

Estudio de una Red de Criminales a Partir del Uso de Herramientas del Análisis de Redes Sociales

Facundo Emina

Juan Pablo Fiorenza

Redes complejas con aplicaciones a sistemas biológicos
Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

Resumen

En este trabajo se analiza una red de crímenes elaborada a partir de las personas vinculadas con 5 homicidios en la ciudad de St. Louis, Missouri, EE.UU. El uso del análisis de redes sociales (también llamado SNA, por sus siglas en inglés) en redes criminales se encuentra muy extendido a nivel internacional con el objetivo de lidiar y resolver los distintos problemas en relación con el crimen de cualquier tipo. A su vez, muchas veces suele ser difícil tratar con redes de este tipo debido a la confidencialidad de la información, y la falta usual de la misma. En este trabajo se utilizan algunas de las herramientas del SNA, como medidas de centralidad y coeficientes de similaridad entre nodos, para poder establecer relaciones entre sujetos centrales y los roles que cumplieron en sus crímenes.

1. Introducción

Uno de los posibles usos del análisis de redes sociales viene dado por el estudio de redes criminales. En la elaboración de éstas, por diversas razones no toda la información se encuentra registrada, por ejemplo, puede suceder que la relación entre dos criminales no se conozca, o no se tenga registro de los lugares que estos frecuentan, etc [1][2][3]. Debido a esto, resulta de interés determinar algún método que permita obtener información extra sobre los crímenes, o las personas involucradas en éstos, a partir de las características que presenta la red y su topología [1][2][4]. Se puede analizar la existencia de grupos de criminales que actúen en conjunto, esto se observa más en las redes de crimen organizado[4]; o predecir la existencia de un dado enlace, el cuál dependiendo de cómo fue elaborada la red denotará hechos distintos, por ejemplo, puede dar cuenta del intercambio de información entre dos sujetos [1][3][4].

Por otro lado, en otro tipo de redes, elaboradas a partir de bases de datos donde no se conoce quién cometió un determinado crimen, sino más bien aquellas personas que estuvieron de alguna manera involucradas en estos ya sea como sospechosas, víctimas o de otra forma, pregunta usual que se suele plantear es si existe algún parámetro para un dado nodo en la red que diga si éste es

importante. Para ello, utilizar medidas de centralidad topológica suele ser una herramienta utilizada frecuentemente para resolver las posibles incógnitas del caso [2][5].

Encontrar la existencia de nodos estructuralmente similares en una red criminal podría dar información de la existencia de personas con mismos intereses u objetivos[1]. Asimismo, ubicar grupos de personas que realicen crímenes en conjunto podría dar información útil a la hora de querer predecir un futuro crimen, o bien, investigar e identificar culpables en un determinado caso. Dependiendo de la información que brinde la red, uno podría realizar diversas hipótesis sobre el significado de una similitud entre un par de nodos. Muchos estudios muestran que los nodos tienen tendencia a estar conectados cuando comparten ciertas características; y a partir de dichas similitudes se tratan de predecir enlaces futuros [3][4][5][6][7][8].

En nuestro trabajo, contamos con una red de crímenes, elaborada a partir de la información que se poseía de todos aquellos crímenes relacionados con 5 homicidios que ocurrieron en St. Louis, Missouri, EE.UU. Para contextualizar la red, en los años 90 Rick Rosenfeld y Norm White usaron los registros de la policía para coleccionar información sobre crímenes en dicha ciudad. Comenzaron con 5 homicidios y guardaron los nombres de todos los individuos que hubiesen estado involucrados como víctimas, sospechosos o testigos. Luego, exploraron todos los archivos y adjuntaron todos los otros crímenes en los que esos mismos individuos habían estado involucrados. Además, se registró también los sexos de cada persona involucrada. La información fue extraída de la base de datos registrada en la Bibliografía [9]. En la red poseíamos la información de qué sujetos fueron sospechosos en cada crimen, quiénes víctimas y quiénes testigos. No se contó con la información de qué personas fueron los verdaderos criminales. Por lo tanto, la motivación principal del trabajo fue identificar si existía cierta correlación entre la importancia topológica de un nodo, y el rol que este tuvo en la mayoría de los crímenes en los que participó.

