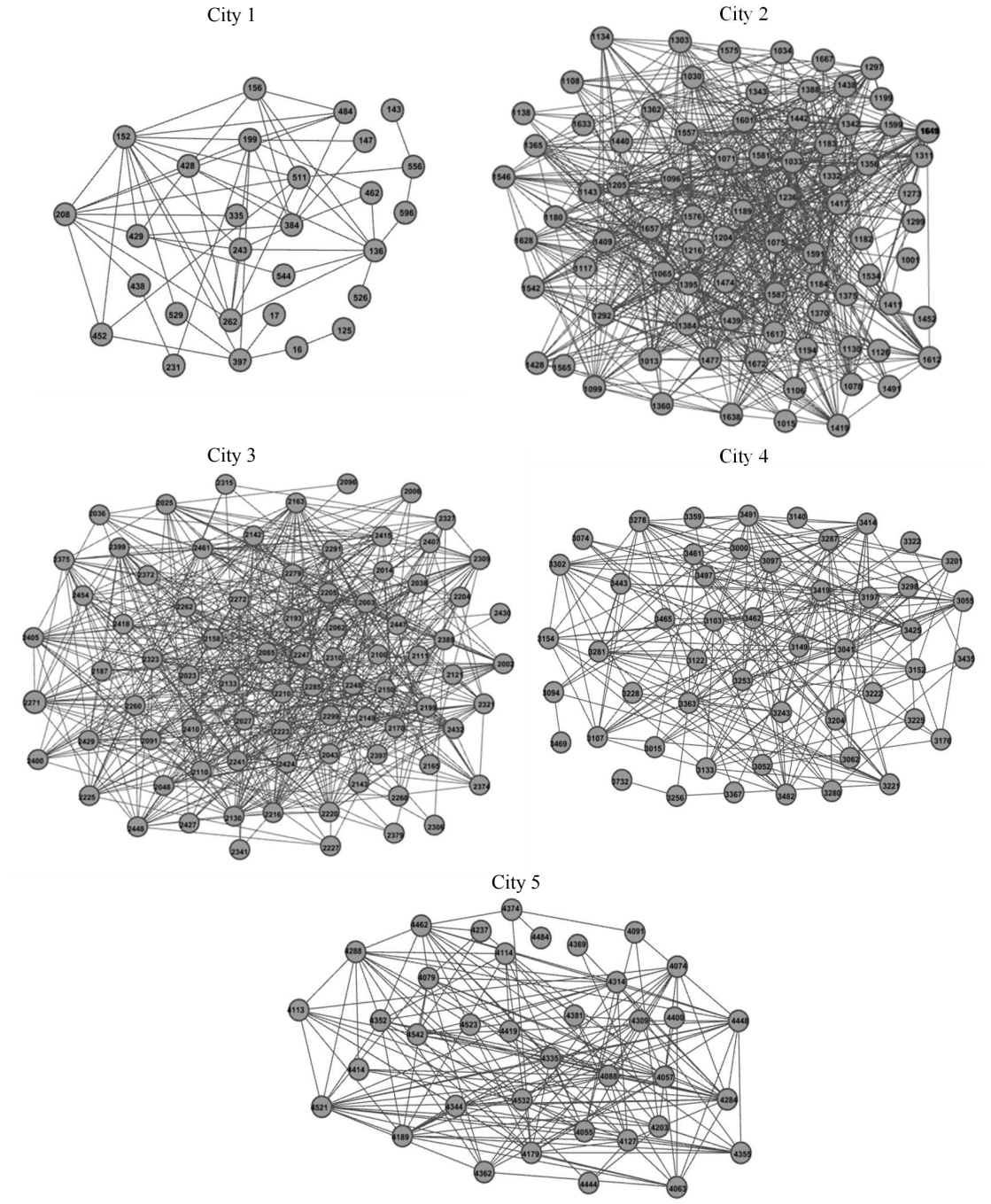


Fig. 2 Student network for each of the five cities.

The first step in developing tests of the effectiveness of the SNCSE using the G.R.E.A.T. program application results just described was to locate a group of students among those ranked by the evaluator who could be independently categorized as strongly suspected of gang membership based on additional information not used in the application, and who would thus serve as a reference for the test.

This group was defined on the basis of its responses to questions in the program surveys clearly pointing to links with gangs or gang-member behavior. Students were considered to be strong suspects

for gang membership if they gave an approving/affirmative or strongly approving/affirmative answer to any of the following five questions in any of the surveys:

1. How do you feel about having friends in gangs?
2. How do you feel about being in a gang yourself?
3. How do you feel about taking part in illegal gang activities?
4. How do you feel about doing whatever the gang leader tells you to do?
5. Have you ever been a gang member?

The number of students thus included in the strong suspect group in each city were 6 of 28 surveyed in City 1, 13 of 84 surveyed in City 2, 22 of 82 surveyed in City 3, 18 of 51 surveyed in City 4 and 10 of 37 surveyed in City 5.

The tests themselves consisted of comparing the number and order of strong suspects in the SNCSE rankings produced by the methodology developed above with the corresponding results generated by a set of SNA centrality measures previously employed in a different youth gang study (Short Jr and Hughes, 2006). These measures were: degree, betweenness, closeness, eigenvector, density, effective size and constraint. The SNCSE rankings were further compared with those obtained by the Hits and Page Rank node evaluation algorithms. The propensity to youth gang membership was also tested as an evaluator in order to separate out the particular significance of including network links in the SNCSE.

The actual comparisons consisted in graphing on lift charts the percentage of total strong suspects the different evaluators included in their respective rankings of all the students, with strong suspect percentages on the vertical axis and the cumulative percentage of all students ordered by ranking from top to bottom along the horizontal axis. A chart for each city showing the results of the SNCSE and the various other evaluators is displayed in Fig.3.

As can be seen, the SNCSE was able to include strong suspects "earlier" than the other evaluators, that is, concentrate more of the strong suspects in higher positions in the rankings. This demonstrates the proposed evaluator's greater effectiveness for police work in that investigative efforts focussing on a relatively small proportion of individuals highly ranked by the SNCSE would include a greater percentage of the strong suspects than the same proportion highly ranked by the other evaluators. This superior performance was due to the incorporation of the students' gang membership propensities.

The lift chart comparisons of the SNCSE with the propensity to gang membership (Pcg) as an evaluator (the latter included, as noted, to separate out the effect of propensity from network linkages) found that the two evaluators behaved similarly on the rankings of all students but that the SNCSE behaved as well or better for rankings of no more than the top 20 %. To bring out more clearly their respective performances, a finer test was designed that compared how the strong suspects were ordered within a specified ranking percentage. Given that good results can be expected with a relatively low proportion of the total, the top-ranked 20 % of students was chosen. At this level, the two evaluators included the same number of strong suspects.

Since the identification of the strong suspect reference group did not differentiate their individual "degrees" of suspect status, the test contrasted the orderings by indicating how many strong suspects the SNCSE ranked in a higher, lower or equal position to the Pcg's rankings of them. The results, set out in Table 2, show that SNCSE performed better on all three indicators. This may be attributed to its incorporation of the students' links, which contributed valuable information that complemented the propensity data.

Results of SNCSE vs Pcg					
Indicator	City1	City2	City 3	City 4	City 5
Higher position	0	3	4	5	4
Lower position	0	2	4	3	1
Equal position	1	2	1	1	0
Strong suspects in top 20%	1	7	9	9	5
Total strong suspects	6	13	22	18	10

Table 2 Effectiveness Test: Relative Positions of Strong Suspects in Top 20 % Of SNCSE And Pcg Rankings, by City

By concentrating on the higher-ranked individuals we are in effect focussing the analysis on those most likely to join a criminal group. It is therefore probable that other individuals who associate directly

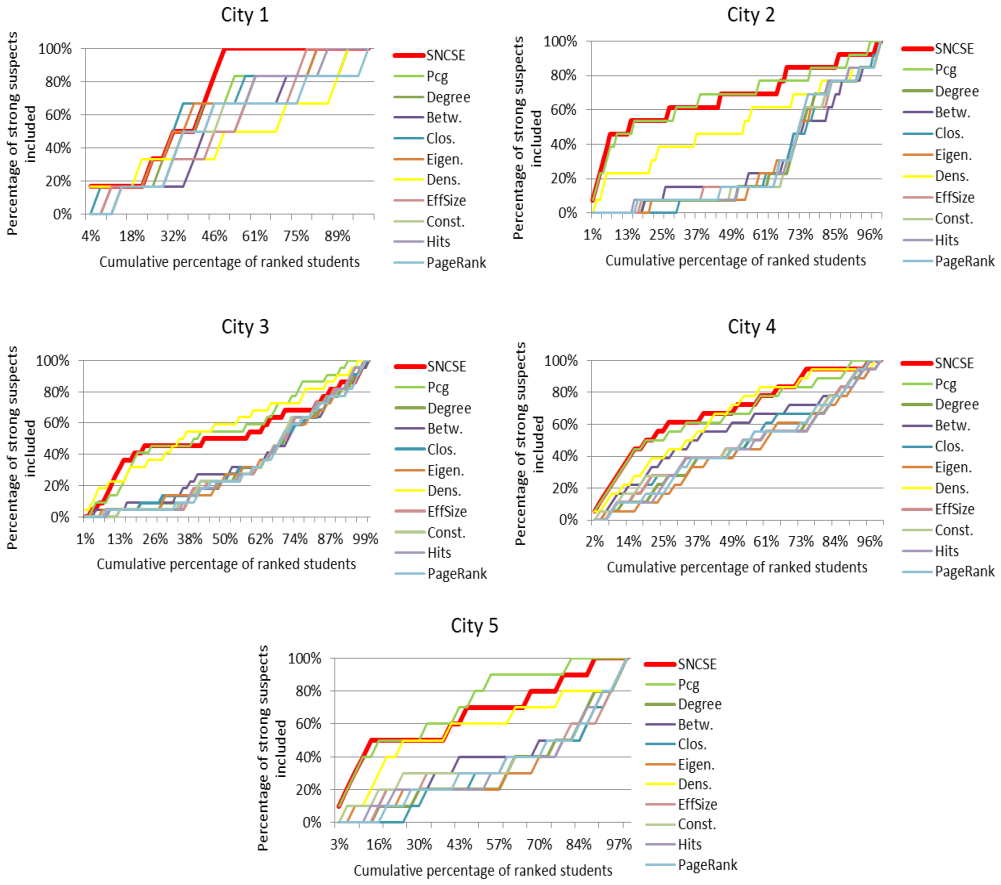


Fig. 3 Effectiveness Test: Inclusion of Strong Suspects in SNCSE and Other Evaluator Rankings, by City

with the higher-ranked ones also belong to such a group. This being so, a subnetwork constructed from the ego networks of the n higher-ranked students could be useful as a base network for broadening investigative work aimed at preventing gang formation. Examples of such subnetworks, built from the ego networks of the three top-ranked students (indicated in red) for each city network, are shown in Fig.4.

4.2 Application: Chilean Criminal Investigation Police

This application utilized information on petty drug offences (possession, use or sale) provided by Chile's Criminal Investigation Police. No effectiveness testing could be conducted in this case due to various institutional restrictions on the type and quantity of data supplied but the application is nevertheless useful as a demonstration that it could be practically implemented in actual police work where such restrictions would obviously not apply.

The data were arranged in two columns, one called IID containing a unique identifier for each individual and the other denoted CID containing a unique identifier for the corresponding court case. An individual could be linked to one or more cases and a case to one or more individuals.

A social network was built using this information to evaluate the importance of individuals in terms of their drug-trafficking group human capital. The network linked individuals who were involved in the same court case and had therefore been prosecuted for the same crime, this being considered evidence that a link existed between them. To establish these linkages, an incidence matrix *IID-CID* was constructed that associated individuals with the court cases in which they were prosecuted. This matrix was then

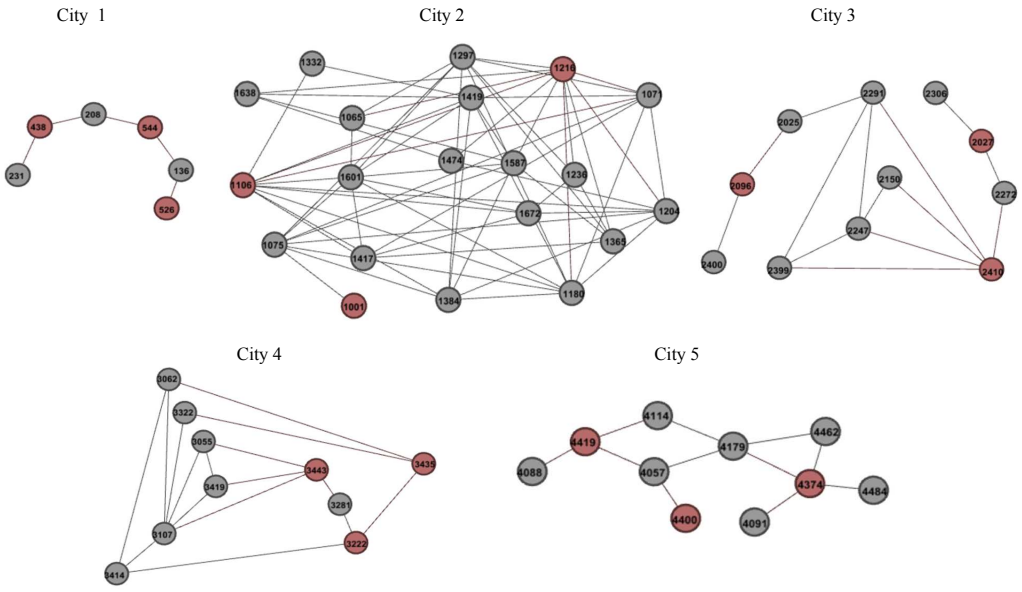


Fig. 4 Subnetworks built from the ego networks of the three top-ranked students in each city.

used to construct the IDD-IDD incidence matrix, which indicated whether any pair of individuals were linked through a shared court case. The existence of such a link was determined by the following function:

$$sl_{ij} = \begin{cases} 1 \text{ if } CID_i = CID_j \quad \forall i, j \in G \\ 0 \text{ if } not \end{cases} \tag{18}$$

On the basis of incidence matrix *IDD-IDD*, a network of 1,132 individuals was obtained as shown in Fig.5. It was composed of a set of disconnected networks containing at least four members each. [3].

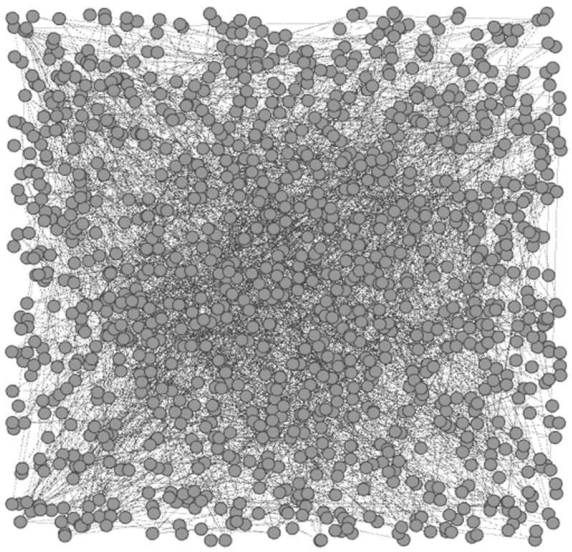


Fig. 5 Petty Drug Offence Network (Possession, Use or Sale)

[3] This minimum number was suggested by the Criminal Investigation Police, who use it as their criterion for defining a criminal group.

The *propensity to membership in a drug-trafficking group (Pcg)* was determined using the following equation:

$$Pcg_i = A_{1i}^{\left(\frac{A_{2i}}{A_{1i}+A_{2i}}\right)} \tag{19}$$

where A_{1i} is the average number of individuals prosecuted in the same court case as individual i, and A_{2i} is the number of cases in which individual i was prosecuted.

The matrix E associated with the network was then obtained, and since the eigenvalues were found to be less than 1 in modulus, the SNCSE was applied by solving the system of linear equations Eq.(10). The ranking thus generated could be used to ensure investigative resources are focussed from the beginning on the most likely suspects, and was made available to the Criminal Investigation Police.

By constructing ego networks based on the top-ranked individuals, investigative work could be extended if necessary on the basis of this original set of strongest suspects. Ego networks based on the 10 top-ranked individuals are shown in Fig. 6.

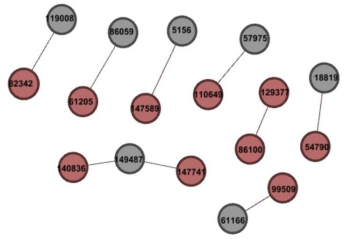


Fig. 6 Individuals with highest human capital levels for membership in a petty drug-trafficking network.

Finally, this application illustrates the high degree of flexibility offered by the SNCSE in that it can be used with a set of disconnected networks.

5 Conclusions and future research

Police investigation of criminal groups requires that large quantities of resources be focussed on key individuals in order to detect the existence of such groups or prevent their formation. By conceptualizing criminal groups as social networks, the identification of these individuals can be approached using node importance evaluators. This study proposed a novel evaluator called the Social Network Criminal Suspect Evaluator (SNCSE), which in addition to network links, incorporates the propensity of each network member to belong to a criminal group.

The SNCSE was tested in two real-world applications that showed it was more effective than traditional evaluators. In the first application, which used data from a youth gang prevention program, the various evaluators were tested by having them rank students known from separate information to be strongly suspected of gang membership. The proposed evaluator demonstrated its superior performance in two particular aspects:

- SNCSE included a greater number of strong suspects among the top-ranked students, and positioned them higher within that ranking, than did the traditional evaluators, none of which incorporated propensity to belong to a criminal group.
- SNCSE included an equal or greater number of strong suspects among the top-ranked students, and positioned them higher within that ranking, than did the propensity to belong to a criminal group alone (i.e., excluding consideration of their network links).

This higher positioning of strong suspects among the top-ranked individuals is an outcome that could have a significant positive impact on the short-term results of a police investigation and thus also on the efficient use of investigative resources.

Also, since the SNCSE incorporates social relations into its evaluations, the top-ranked individuals tend to be those who have the most significant network links. This suggests that those who are

linked to such individuals may also be associated with criminal groups. Subnetworks consisting of the highest-ranked individuals' ego networks could therefore be used as a base for broadening criminal group investigations.

The second application, in which the SNCSE was used with real police data on petty drug offenders, revealed the proposed evaluator's ability to be used with disconnected networks. This empirical conclusion complements the theoretical demonstration that the SNCSE can be applied, depending on the satisfaction of certain conditions, either by solving a system of linear equations or by solving a system of difference equations using an algorithm. Thus, there is a diversity of situations in which the evaluator can be employed, testimony to its flexibility and wide applicability.

As regards future research, the following extensions would be of particular interest:

- The development of a methodology for detecting communities in the subnetworks comprising the ego networks of each of the n highest-ranked individuals. This would strengthen efforts to identify criminal groups.
- The incorporation into the proposed new evaluator of the propensities to various types of group crimes. This would enrich the information used in node evaluation and enable various subnetworks composed of the ego networks of each of the n highest-ranked individuals to be generated in accordance with the importance level assigned to each propensity.
- The application of the proposed new evaluator to tasks such as the detection of fraud rings in social networks.

Acknowledgments

The authors gratefully acknowledge the support of the Santiago-based "Instituto Sistemas Complejos de Ingeniería" (ICM: P-05-004-F, CONICYT: FBO16; www.isci.cl); the Anillo project ACT87 "Quantitative methods in security", managed by Santiago-based CEAMOS (www.ceamos.cl); and the PhD program in Engineering Systems at the Universidad de Chile. The first author also acknowledges a grant provided by CONICYT for his Ph.D. studies in Engineering Systems at Universidad de Chile.

References

Borgatti, S. P., Jones, C., and Everett, M. G. (1998). Network measures of social capital. *Connections*, 21(2):27–36.

Borgatti, S. P., Mehra, A., Brass, D. J., and Labianca, G. (2009). Network analysis in the social sciences. *Science*, 323(5916):892–895.

Chang, J.-J., Lu, H.-C., and Chen, M. (2005). Organized crime or individual crime? endogenous size of a criminal organization and the optimal law enforcement. *Economic Inquiry*, 43(3):661–675.

Cheng, Q., Huang, J., Liu, Z., and Zhu, C. (2011). Evaluation method of effect from network attack considering node multi-property feature. In *Mechatronic Science, Electric Engineering and Computer (MEC), 2011 International Conference on*, pages 1947–1952. IEEE.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American journal of sociology*, 94:S95–S120.

Decker, S. H., Pyrooz, D. C., Sweeten, G., and Moule Jr, R. K. (2014). Validating self-nomination in gang research: Assessing differences in gang embeddedness across non-, current, and former gang members. *Journal of Quantitative Criminology*, pages 1–22.

Ding, L. and Dixon, B. (2008). Using an edge-dual graph and k-connectivity to identify strong connections in social networks. In *Proceedings of the 46th Annual Southeast Regional Conference on XX*, pages 475–480. ACM.

Ding, L., Steil, D., Dixon, B., Parrish, A., and Brown, D. (2011). A relation context oriented approach to identify strong ties in social networks. *Knowledge-Based Systems*, 24(8):1187–1195.

Dnes, A. W. and Garoupa, N. (2010). Behavior, human capital and the formation of gangs. *Kyklos*, 63(4):517–529.

Esbensen, F.-A. (2003). Evaluation of the gang resistance education and training (great) program in the united states, 1995–1999. *Ann Arbor, MI: Inter-University Consortium for Political and Social Research.*

Getoor, L. and Diehl, C. P. (2005). Link mining: a survey. *ACM SIGKDD Explorations Newsletter*, 7(2):3–12.

Gibson, C. L., Miller, J. M., Jennings, W. G., Swatt, M., and Gover, A. (2009). Using propensity score matching to understand the relationship between gang membership and violent victimization: A research note. *Justice Quarterly*, 26(4):625–643.

Landau, L. L. D. and Lifshits, E. E. M. (1975). *The classical theory of fields*, volume 2. Butterworth-Heinemann.

Lauchs, M. A., Keast, R. L., and Le, V. (2012). Social network analysis of terrorist networks: can it add value? *Pakistan Journal of Criminology*, 3(3):21–32.

Lauw, H. W., Lim, E.-P., Pang, H., and Tan, T.-T. (2005). Social network discovery by mining spatio-temporal events. *Computational & Mathematical Organization Theory*, 11(2):97–118.

Le, L. and Hewei, Y. (2010). A new method for evaluating node importance in complex networks based on data field theory. In *Networking and Distributed Computing (ICNDC), 2010 First International Conference on*, pages 133–136. IEEE.

Leontief, W. (1986). *Input output economics*. Oxford University Press.

Liu, Y. Y., Yang, M., Ramsay, M., Li, X. S., and Coid, J. W. (2011). A comparison of logistic regression, classification and regression tree, and neural networks models in predicting violent re-offending. *Journal of Quantitative Criminology*, 27(4):547–573.