En la última sección del trabajo, se estudia la similitud estructural entre nodos. En nuestro caso, enlaces entre nodos dentro de la red pesada representan haber participado en un mismo crimen (sin importar el rol). Debido a que los enlaces en nuestra red bipartita sólo aportan la información de coparticipación en un crimen junto con sus roles respectivos en el mismo, y bajo la hipótesis ad-hoc de que una mayor frecuencia de rol sospechoso en crímenes aumenta la probabilidad de la persona de ser criminal, se analiza la similitud estructural basada en la cantidad de crímenes compartidos. Esto se debe a que este parámetro podría estar hablándonos de la probabilidad de que esas dos personas hayan participado conjuntamente como criminales en los casos donde se vieron involucrados.

Por último, queremos recalcar que a lo largo de todo el trabajo se trabajó sobre la hipótesis de que personas criminales poseen en su historial un número alto de crímenes en los que fue sospechoso, por lo tanto, en base a la frecuencia con la que una persona fue sospechosa en diversos crímenes, se plantea qué tan probable es que esa persona sea realmente un criminal. Debido a que en este trabajo **no se busca ejercer ningún tipo de juicio sobre personas**, sino, entender mejor y explorar la herramienta del análisis de redes sociales (en este caso, aplicado a una red criminal), no se trató generar ningún tipo de conclusiones acerca de quiénes fueron criminales. En otras palabras, la idea es mostrar un análisis que podría llevarse a cabo en caso de que la hipótesis planteada fuese

correcta. Además, se suma el hecho de que los resultados obtenidos no pudieron verificarse con la información real, debido a la ausencia de la misma a disposición del público, por lo que tampoco debe interpretarse el siguiente método de análisis como resolutivo al momento de hallar a los criminales de una red.

2. Caracterización de la red

2.1. Redes y Metadata

Para la elaboración de la red contamos con la información de todas aquellas personas que estaban involucradas en los crímenes registrados. Con ésta se armó una red bipartita compuesta por nodos de dos clases distintas: crímenes y personas. Los enlaces representan la participación de una persona en un crimen, y sólo se permiten enlaces entre nodos de clases distintas. Por otro lado, también se tuvo la información del rol que cada persona ocupó en los crímenes en los que participó, estando esta información contenida en el enlace. Los enlaces duales son aquellos donde la persona se encuentra involucrada en el crimen como sospechosa y víctima simultáneamente.

Personas		Enlaces				Crímenes
Masculino	Femenino	Sospechosos	Victimas	Testigos	Duales	
558 (67,31 %)	271 (32,69 %)	682	558	195	41	551

Cuadro 1: Características de la red.

A lo largo del trabajo se utilizó principalmente dos redes derivadas de los datos. La primera es la red bipartita ya comentada. La otra red consiste en una proyección de esta última sobre las personas. Los enlaces entre personas indican crímenes compartidos y el peso de estos la cantidad de crímenes que compartieron (personas que hayan estado involucradas en un sólo crimen, tendrán un enlace de peso 1, mientras que personas involucradas en 3 crímenes simultáneamente tendrán un enlace de peso 3). En esta última red, se pierde la información del rol a cambio de obtener una red que represente conexiones entre personas.

En cuanto a las características topológicas de la red, en la siguiente tabla se detallan el número de nodos y enlaces de la componente gigante de cada una.

Red (comp. gig.)	Número de nodos	Número de enlaces
Red Bipartita	1263	1377
Red Proyectada	754	2127

Cuadro 2: Tamaño componente gigante para cada red.

Ambas redes poseen una distribución de grado de tipo cola pesada como se observa en la Fig. (1). Para la red bipartita, la distribución de grado se analizó para las dos clases de nodos, y ambas dieron el mismo tipo de distribución. Para el caso de la red pesada, se tomaron dos medidas de grado

distintas, en donde se tuvo en cuenta el peso de los enlaces en una y en otra no, y su distribución también se muestra en la figura.