McGloin, J. M. and Kirk, D. S. (2010). Social network analysis. In Piquero, A. R. and Weisburd, D., editors, *Handbook of quantitative criminology*, pages 209–224. Springer.

McIllwain, J. S. (1999). Organized crime: A social network approach. *Crime, Law and Social Change*, 32(4):301–323.

Melde, C. and Esbensen, F.-A. (2012). Gangs and violence: Disentangling the impact of gang membership on the level and nature of offending. *Journal of quantitative criminology*, 29:1–24.

Nan, H., Wen-Yan, G., et al. (2007). Evaluate nodes importance in the network using data field theory. In *Convergence Information Technology, 2007. International Conference on*, pages 1225–1234. IEEE.

Qin, J., Xu, J. J., Hu, D., Sageman, M., and Chen, H. (2005). Analyzing terrorist networks: A case study of the global salafi jihad network. In *Intelligence and Security Informatics*, volume 3495, pages 287–304. Springer Berlin Heidelberg.

Rhodes, C. and Keefe, E. (2007). Social network topology: a bayesian approach. *Journal of the operational research society*, 58(12):1605–1611.

Robins, G. (2009). Understanding individual behaviors within covert networks: the interplay of individual qualities, psychological predispositions, and network effects. *Trends in Organized Crime*, 12(2):166–187.

Schultz, T. W. (1961). Investment in human capital. *The American economic review*, 51(1):1–17.

Shaikh, M. A. and Jiaxin, W. (2006). Investigative data mining: Identifying key nodes in terrorist networks. In *Multitopic Conference, 2006. INMIC'06. IEEE*, pages 201–206.

Shaikh, M. A. and Jiaxin, W. (2008). Network structure mining: locating and isolating core members in covert terrorist networks. *WSEAS Transactions on Information Science and Applications*, 5(6):1011–1020.

Shaikh, M. A., Wang, J., Yang, Z., and Song, Y. (2007). Graph structural mining in terrorist networks. In Alhajj, R., Gao, H., Li, J., Li, X., and Zaïane, O., editors, *Advanced Data Mining and Applications*, volume 4632, pages 570–577. Springer Berlin Heidelberg.

Short Jr, J. F. and Hughes, L. A. (2006). *Studying youth gangs*. Rowman Altamira.

Solow, R. (1952). On the structure of linear models. *Econometrica: Journal of the Econometric Society*, 20(1):29–46.

van der Hulst, R. C. (2009). Introduction to social network analysis (sna) as an investigative tool. *Trends in Organized Crime*, 12(2):101–121.

Wang, M. and Pan, W. (2012). A comparative study of network centrality metrics in identifying key classes in software. *Journal of Computational Information Systems*, 8(24):10205–10212.

Wang, T., Han, Y., and Wu, J. (2010). Evaluate nodes importance in directed network using topological potential. In *Information Engineering and Computer Science (ICIECS), 2010 2nd International Conference on*, pages 1–4. IEEE.

Wasserman, S. (1994). *Social network analysis: Methods and applications*, volume 8. Cambridge university press.

Xu, J. and Chen, H. (2003). Untangling criminal networks: A case study. In *Intelligence and Security Informatics*, volume 2665, pages 232–248. Springer.

Xu, J., Marshall, B., Kaza, S., and Chen, H. (2004). Analyzing and visualizing criminal network dynamics: A case study. In *Intelligence and Security Informatics*, pages 359–377. Springer.

Xu, J. J. and Chen, H. (2004). Fighting organized crimes: using shortest-path algorithms to identify associations in criminal networks. *Decision Support Systems*, 38(3):473–487.

Yang, C. C. (2008). Knowledge discovery and information visualization for terrorist social networks. In *Intelligence and security informatics*, volume 135, pages 45–64. Springer Berlin Heidelberg.

Yang, C. C., Liu, N., and Sageman, M. (2006). Analyzing the terrorist social networks with visualization tools. In *Intelligence and security informatics*, volume 3975, pages 331–342. Springer Berlin Heidelberg.

Capítulo 3

A Decision Support Model for Detecting Criminal Network Associations Using Personal Attributes

A Decision Support Model for Detecting Criminal Network Associations Using Personal Attributes

Fredy Troncoso[a,*], Richard Weber[b]

[a]*Departamento de Ingeniería Industrial , Facultad de Ingeniería, Universidad del Bío- Bío, Concepción, Chile.*
[b]*Departamento de Ingeniería Industrial, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Santiago, Chile.*

Abstract

Understanding criminal groups as social networks has led to the design of powerful systems for decision support in police investigative work. Tools using the methods of social network analysis have proven particularly effective in the identification of associations between individuals whose relationships are not otherwise evident. This identification is typically based on the links between the individuals, taking no account of other relevant information such as individual attributes. The present study proposes a new model for identifying criminal associations that incorporates this type of data. Built around what is denoted a linear association model, this approach identifies the principal association between two individuals by maximizing the utility function of one of them. The model is tested on a case involving youth gangs by comparing its performance to that of an existing algorithm for identifying associations. The results demonstrate the proposed model's effectiveness as well as its flexibility in generating different association alternatives, a particularly useful feature that contributes to a more efficient use of police resources.

Keywords:
Crime Analytics, Social Networks, Association, Rational Choice, Criminal Propensity

1. Introduction

Police investigations must often mobilize large quantities of human and technical resources to track down the person or persons responsible for a crime [1]. As criminal behavior becomes more complex, such efforts demand the application of ever greater levels of knowledge, technology, experience and time. Investigations typically begin with a set of suspects, and if the set is large then so will be the number of investigative alternatives to be pursued. Each of these alternatives will require different types and amounts of resources, with no guarantee that any of them will produce useful results. Given that resources are always scarce, there is an obvious need for systems that support decision-making in the identification of the alternatives most likely to produce satisfactory results and the efficient employment of the resources available.

In criminal group investigation, an approach that has proven its effectiveness is to consider such groups as social networks amenable to existing methods of social network analysis [2]. Thus, criminal groups are represented as networks in which the nodes are individuals and the arcs are the links between them. A particular network is specified by identifying these links using data in the initial information base created when an investigation is launched.

To the best of our knowledge, in the general network analysis methods developed to date for crime analysis, the principal sources of information are those that provide the necessary data for determining the links between individuals of interest. However, the aforementioned initial information base used in criminal group investigation will typically contain a set of personal data on the suspects that are not yet employed by

*Corresponding author.
Email addresses: ftroncos@ubiobio.cl (Fredy Troncoso), rweber@dii.uchile.cl (Richard Weber)

these methods and which have the potential of enriching and complementing the analysis and thus improving investigative work.

Many network analysis methods identifying associations between individuals are then used to uncover relationships between individuals whose links are not otherwise evident. The discovery of these relationships can be very useful in the identification of criminal groups. Proceeding along similar lines, the present study develops a new method of associating individuals that takes into account not only the links between them but also their personal information. This information is summarized and represented in terms of a value indicating each individual's propensity to commit certain types of group crimes. The proposed method offers greater support for police analysis of criminal groups while adding flexibility to investigative efforts due to its ability to generate a set of alternatives by simply varying a certain key parameter. This flexibility potentially translates into a more efficient use of investigative resources.

The rest of this article is set out in four sections. Section 2 provides some background on the application of social network analysis to criminal groups, reviewing the traditional tools used in this motivating a new model that identifies criminal associations incorporating personal information on network members. Section 3 then introduces the theory and assumptions behind a proposed new model which uses integer linear programming to determine the best criminal association. Section 4 describes a particular application of the proposed model to a real case and tests its effectiveness. Finally, Section 5 presents our conclusions and some suggestions for further development.

2. Criminal Groups and Social Network Analysis

A social network may be defined as *the relational structure of a group or larger social system, consisting of the pattern of relationships among a collection of actors* [3]. Representing a social system as a social network is a reflection of one of the most fruitful ideas in the social sciences, which is that individuals are embedded in webs of social relations and interactions [4].

A criminal group can be understood as a social network in which the nodes represent criminals and the arcs are the links between them. These links act as channels for the transfer or flow of material and/or non-material resources [5]. On this view, criminal group members and their actions are seen as units that are interdependent rather than autonomous.

The social network for any given criminal group is not explicit therefore a representation of it is a fundamental aspect of the social network approach in police investigation. A key element of this task is the definition of a representative link between the various suspects under consideration, which is achieved by analyzing the data collected for the purpose. This information may have been extracted from a range of media such as databases, written records such as suspect statements, bank account records, visual recordings, electronic mails, photographs and mobile phone calls [6]. The collection of these data and their transformation into links is traditionally known as link analysis [7] [8] [9] and tends to be very labor-intensive and time-consuming [1]. Indeed, this activity is considered one of the main problems in social network spatio-temporal data mining [10]. The magnitude of the link between individuals is commonly represented by a value between 0 and 1, where 0 represents no relationship between two individuals and 1 the highest relationship. Various techniques are available for specifying a representative link between two individuals but the four principal ones are the following [11]:

- Self-Report: establishes a link based on the reports of the individual actors.
- Communication: establishes a link based on the communication or transfer of resources between individuals. Evidence of communication thus indicates association.
- Similarity: establishes a link on the basis that friends tend to be similar in their social conduct. Similar behavior among individuals may therefore indicate association.
- Co-occurrence: establishes a link on the notion that if references to individuals occur together more often than mere chance would explain, the individuals may be related.

A social network specified through the process of establishing links will represent the corresponding social system as it was in the period the information on the set of individuals in the network was recorded. To ensure that network analysis is effective, therefore, a network must be updated as new information on the individual members becomes available [12]. This dynamic will be reflected in changes in the network's membership and its social relations.

Once a network has been obtained for a police investigation, the analysis focuses on identifying criminal structures, key individuals and other important members based on the links that have been established. Two classes of techniques for extracting information from networks are node evaluation and identification of associations. Among the node evaluators traditionally used [5] are centrality measures taken from *Social Network Analysis (SNA)* and node evaluation algorithms. The most common centrality measures [13, 14] are degree, closeness, betweenness, density and eigenvector, while algorithms include Page Rank [15] and Hits [16], both frequently employed, and the recently developed Topological Potential [17].

As for the identification of associations, this reveals relationships between suspects that are not immediately identifiable but whose specification may be essential to obtaining good results. A method proposed to identifying associations, is the modified shortest-path algorithm [18]. In this method the association between two individuals in a network is represented by the shortest path with the highest sum of the magnitude of the links between individuals. This approach is based on the idea that when considering the magnitude of the link as a value between 0 and 1 this can be considered also as a probability measure expressing the likelihood that two individuals are related, assuming these links are independent events. Thus, if two nodes are not directly connected but are associated through a path consisting of intermediate nodes and links, the probability of this association is the product of the probabilities of these links. The best association is the one that is the most probable.

To find this most probable association between two individuals using a shortest-path algorithm, the following logarithmic transformation to the magnitude of the network links has been proposed [18]:

$$l = -ln\, w \quad 0 < w \leq 1 \tag{1}$$

where w represents the probability that two individuals are related.

This transformation generates a new graph of the network in which the nodes remain the same. The new graph has the following important properties: the shortest path (determined by the sum of probability values) between a pair of nodes i and j generates the path with the highest value for the product of the probabilities and therefore the highest probability of all possible paths between these nodes. The respective proof can be found in AppendixA.

In a similar vein, another method suggests the measuring the strength of the association between two individuals through transformation of the original network into an edge-dual network in which each unique relation between two nodes has been replaced by a so-called relation node [19] [20]. The strength of these relations is then calculated using the concept of k-connectivity, an indicator for measuring a network's cohesion. A network is said to be k-connected if k is the minimum number of nodes whose removal will divide the graph into two or more pieces. More specifically, the authors apply a local approach to node connectivity. Given two nodes i and j, their association strength is determined by the minimum number k of relation nodes that must be removed to disconnect i and j. This implies that the greater the value of k, the greater the strength of the two nodes' association. The connectivity problem is solved using a modification of the traditional maximum flow algorithm. Whereas the traditional version calculates maximum node connectivity from a source node to a destination node, the modification calculates it as the number of relation nodes that must be removed to disconnect the source node from a destination node. The techniques to establish association in a criminal network are focused mainly on relations among individuals, that is, the magnitude of the links between nodes of the network. The approach we propose in the next section goes beyond existing models for association detection, because it incorporates individual information, included in the nodes, to the analysis of criminal groups.

3. A new association model including individual atributes

In this section we develop a new model that incorporates individual attributes to search for associations between two nodes of a network. In what follows, Subsection 3.1 provides a general view of our new approach. Subsection 3.2 presents a new criterion to identify associations incorporating information on links as well as on individual attributes at a node level. Based on this new criterion, Subsection 3.3 formally states the proposed Integer Linear Program (ILP) that determines the subset of individuals E and the set of links A_E which form the best association between two individuals.

3.1. The new association approach - A general view

Since a social network may associate two members through various different paths, a method is needed that will identify the path which represents the "best" association between these two members. In general terms, it will be the path that most strongly unites two network members. The new method we propose for identifying the best association in a network combines a measurement of links with a measurement of the nodes themselves constructed from the respective individual attributes.

The individual attributes consist of data that reflect an individual's acquisition of knowledge and skills needed for committing certain types of crime. These attributes can then be represented in terms of an indicator that expresses the individual's propensity to commit a crime or criminal propensity. In the present context, criminal propensity will mean the propensity to belong to any criminal group (Pcg), and this may be expressed as:

$$Pcg_i = r(S_i) \tag{2}$$

where S_i is the set of relevant attributes of individual i and r is some function, elected under a certain context, that transforms this set into a propensity value. Thus, the proposed model identifies the best association using two measures, the magnitude of links and the nodes criminal propensity.

The use of attributes has been used in the context of a social network, but with the aim to inferring the existence of links in a cover social network using a bayesian approach [21].

In the present context, the search for the best association can be interpreted as the process of forming a criminal group in which a decision-maker or planner plans a group crime and chooses the other individuals who will participate in it based on their criminal abilities and their trustworthy. The criminal abilities are represented by the criminal propensity and the trustworthiness between two individuals is represented by the magnitude of their link.

To find the best association between the planner s and the receiver d we assume that the planner acts rationally. This implies that the individuals he/she chooses to be members of the group must have enough criminal ability and provide sufficient trustworthiness to the group to ensure two main goals:

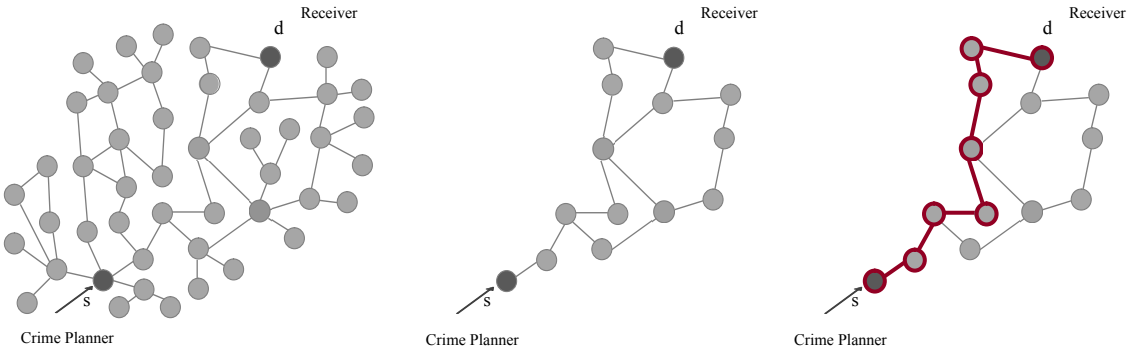
- The crime is carried out and
- The planners utility is maximized

To formulate the planner's rational choice of the best criminal group, we propose an integer linear programming model (ILP) that determines the group which maximizes the planner's utility function subject to a budget constraint plus additional conditions to achieve a path between s and d. The planners utility function thus combines two criteria to determine the incorporation of individuals into an association, high overall criminal ability and high total trustworthiness to the group.

To see this more clearly, consider a given pair of individuals assumed to be part of a criminal group whose best association we want to find. One of the two individuals is assigned the crime planner's role and is denoted s. The planner chooses the group members, decides how the illicit proceeds of the crime (hereafter simply "the proceeds") will be divided up and assumes the risks involved. The second individual plays the role of the "receiver" in the carrying out of the crime and is denoted d. Let us say, for example, that the planned crime is the theft of a car. The criminal group in this case would consist of a chain of individuals in which the planner s plans the theft and then recruits the members who will steal and hide the car, those

who will disassemble it for parts or alter its identification number, and individual d who acts as the receiver or "fence", selling the parts or the entire vehicle.

The identification of this association in our example is depicted visually in Fig.1, where Fig.1a represents the criminal network of the group containing s and d as defined by police analysis, Fig.1b shows all of the possible associations in the network between s and d, and Fig.1c highlights for purposes of illustration one of these possible associations as the best one, that will be the best criminal group that the planner can put together for the crime to be carried out and to maximize his/her utility.



a) Suspect Network

b) Set of possible criminal associations between s and d

c) Best criminal association between s and d

Figure 1: Identifying criminal associations.

3.2. The planner's utility function

In this subsection we formulate the planner's utility function as functions of link magnitude and criminal propensity (Pcg). The formulation of each of its components and the assumptions necessary to formulate is also described.

3.2.1. General form of the utility function

Various definitions of the utility of a criminal organization have been proposed by Becker (1968) [22], Garoupa (2000) [23], Kugler et al. (2005) [24] and Dnes and Garoupa (2010) [25]. Based on these definitions we propose the following conceptual version of a criminal group utility function:

$$U = I\,p - C\,q - W \tag{3}$$

where $I\,p$ is the expected income from a planned illegal activity by the organization, I being the proceeds and p being the probability of carrying out the activity; $C\,q$ is the expected cost to the organization of bribes to buy the members' silence and prevent information on the planned crime from leaking out, C being the maximum bribe the organization is willing to pay and q the probability of a leak; and W is the payout from the proceeds to the members of the organization.

Adapting Eq. (3) to the context of a suspect network $G(N, A)$, where N is the set of individual suspects or nodes and A is the set of arcs joining the individuals, we express the general form of the planner's utility function in terms of three components just described above: Income, Cost of bribes and Payout to group members. Thus,

$$U = I\,Pr(\sum_{i \in E} Pcg_i) - C\,Pr(\sum_{(i,j) \in A_E} (1 - sl_{ij})) - \sum_{i \in E} W_i(Pcg_i) \tag{4}$$

where Pcg_i is the propensity of each individual $i \in N$ to belong to a criminal group , sl_{ij} is the magnitude of the link $(i,j) \in A$ between two individuals i and j. $E \subseteq N$ is the set of individuals forming a particular association and $A_E := \{(i,j) \in A; i,j \in E\}$.

Below we will define and justify an expression for each of the three components of this general form utility function, incorporating our two measures Pcg_i and sl_{ij}.

3.2.2. Expected income

In Eq.(4), expected income is represented by $I\,Pr(\sum_{i\in E} Pcg_i)$, where $Pr(\sum_{i\in E} Pcg_i)$ is the probability of carrying out a planned crime, the equivalent of p in Eq.(3). This probability depends directly on the criminal propensity $\sum_{i\in E} Pcg_i$. The greater the propensity, the greater the criminal ability of the association and the greater, in turn, the probability that a planned crime will be carried out. We now model the planner's income function using the following assumption:

Assumption 1. *We assume that the probability of carrying out the crime is equal to the proportion of criminal propensity present in the select group.*

Based on this assumption we model the planner's income function as:

$$I\,Pr(\sum_{i\in E} Pcg_i) = I\frac{\sum_{i\in E} Pcg_i}{\sum_{i\in N} Pcg_i} \tag{5}$$

In this equation, the fractional term is an expression of the probability $Pr(\sum_{i\in E} Pcg_i)$ of carrying out the planned crime, which is based on the following assumption:

To use Eq.(5) as the planner's income function, we suppose that when the criminal abilities of all available individuals in the network are incorporated into the criminal group, the planned crime is going to be carrying out and the probability $Pr(\sum_{i\in E} Pcg_i)$ will be 1 for $E = N$. This probability will reach its minimum when none of them are incorporated.

3.2.3. Expected cost of bribes to prevent leaks

In Eq.(4), the expected bribe cost to prevent leaks is represented by $C\,Pr(\sum_{(i,j)\in A_E}(1 - sl_{ij}))$, where $Pr(\sum_{(i,j)\in A_E}(1 - sl_{ij}))$ is the probability of a leak given the links A_E. This probability corresponds to q from Eq.(3) adapted to our case. The motivation for basing this probability on the link magnitude term sl_{ij} is that the latter is a measure of social closeness between i and j so that $1-sl_{ij}$ is naturally just the opposite, that is, a measure of social "farness" and therefore also of mistrust (i.e., the absence of trust) between them. Understood in this manner, total mistrust in an association is given by $\sum_{(i,j)\in A_E}(1-sl_{ij})$. Thus, we assume the probability of a leak for a given association depends directly on the total mistrust in that association. The greater this probability, the greater the proportion of the maximum bribe C the association will have to pay. We now model the planner's bribe cost using the following assumption:

Assumption 2. *The planner is willing to assign a maximum amount for paying bribes C defined as a percentage γ of the proceeds I. The maximum amount will be paid when the maximum probability of leak occurs, that is in the association of greatest mistrust between the planner s and the receiver d.*

To see this assumption more clearly, suppose the planner needs to incorporate all of the available criminal ability – in effect – to the entire network. As we assume that the planner acts rationally, he/she chooses the path that provides the least mistrust to the group because he/she will always prefer to obtain the minimum probability of a leak. If the planner makes a bad planification and chooses a different path, this will increase the mistrust in the group and the probability that a leak occurs. Under this context the worst planification is made when the planner chooses the path of greatest mistrust, so the probability that a leak occurs will be the maximal.

Based on this assumption we model the following as the planner's bribe cost function:

$$C\,Pr(\sum_{(i,j)\in A_E}(1-sl_{ij})) = I\gamma\frac{\sum_{(i,j)\in A_E}(1-sl_{ij})}{M_{Path}} \tag{6}$$

where M_{Path} is the total social distance of the path of greatest mistrust or maximum social distance, formally defined in Eq.(7) [26].

$$M_{Path} = \sum_{(i,j)\in A}(1-sl_{ij})Z_{ij} \tag{7}$$

where Z_{ij} is a variable that is 1 when the arc $(i,j) \in A$ is in path of greatest mistrust or maximum social distance and 0 otherwise. This variable is obtained by solving the following maximization problem:

$$Max \sum_{(i,j)\in A}(1-sl_{ij})Z_{ij} \tag{8}$$

$$\sum_{i\in N:i\neq s} Z_{ji} = 1 \;\; \forall\, j \in N\backslash\{d\} \tag{9}$$

$$\sum_{i\in N:i\neq d} Z_{ij} = 1 \;\; \forall\, j \in N\backslash\{s\} \tag{10}$$

$$\sum_{i,j\in L} Z_{ij} = |L| - 1 \;\; \forall\, L \subseteq N\backslash\{s,d\} : |L| \geq 2 \tag{11}$$

$$Z_{ij} \in \{0,1\}\,\forall\,(i,j) \in A \tag{12}$$

Where the constraint of Eq.(9) imposes that each node (except d) must have a link leading from it and the constraint of Eq.(10) requires that each node (except s) must have a link leading to it. Eq.(11) describes a set of constraints that avoids subtours.

3.2.4. Payout to criminal group members

In Eq.(4), the payout from the proceeds to the criminal group members is represented by $\sum_{i\in E} W_i(Pcg_i)$, which corresponds to W in Eq.(3). The term $W_i(Pcg_i)$ represents the payout each group member i receives as a function of his/her criminal ability [27] given by Pcg_i.

We propose the following function for criminal group member payout:

$$\sum_{i\in E} W_i(Pcg_i) = w_p \sum_{i\in E} Pcg_i \tag{13}$$

where we propose w_p as the rate that represents the planner is willing to pay per unit of Pcg. To obtain w_p we assume the following assumption:

Assumption 3. *For any association (group) the planner chooses the utility he or she will expect to receive is at least equal to his or her criminal ability given by Pcg_s.*

This assumption implies that for the worst association possible the group income must be at least as much as the planner could obtain individually. This is expressed by:

$$I - w_p \sum_{i\in N:i\neq s} Pcg_i - I\gamma \geq I\frac{Pcg_s}{\sum_{i\in N} Pcg_i} \tag{14}$$

The left-hand side of Eq.(14) represents the proceeds less the costs incurred by the planner for the path of greatest mistrust and the right-hand side is the planner's minimum expected utility.

The upper bound for the rate the planner is willing to pay per criminal ability unit to group members, is as follows:

$$w_p \leq \frac{I}{\sum_{i \in N: i \neq s} Pcg_i} \left(1 - \gamma - \frac{Pcg_s}{\sum_{i \in N} Pcg_i}\right) \tag{15}$$

If for any reason the planner chooses the path of greatest mistrust, the rates obtained for w_p ensure the planner receives a utility at least commensurate with his or her criminal ability.

Assuming the planner pays the best rate possible w_p per criminal ability unit in order to ensure the chosen individuals join the group, including himself/herself, the rate is given by

$$w_p = \frac{I}{\sum_{i \in N: i \neq s} Pcg_i} \left(1 - \gamma - \frac{Pcg_s}{\sum_{i \in N} Pcg_i}\right) \tag{16}$$

The expressions Eq.(5), (6) and (13) for the three components of the planner's utility function are now substituted into the function's general form given by Eq.(4), yielding the following:

$$U = I \frac{\sum_{i \in E} Pcg_i}{\sum_{i \in N} Pcg_i} - I\gamma \frac{\sum_{(i,j) \in A_E} (1 - sl_{ij})}{M_{Path}} - w_p \sum_{i \in E} Pcg_i \tag{17}$$

3.3. Linear Rational Association Model: The New Association Model

We now present an Integer Linear Program (ILP) model that determines the subset of individuals E and the set of arcs A_E which form the best association between individuals s (planner) and d (receiver). First, however, we must express the planners utility function given by Eq.(17) as a linear objective function, for which we define the following sets of decision variables:

$$X_{ij} = \begin{cases} 1 & \text{if } (i,j) \in A_E \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

$$Y_i = \begin{cases} 1 & \text{if } i \in E \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

where Eq.(18) indicates whether the link between individuals i and j is or is not in the association and Eq.(19) indicates whether individual i is or is not in it.

Using these sets of decision variables, we model the planner's linear utility function as

$$U = I \frac{\sum_{i \in N} Pcg_i Y_i}{\sum_{i \in N} Pcg_i} - I\gamma \frac{\sum_{(i,j) \in A} (1 - sl_{ij}) X_{ij}}{M_{Path}} - w_p \sum_{i \in N} Pcg_i Y_i \tag{20}$$

If we then substitute the expression for w_p derived above as Eq. (16) into Eq. (20) (see AppendixB), we arrive at the final form of the planner's utility function:

$$U = \frac{I\gamma}{\sum_{i \in N: i \neq s} Pcg_i} \sum_{i \in N} Pcg_i Y_i - \frac{I\gamma}{M_{Path}} \sum_{(i,j) \in A} (1 - sl_{ij}) X_{ij} \tag{21}$$

It is this the planner's utility function that we use in our ILP model to seeking the best criminal group, which is the group that exhibits the best association between individuals s (planner) and d (receiver) in a network as described above. Since $I\gamma$ is a positive constant that appears in both terms of the planner's utility function, we can divide by this constant without altering the optimal solution. The complete formulation, denoted the Linear Rational Association Model (LiRAM), is as follows:

$$Max\ U = \frac{\sum_{i \in N} Pcg_i Y_i}{\sum_{i \in N: i \neq s} Pcg_i} - \frac{\sum_{(i,j) \in A} (1 - sl_{ij}) X_{ij}}{M_{Path}} \tag{22}$$

s/t

$$\sum_{j \in N} X_{sj} = 1 \tag{23}$$

$$\sum_{i \in N} X_{id} = 1 \tag{24}$$

$$\sum_{i \in N: i \neq d} X_{ij} = \sum_{k \in N: k \neq s} X_{jk} \; \forall j \in N \backslash \{s, d\} \tag{25}$$

$$Y_s = 1 \tag{26}$$

$$\sum_{i \in N} X_{ij} = Y_j \; \forall j \in N \backslash \{s\} \tag{27}$$

$$w_p \sum_{i \in N} Pcg_i Y_i \leq \varphi I \tag{28}$$

$$\sum_{i,j \in L} X_{ij} = |L| - 1 \; \forall L \subseteq N \backslash \{s, d\} : |L| \geq 2 \tag{29}$$

$$X_{ij} \in \{0, 1\} \forall (i, j) \in A \tag{30}$$

$$Y_i \in \{0, 1\} \forall i \in N \tag{31}$$

The constraints in *LiRAM* stated by (23) through (27) assure that the planner will choose a single path to associate with the receiver. Constraint (28) represents the fact that in making this choice the planner is willing to pay out a share φ of the illicit proceeds to the members of the association, and constraint (29) eliminates solutions with subtours.

The maximization problem faced by the planner comes down to securing for himself/herself as much as possible of the expected income destined for paying bribes (the first term of the utility function Eq.(21)) and paying as little as possible of this amount in bribes to the other group members (the second term of the function). This can be seen more clearly by analyzing two extreme situations as follows:

- If we consider the situation in which the mistrust between the individuals in the network is 0 ($1 - sl_{ij} = 0, \; \forall (i, j) \in A$) and the planner does not need to pay any bribes, the planner's problem then consists in choosing those group members that allow him or her to keep as much as possible of the income destined for paying bribes. In other words, the planner will include in the group as much criminal ability as possible, an amount that is restricted by the share φI he or she is willing to pay out to group members as expressed by constraint (28).

- If we consider only the second term of the objective function, that is, if we exclude the criminal ability Pcg, the planner's problem then consists of choosing those group members situated along the path of least total mistrust. This allow to the planner, pay as little as possible of the income destined to bribes.

In the next section we examine a real-world application of LiRAM taking into account the two technical details just discussed in order to demonstrate the model's effectiveness.

4. Application of the Linear Rational Association Model

In this section we present an application of the proposed Linear Rational Association Model and discuss its results. Subsection 4.1 describes the dataset used to built the network, to obtain the propensity to belong to a criminal group (Pcg), and to prove LiRAMs effectiveness. Subsection 4.2 shows the network construction and the calculation of its links strengths. How to determine an individuals propensity to belong to a criminal group is outlined in Subsection 4.3 After this preparation, we show the application of LiRAM to the respective criminal network and discuss its results in Subsection 4.4.

4.1. Description of dataset

The LiRAM model was applied to a real-world case of youth gang members in the United States. The literature has indicated that gangs are very loosely organized, and more recent works have demonstrated that gangs are social networks rather than structured groups [28]. The necessary information was drawn from an evaluation survey of a program for youth gang prevention called Gang Resistance Education and Training (G.R.E.A.T.) that was conducted under the auspices of the University of Nebraska during 1995-99 [29]. The survey was administered to 3,500 students for six periods at the secondary schools where the program was implemented and information was gathered on their social and gang-related behaviors as well as personal and family characteristics. Data from this survey were used in the present study to construct a social network, to determine each students propensity (Pcg) to belong to a criminal group (here, youth gang), and then to identify strong suspects in order to test LiRAMs effectiveness.

- *Data for network construction and link magnitude determination:* to construct the social network and determine the link magnitude, we used the information of students from a particular city where the survey was administered. Only 30 out of 584 students provided all necessary answers which were finally used for our purpose. These were answers to 29 social behavior questions, some of which are:
  - If you had to choose between studying to get a good grade on a test or going out with your friends, which would you do?
  - How many of your current friends have been thought of as good students?
  - How many of your current friends have purposely damaged or destroyed property that did not belong to them?
  - How many of your current friends have stolen or tried to steal a motor vehicle?
  - How many of your current friends have attacked someone with a weapon?
- *Data to determine propensity:* In order to determine each students propensity (Pcg) to belong to a criminal group (here, youth gang), we considered 75 out of 3,500 students who declared to belong to a gang now. With these cases we trained a model that provides the propensity Pcg for all students of the constructed network.

  We used the following 7 attributes to train the model [30] [31]:

  1. Sex
  2. Race
  3. Age
  4. Self Control
  5. School Commitments
  6. Delinquency
  7. Violent Victimization

Sex was coded as 0 for female and 1 for male. Race was recoded into a series of dummy variables (Black, Hispanic, Indian, and Other) with White serving as the excluded category. Age is used as a continuous variable ranging from 10 years to 14 years. Self control is composed of two attitudinal dimensions: impulsivity and risk seeking. The measure of each dimension consisted of a four-item scale with Likert-type response. School commitment was alignment with positive academic attitudes, using Likert-type responses. Delinquency was measured by various self-reported illegal and delinquent acts, such as e.g. questions about purposely damaged property and violent crime, minor delinquency and drug dealing. Violent Victimization was measured as the total number of times during the previous year each subject was hit by someone trying to hurt them, had a weapon or force used against them in an effort to obtain money or property, or had been attacked by someone with a weapon or by someone trying to seriously hurt or kill them.

- *Data to identify strong suspects*: to prove LiRAMs effectiveness, it was necessary to identify a group of students in the network who could be categorized as strong suspects for gang membership based on information independent of that used by the model. The identification of that group was based on responses by the students to questions in the G.R.E.A.T. program surveys that clearly pointed to links with gangs or gang-members behavior. Students were considered to be strong suspects for gang membership if they had an approving/affirmative or strongly approving/affirmative answer to any of the following five questions in any of the surveys:

  1. How do you feel about having friends in gangs?
  2. How do you feel about being in a gang yourself?
  3. How do you feel about taking part in illegal gang activities?
  4. How do you feel about doing whatever the gang leader tells you to do?
  5. Have you ever been a gang member?

4.2. Construction of the network and determination of link strengths

Based on the data for network construction and link magnitude determination introduced in Section 4.1, we estimated the social distance among the set G of 30 students as the Euclidean distance between the vector of each student's responses to a set P of 29 social behavior questions, as shown in the following equation:

$$D_{ij} = \sqrt{\sum_{p \in P} (As_{pi} - As_{pj})^2} \;\; \forall i, j \in G \tag{32}$$

where As_{pi} is the response of student i to question $p \in P$.

Since similar social behavior between two individuals implies a high strength of their respective link [11], we propose to construct the social network and calculate a links strength according to Eq.(33).

$$sl_{ij} = \begin{cases} 1 - \frac{D_{ij}}{max D_{ij}} & \text{if } D_{ij} < \beta \quad \forall i, j \in G \\ 0 & \text{otherwise} \end{cases} \tag{33}$$

where β is the mean of the distances between the students given by:

$$\beta = \frac{\sum_{i,j \in G: i \neq j} D_{ij}}{(|G| - 1)\,|G|} \tag{34}$$

The social network thus constructed for the social interactions of these students contained a set N of 22 nodes, each one representing a student, and a set A of 54 arcs between them. The network is depicted in Fig.2.

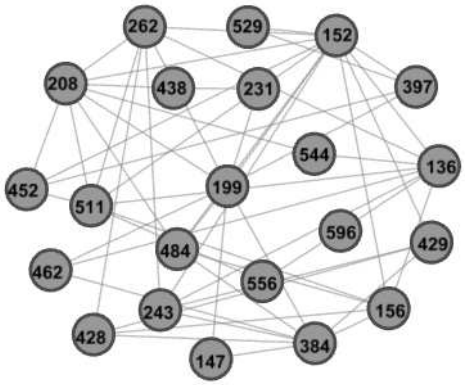


Figure 2: Student network.

4.3. Determination of the propensity to belong to a criminal group (Pcg)

To represent each student's propensity (Pcg) to belong to a criminal group (here, youth gang) we used the score generated by a existing classification model [32]. A classification model is a data mining technique that can be trained to identify an individual's membership to a class using information contained in an attribute vector that defines the individual's particular characteristics. Each classifier generates an output that can be transformed into a score between 0 and 1 and which could be interpreted as probability to belong to the class of interest, here a youth gang.

The classifiers we considered were *Support Vector Machine, Naïve Bayes, Neural Net Auto MPL, Logistic Regression and K-Nearest Neighbor*, using their implementation in RapidMiner 5.0.[33]

To train and validate the classifiers, we considered the 75 students who declared to Belong to a gang now, as introduced in Subsection 4.1.

The classifier that performed best was K-Nearest Neighbor with K=5, and thus was chosen to estimate Pcg. In the training stage, this classifier demonstrated an accuracy of 79.55 % +/- 4.34 % and a precision of 79.74 % while in the validation stage, its maximum accuracy was 82.22 % and its precision was 75 %. We performed 10-fold cross validation on a balanced dataset of 300 students, 150 of which stated they belonged to a youth gang. Testing was conducted on an independent dataset composed of 90 students, 20 of which declared they were youth gang members.

Finally, we applied the trained K-NN classifier to the 22 students of the network to obtain their propensity score Pcg.

4.4. LiRAM application and test of its effectiveness

Once the social network and each student's *Pcg* were obtained, the LiRAM was applied to the network to test its effectiveness in identifying the best associations between pairs of students. The test criterion was the number of strong suspects the model was able to include in the associations it identified compared to the number found by a modified shortest-path algorithm (SPA) such as the one discussed in Section 2 that identifies best associations as those with the highest link weight [18]. The strong suspects were determinated as shown in Subsection 4.1. This approach directly reflects the central purpose of LiRAM, which is to identify a small number of individuals among a large population of possible suspects as the best candidates for police investigation.

A total of 5 students out of the 22 in the network were thus identified as strong suspects, and are marked as pentagon in Fig.3a.

The next step would be to generate associations using LiRAM and the modified SPA between all pairs of students among the remaining 17. Since a planner is expected to have many connections, without loss of generality, we will focus on the four nodes with the greatest number of link; those are nodes 136, 152, 208, and 384.

For each of these 4 students, associations were generated only with other students they were not directly linked to and who were not among the 5 strong suspects. In the case of student 136, whose results we will use as an example and examine in detail, there were 10 such other students and therefore associations were generated between student 136 and each of the 10.

LiRAM was applied to these 10 instances using 5 different values of the maximum payout share φ. The values chosen were 10%, 20%, 30%, 40% and 50%. No higher percentages were used since above 50% the number of individuals included in the association grew too large and accuracy declined.

The results of the associations generated for student 136 are displayed in Table 1. The values shown indicate the number of strong suspects among the total number of students included by the two methods in the associations between 136 and each of the other 10 students as just described above. For the association between student 136 and 397, for example, LiRAM produced a result of 2/5 with $\varphi = 30\%$, meaning the model included 5 students of which 2 (452 and 544) were strong suspects. This result is illustrated in Fig.3b. By contrast, for the same example the modified SPA found no strong suspects among the 3 individuals it included in the association, depicted in Fig.3c.

Number of strong suspects included in associations						
Association	LiRAM					Modified SPA
	$\varphi = 50\%$	$\varphi = 40\%$	$\varphi = 30\%$	$\varphi = 20\%$	$\varphi = 10\%$	
136-147	2/8	1/8	2/5	0/6	1/4	1/4
136-208	1/11	2/10	1/6	1/3	0/3	0/3
136-231	1/11	1/5	0/6	1/5	not feasible	1/4
136-384	2/8	1/8	1/6	0/7	1/3	1/3
136-397	2/9	2/9	2/5	1/5	0/4	0/3
136-428	3/9	3/9	1/7	1/7	1/3	1/3
136-429	2/9	2/9	1/6	0/6	0/4	0/4
136-438	3/8	2/6	1/6	0/4	not feasible	0/4
136-529	2/10	2/9	2/6	1/6	0/5	0/4
136-556	2/10	3/10	2/6	1/6	0/3	0/3
Performance	22.3%	22.8%	22.7%	12.1%	9.2%	11.6%

Table 1: Strong suspects included in associations with student 136: LiRAM vs modified SPA.

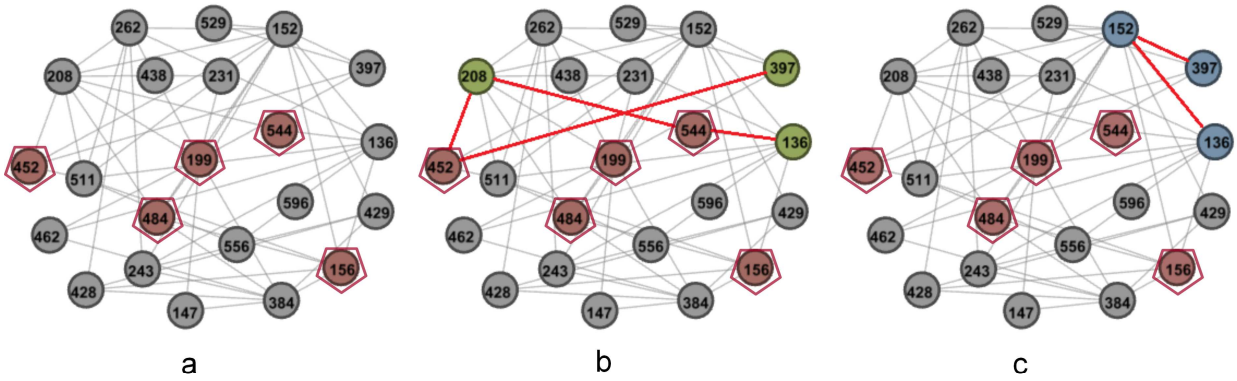


Figure 3: Strong suspects in student network.

The results in Table 1 reflect the alternatives facing a planner trying to choose the best criminal group to represent the best association between two individuals. In the case of association 136-438, for a low payout share $\varphi = 10\%$ and given the rate w_p per unit of criminal propensity in Eq.(28), there are no individuals with whom a criminal group can be formed because the minimum payout per unit of any association is greater

than $0.1I$. But as the value of φ is increased, various groups of different sizes and member composition are identified. For example, when $\varphi = 30\%$ the group has 6 members of which 1 is a strong suspect but when $\varphi = 40\%$, there are 2 strong suspects among the 6 members identified. Thus, by increasing his or her willingness to pay, the planner can replace one member of low criminal ability with one of higher ability who is a strong suspect. By further raising the payout share to $\varphi = 50\%$, the planner will have an 8-member association with 3 of higher ability, that is, strong suspects. In cases where the outcome was "not feasible", there is no solution for the planner's willingness to pay.

LiRAM's ability to generate different best groups by varying parameter φ can lend a significant measure of flexibility to the investigative process. Thus, investigations can be initiated modestly with a small number of suspects by setting a low value of φ that can later be increased as required depending on the results obtained. This characteristic has obvious potential for achieving efficiencies in the assignment of investigative resources.

As for the modified SPA, its results were similar to those of LiRAM at low values of φ. This similarity may be explained by the fact that when the planner's willingness to pay is low, the possibilities of association are more limited and the choice of a group will emphasize lower mistrust levels and greater total link weight. On these criteria, the two methods are likely to behave similarly.

The general performance of the two methods is compared in the bottom row of Table 1 in terms of the average percentage of strong suspects included in the associations between 136 and the other 10 students. For LiRAM, separate percentages are given for each value of φ. These results are plotted as Association 1 in Fig. 4, which also graphs the performance of the associations formed with the other three members of the representative set of non-suspects, students 152, 208 and 384 (respectively, Associations 2, 3 and 4). As can be seen, for Associations 1 through 3, LiRAM always performed better than the SPA at φ values above 20%.

For Association 4, however, although LiRAM averaged about as well as it did for the other three associations, with about 15% of strong suspects included, the SPA's average performance jumped to about 20%. This latter figure was just below the LiRAM result at a φ value of 30%, about the same at 50% but significantly better at all other values.

In such cases, LiRAM's results can be improved considerably by a direct but simple intervention in the association patterns. A close look at Association 4 reveals that when $\varphi = 40\%$, student 147 is frequently chosen as the first node after student 384 in the association path while at φ values of 10%, 20% and 50%, student 429 is frequently selected in this position. To generate a set of alternatives for this case, therefore, LiRAM was reexecuted with the frequent student as just named for each φ case excluded. The outcomes, shown in the last graph in Figure 4, indicate that LiRAM performed better than the SPA for every case except $\varphi = 20\%$. This suggests that intervening to avoid a repetitive association pattern as described provides a simple method of improving an initially unsatisfactory set of investigative results.

5. Conclusions and future research

This paper proposed a model of strengthening and enriching existing methods of social network analysis used to determine whether there exists an association between two individuals suspected of participation in group crimes, and if so, the identity of the association's other members. Under this new approach, network nodes representing criminal group members are assigned values derived from individual attributes reflecting the members' individual criminal abilities and therefore their criminal propensities. The magnitudes of the links joining the members are considered to reflect their individual degrees of trustworthiness. Associations are then the result of a decision process in which a maximizing crime planner chooses the best individuals to associate with on the basis of a personal utility function that specifies the tradeoff between the criminal propensity and trustworthiness measures.

The maximization itself is determined by an integer linear model denoted LiRAM (Linear Rational Association Model) that incorporates the two measures as well as other relevant factors. It also contains a parameter that determines the proportion of the crime proceeds the crime planner pays out to the group members as a function of their individual criminal propensities. For each parameter value, LiRAM determines the best association between any pair of suspect individuals and therefore the other individuals constituting the association.

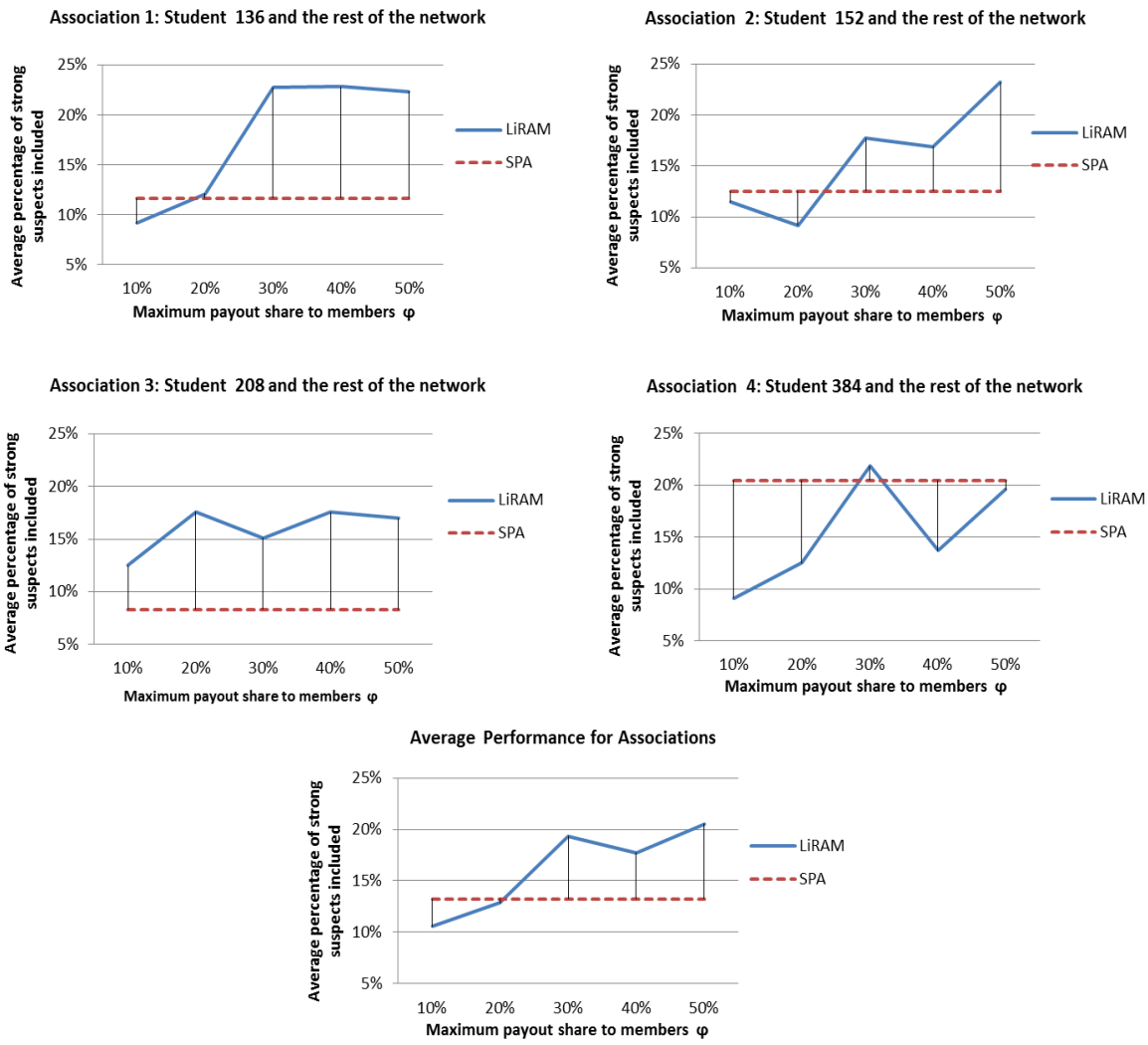


Figure 4: Performance of LiRAM and modified SPA by maximum payout share for representative student associations.

To test the model, it was applied to real data gathered from surveys taken by a youth gang program in the United States. The results were compared to those obtained using a modified shortest-path algorithm. The principal test findings may be summarized as follows:

- LiRAM is effective in finding the best association between individuals in a social network using information on their individual attributes. The model proved able to incorporate individuals independently categorized as strong suspects in the majority of the best associations identified.
- LiRAM lends flexibility to police investigation through its ability to generate different sets of association alternatives simply by varying the above-described parameter governing the payout (φ). Thus, an investigation can begin with a given alternative and explore others as required.
- If the first set of associations generated by LiRAM is less than satisfactory, the results can be improved by intervening in the association to eliminate repetitive patterns and then running the model again.

- The modified shortest-path algorithm, which only generates a single alternative, was often more effective than LiRAM when the latter was used with low parameter values while the proposed model always performed better at higher values. This suggests the two approaches are complementary and thus could be applied in tandem, with LiRAM being the preferred option as the number of desired alternatives for investigation increases.

In a future article we will extend the present work to include the following:

- Assume the magnitude of the link as a probability measure expressing the likelihood that two individuals are related, assuming these links are mutually independent events [18]. This assumption will generate a non linear function of expected cost of bribes that could improve the proposed model effectiveness.
- The incorporation of criminal propensities of different types of complementaries groupal crimes. This allows analyze different associations depending to the importance assigned to different types of crimes, providing differents scenarios to police analysis.
- If also neighbourhood information for suspects would be available, this could be used to further improve the network structure and link information, as has been proposed e.g. in [34].

Acknowledgments

The authors gratefully acknowledge the support of the Santiago-based "Instituto Sistemas Complejos de Ingeniería" (ICM: P-05-004-F, CONICYT: FBO16; www.isci.cl); the Anillo project ACT87 "Quantitative methods in security", managed by Santiago-based CEAMOS (www.ceamos.cl); and the Ph.D. program in engineering systems at the Universidad de Chile. The first author was the recipient of a CONICYT grant number 21120226 to pursue doctoral studies in engineering systems at the Universidad de Chile.

References

[1] J. J. Xu, H. Chen, Crimenet explorer: A framework for criminal network knowledge discovery, ACM Trans. Inf. Syst. 23 (2) (2005) 201–226.

[2] J. Xu, H. Chen, Untangling criminal networks: A case study, in: Intelligence and Security Informatics, Springer, 2003, pp. 232–248.

[3] S. Wasserman, Social network analysis: Methods and applications, Vol. 8, Cambridge university press, 1994.

[4] S. P. Borgatti, A. Mehra, D. J. Brass, G. Labianca, Network analysis in the social sciences, Science 323 (5916) (2009) 892–895.

[5] J. S. McIllwain, Organized crime: A social network approach, Crime, Law and Social Change 32 (4) (1999) 301–323.

[6] E. Ferrara, P. D. Meo, S. Catanese, G. Fiumara, Detecting criminal organizations in mobile phone networks, Expert Systems with Applications 41 (13) (2014) 5733 – 5750.

[7] M. K. Sparrow, The application of network analysis to criminal intelligence: An assessment of the prospects, Social Networks 13 (3) (1991) 251 – 274.

[8] R. Hauck, H. Atabakhsb, P. Ongvasith, H. Gupta, H. Chen, Using coplink to analyze criminal-justice data, Computer 35 (3) (2002) 30–37.

[9] H. Chen, K. Lynch, Automatic construction of networks of concepts characterizing document databases, Systems, Man and Cybernetics, IEEE Transactions on 22 (5) (1992) 885–902.

[10] L. Getoor, C. P. Diehl, Link mining: a survey, ACM SIGKDD Explorations Newsletter 7 (2) (2005) 3–12.

[11] H. W. Lauw, E.-P. Lim, H. Pang, T.-T. Tan, Social network discovery by mining spatio-temporal events, Computational & Mathematical Organization Theory 11 (2) (2005) 97–118.

[12] J. Xu, H. Chen, Criminal network analysis and visualization, Commun. ACM 48 (6) (2005) 100–107.

[13] J. M. McGloin, D. S. Kirk, Social network analysis, in: Handbook of quantitative criminology, Springer, 2010, pp. 209–224.

[14] R. C. van der Hulst, Introduction to social network analysis (sna) as an investigative tool, Trends in Organized Crime 12 (2) (2009) 101–121.

[15] L. Page, S. Brin, R. Motwani, T. Winograd, The pagerank citation ranking: Bringing order to the web., Technical Report 1999-66, Stanford InfoLab, previous number = SIDL-WP-1999-0120 (November 1999).

[16] J. M. Kleinberg, Authoritative sources in a hyperlinked environment, J. ACM 46 (5) (1999) 604–632.

[17] M. Wang, W. Pan, A comparative study of network centrality metrics in identifying key classes in software, Journal of Computational Information Systems 8 (24) (2012) 10205–10212.

[18] J. J. Xu, H. Chen, Fighting organized crimes: using shortest-path algorithms to identify associations in criminal networks, Decision Support Systems 38 (3) (2004) 473–487.

[19] L. Ding, B. Dixon, Using an edge-dual graph and k-connectivity to identify strong connections in social networks, in: Proceedings of the 46th Annual Southeast Regional Conference on XX, ACM-SE 46, ACM, New York, NY, USA, 2008, pp. 475–480.

[20] L. Ding, D. Steil, B. Dixon, A. Parrish, D. Brown, A relation context oriented approach to identify strong ties in social networks, Knowledge-Based Systems 24 (8) (2011) 1187–1195.

[21] C. Rhodes, E. Keefe, Social network topology: a bayesian approach, Journal of the operational research society 58 (12) (2007) 1605–1611.

[22] S. Becker Gary, Crime and punishment: An economic approach, Journal of Political Economy 76 (2) (1968) 169–217.

[23] N. Garoupa, The economics of organized crime and optimal law enforcement, Economic Inquiry 38 (2) (2000) 278–288.

[24] M. Kugler, T. Verdier, Y. Zenou, Organized crime, corruption and punishment, Journal of Public Economics 89 (9) (2005) 1639–1663.

[25] A. W. Dnes, N. Garoupa, Behavior, human capital and the formation of gangs, Kyklos 63 (4) (2010) 517–529.

[26] C. Berge, E. Minieka, Graphs and hypergraphs, Vol. 7, North-Holland Amsterdam, 1973.

[27] J.-J. Chang, H.-C. Lu, M. Chen, Organized crime or individual crime? endogenous size of a criminal organization and the optimal law enforcement, Economic Inquiry 43 (3) (2005) 661–675.

[28] C. Bolden, Gangs and social networks, in: G. Bruinsma, D. Weisburd (Eds.), Encyclopedia of Criminology and Criminal Justice, Springer New York, 2014, pp. 1857–1866.

[29] F.-A. Esbensen, Evaluation of the gang resistance education and training (great) program in the united states, 1995–1999, Summary, URL: www. icpsr. umich. edu 8080.

[30] C. L. Gibson, J. M. Miller, W. G. Jennings, M. Swatt, A. Gover, Using propensity score matching to understand the relationship between gang membership and violent victimization: A research note, Justice Quarterly 26 (4) (2009) 625–643.

[31] C. Melde, F.-A. Esbensen, Gangs and violence: Disentangling the impact of gang membership on the level and nature of offending, Journal of Quantitative Criminology 29 (2) (2013) 143–166.

[32] Y. Y. Liu, M. Yang, M. Ramsay, X. S. Li, J. W. Coid, A comparison of logistic regression, classification and regression tree, and neural networks models in predicting violent re-offending, Journal of Quantitative Criminology 27 (4) (2011) 547–573.

[33] M. Hofmann, R. Klinkenberg, RapidMiner: Data Mining Use Cases and Business Analytics Applications, Chapman & Hall/CRC, 2013.

[34] J. R. Hipp, C. T. Butts, R. Acton, N. N. Nagle, A. Boessen, Extrapolative simulation of neighborhood networks based on population spatial distribution: Do they predict crime?, Social Networks 35 (4) (2013) 614 – 625.

AppendixA.

Below we set out the proof for the network properties resulting from the logarithmic transformation of the arcs. To maintain consistency with the original paper we use the terminology chosen by the authors [18], in particular the term axiom.

Axiom 1. *All links in the new graph are nonnegative numbers.*

- *Proof* : Since $0 < w \leq 1$, thus $ln\, w \leq 0$, which implies $-ln\, w \geq 0$.

Axiom 2. *A lower link weight in the new graph corresponds to a higher link weight in the original network.*

- *Proof* : Let $w_1 < w_2$, then $-ln\, w_1 < -ln\, w_2$ or $ln\, w_1 > ln\, w_2$. Since $ln\, w$ is a monotonic increasing function, it follows that $w_1 > w_2$.

Axiom 3. *The shortest path (using summation of social distance) between a pair of nodes in the new graph generates a path with the maximum link*

- *Proof* : Consider the shortest path, say P, between a pair of nodes A and B. P consists of a set of links with weight $(l_1, l_2, ..., l_p)$, $1 \leq p \leq n$, where n is the total number of nodes in this graph. The total length of this path is $\sum_{i=1}^{p} l_i$. Consider another path between node A and node B, say Q, consisting of another set of links with weight $(l'_1, l'_2, ..., l'_q)$, $1 \leq q \leq n$. The total length is $\sum_{i=1^p} l'_i$.

Therefore, this transformation enables us to use conventional shortest-path algorithms for identifying the strongest associations between a pair of nodes or entities in a suspect network.

AppendixB.

The final form of the planners utility function, given in the main body of the text as Eq.(21), is derived from Eq.(20) as follows:

Recall that Eq.(20)) is

$$U = I\frac{\sum_{i\in N} Pcg_i Y_i}{\sum_{i\in N} Pcg_i} - I\gamma\frac{\sum_{(i,j)\in A}(1-sl_{ij})X_{ij}}{M_{Path}} - w_p\sum_{i\in N} Pcg_i Y_i \tag{B.1}$$

and assuming the planner is willing to pay the highest possible rate per unit of criminal ability to ensure the chosen individuals join the group, w_p is given by

$$w_p = \frac{I}{\sum_{i\in N:i\neq s} Pcg_i}\left(1-\gamma-\frac{Pcg_s}{\sum_{i\in N} Pcg_i}\right) \tag{B.2}$$

Substituting w_p into the utility function, the latter becomes

$$U = I\frac{\sum_{i\in N} Pcg_i Y_i}{\sum_{i\in N} Pcg_i} - \frac{I\gamma}{M_{Path}}\sum_{(i,j)\in A}(1-sl_{ij})X_{ij} - \frac{I}{\sum_{i\in N:i\neq s} Pcg_i}\left(1-\gamma-\frac{Pcg_s}{\sum_{i\in N} Pcg_i}\right)\sum_{i\in N} Pcg_i Y_i \tag{B.3}$$

This can be rewritten as

$$U = I\sum_{i\in N} Pcg_i Y_i\left(\frac{1}{\sum_{i\in N} Pcg_i} - \frac{1}{\sum_{i\in N:i\neq s} Pcg_i} + \frac{\gamma}{\sum_{i\in N:i\neq s} Pcg_i} + \frac{Pcg_s}{\sum_{i\in N} Pcg_i \sum_{i\in N:i\neq s} Pcg_i}\right) - \frac{I\gamma}{M_{Path}}\sum_{(i,j)\in A}(1-sl_{ij})X_{ij} \tag{B.4}$$

Since

$$Pcg_s = \sum_{i\in N} Pcg_i - \sum_{i\in N:i\neq s} Pcg_i \tag{B.5}$$

Substituting this equation into the utility function, we get

$$U = I\sum_{i\in N} Pcg_i Y_i\left(\frac{\sum_{i\in N:i\neq s} Pcg_i - \sum_{i\in N} Pcg_i}{\sum_{i\in N} Pcg_i \sum_{i\in N:i\neq s} Pcg_i} + \frac{\gamma}{\sum_{i\in N:i\neq s} Pcg_i} - \frac{\sum_{i\in N:i\neq s} Pcg_i - \sum_{i\in N} Pcg_i}{\sum_{i\in N} Pcg_i \sum_{i\in N:i\neq s} Pcg_i}\right) - \frac{I\gamma}{M_{Path}}\sum_{(i,j)\in A}(1-sl_{ij})X_{ij} \tag{B.6}$$

Simplifying, the utility function finally becomes

$$U = \frac{I\gamma}{\sum_{i\in N:i\neq s} Pcg_i}\sum_{i\in N} Pcg_i Y_i - \frac{I\gamma}{M_{Path}}\sum_{(i,j)\in A}(1-sl_{ij})X_{ij} \tag{B.7}$$

Bibliografía

[1] Carolyn J Anderson, Stanley Wasserman, and Katherine Faust. Building stochastic blockmodels. *Social Networks*, 14(1–2):137 – 161, 1992. Special Issue on Blockmodels.

[2] Stephen P Borgatti, Ajay Mehra, Daniel J Brass, and Giuseppe Labianca. Network analysis in the social sciences. *science*, 323(5916):892–895, 2009.

[3] Juin-Jen Chang, Huei-Chung Lu, and Mingshen Chen. Organized crime or individual crime? endogenous size of a criminal organization and the optimal law enforcement. *Economic Inquiry*, 43(3):661–675, 2005.

[4] H. Chen and K.J. Lynch. Automatic construction of networks of concepts characterizing document databases. *Systems, Man and Cybernetics, IEEE Transactions on*, 22(5):885–902, Sep 1992.

[5] Li Ding and Brandon Dixon. Using an edge-dual graph and k-connectivity to identify strong connections in social networks. In *Proceedings of the 46th Annual Southeast Regional Conference on XX*, pages 475–480. ACM, 2008.

[6] Li Ding, Dana Steil, Brandon Dixon, Allen Parrish, and David Brown. A relation context oriented approach to identify strong ties in social networks. *Knowledge-Based Systems*, 24(8):1187–1195, 2011.

[7] Antony W Dnes and Nuno Garoupa. Behavior, human capital and the formation of gangs. *Kyklos*, 63(4):517–529, 2010.

[8] Ian Donald and Angela Wilson. Ram raiding: Criminals working in groups. *The social psychology of crime*, pages 189–246, 2000.

[9] James Robert Evans and Edward Minieka. *Optimization algorithms for networks and graphs*, volume 1. CRC Press, 1992.

[10] Linton C. Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215 – 239, 1978–1979.

[11] Nuno Garoupa. The economics of organized crime and optimal law enforcement. *Economic Inquiry*, 38(2):278–288, 2000.

[12] Lise Getoor and Christopher P Diehl. Link mining: a survey. *ACM SIGKDD Explorations Newsletter*, 7(2):3–12, 2005.

[13] Chris L Gibson, J Mitchell Miller, Wesley G Jennings, Marc Swatt, and Angela Gover. Using propensity score matching to understand the relationship between gang membership and violent victimization: A research note. *Justice Quarterly*, 26(4):625–643, 2009.

[14] R.V. Hauck, H. Atabakhsb, P. Ongvasith, H. Gupta, and Hsinchun Chen. Using coplink to analyze criminal-justice data. *Computer*, 35(3):30–37, Mar 2002.

[15] Jon M. Kleinberg. Authoritative sources in a hyperlinked environment. *J. ACM*, 46(5):604–632, September 1999.

[16] Mark A Lauchs, Robyn L Keast, and Vy Le. Social network analysis of terrorist networks: can it add value? *Pakistan Journal of Criminology*, 3(3):21–32, 2012.

[17] Hady W Lauw, Ee-Peng Lim, Hweehwa Pang, and Teck-Tim Tan. Social network discovery by mining spatio-temporal events. *Computational & Mathematical Organization Theory*, 11(2):97–118, 2005.

[18] Jean Marie McGloin and David S Kirk. Social network analysis. In *Handbook of quantitative criminology*, pages 209–224. Springer, 2010.

[19] Jeffrey Scott McIllwain. Organized crime: A social network approach. *Crime, Law and Social Change*, 32(4):301–323, 1999.

[20] Chris Melde and Finn-Aage Esbensen. Gangs and violence: Disentangling the impact of gang membership on the level and nature of offending. *Journal of quantitative criminology*, pages 1–24, 2012.

[21] Jacob Levy Moreno. *Who shall survive? Foundations of sociometry, group psychotherapy and socio-drama* . Beacon House, 1953.

[22] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, November 1999. Previous number = SIDL-WP-1999-0120.

[23] Jialun Qin, Jennifer J Xu, Daning Hu, Marc Sageman, and Hsinchun Chen. Analyzing terrorist networks: A case study of the global salafi jihad network. In *Intelligence and security informatics*, pages 287–304. Springer, 2005.

[24] Britta Ruhnau. Eigenvector-centrality — a node-centrality? *Social Networks*, 22(4):357 – 365, 2000.

[25] Muhammad Akram Shaikh and Wang Jiaxin. Investigative data mining: Identifying key nodes in terrorist networks. In *Multitopic Conference, 2006. INMIC'06. IEEE*, pages 201–206. IEEE, 2006.

[26] Muhammad Akram Shaikh and Wang Jiaxin. Network structure mining: locating and isolating core members in covert terrorist networks. *WSEAS Transactions on Information Science and Applications*, 5(6):1011–1020, 2008.

[27] Muhammad Akram Shaikh, Jiaxin Wang, Zehong Yang, and Yixu Song. Graph structural mining in terrorist networks. In *Advanced Data Mining and Applications*, pages 570–577. Springer, 2007.

[28] Malcolm K. Sparrow. The application of network analysis to criminal intelligence: An assessment of the prospects. *Social Networks*, 13(3):251 – 274, 1991.

[29] Renée C van der Hulst. Introduction to social network analysis (sna) as an investigative tool. *Trends in Organized Crime*, 12(2):101–121, 2009.

[30] Muchou Wang and Weifeng Pan. A comparative study of network centrality metrics in identifying key classes in software. *Journal of Computational Information Systems*, 8(24):10205–10212, 2012.

[31] Stanley Wasserman. *Social network analysis: Methods and applications*, volume 8. Cambridge university press, 1994.

[32] Jennifer Xu and Hsinchun Chen. Untangling criminal networks: A case study. In *Intelligence and Security Informatics*, pages 232–248. Springer, 2003.

[33] Jennifer Xu, Byron Marshall, Siddharth Kaza, and Hsinchun Chen. Analyzing and visualizing criminal network dynamics: A case study. In *Intelligence and Security Informatics*, pages 359–377. Springer, 2004.

[34] Jennifer J Xu and Hsinchun Chen. Fighting organized crimes: using shortest-path algorithms to identify associations in criminal networks. *Decision Support Systems*, 38(3):473–487, 2004.

[35] Jennifer J. Xu and Hsinchun Chen. Crimenet explorer: A framework for criminal network knowledge discovery. *ACM Trans. Inf. Syst.*, 23(2):201–226, April 2005.

[36] Jie Xu. *Mining Static and Dynamic Structural Patterns in Networks for Knowledge Management: A Computational Framework and Case Studies*. PhD thesis, Tucson, AZ, USA, 2005. AAI3168596.

[37] Christopher C Yang. Knowledge discovery and information visualization for terrorist social networks. In *Intelligence and security informatics*, pages 45–64. Springer, 2008.

[38] Christopher C Yang, Nan Liu, and Marc Sageman. Analyzing the terrorist social networks with visualization tools. In *Intelligence and security informatics*, pages 331–342. Springer, 2006.