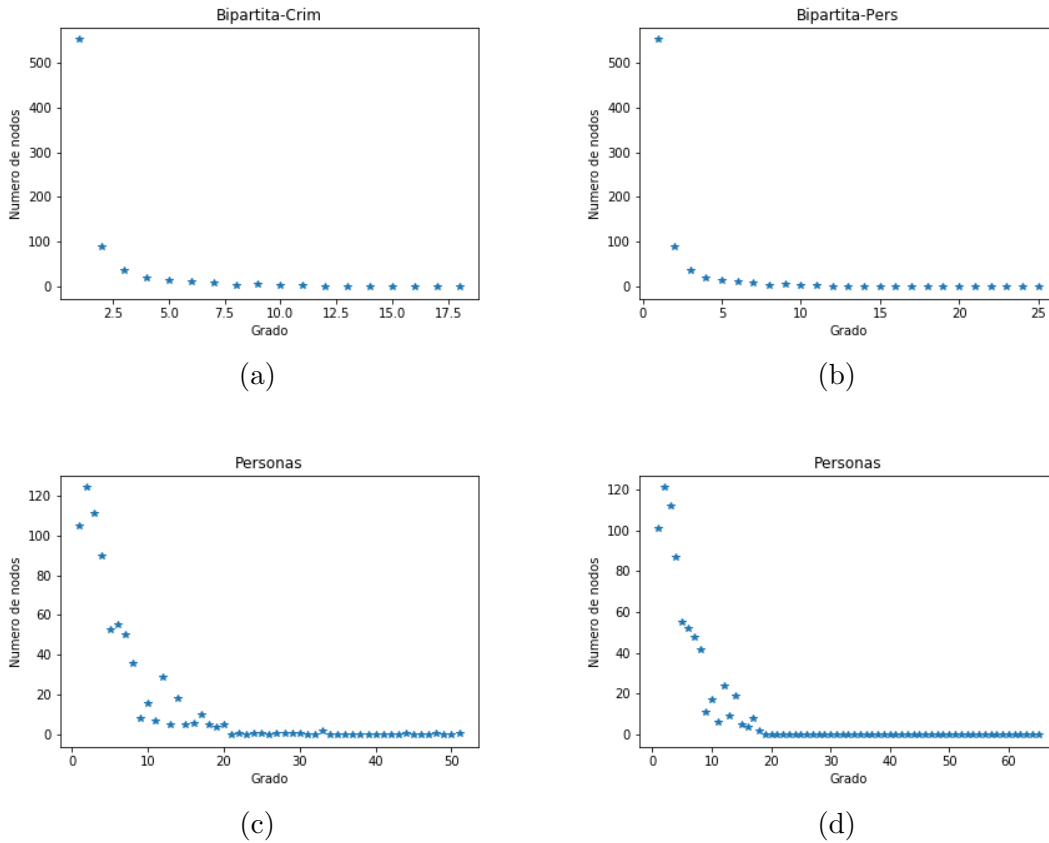


Figura 1: Distribución de grado para crímenes en red bipartita (a), personas en red bipartita (b), Personas sin pesar en red proyectada (c), personas con pesos en red proyectada (d)

A partir de esto, se puede deducir que la red se caracteriza por tener tanto personas como crímenes de alta conectividad, en contraste con una gran población de nodos de baja conectividad.

2.2. Estructura k-shell

En particular, la red proyectada posee una estructura de tipo k-shell. Esta estructura nos dice que la red posee un núcleo denso en conexiones, rodeado por una capa periférica de menor densidad. Usualmente, las redes criminales se caracterizan por tener este tipo de estructura en la que las personas de mayor jerarquía, y por ende, más importantes suelen ubicarse dentro del núcleo [1]. Para detectar esta estructura, lo que se hace es asignarles niveles de capas (k-shell) a cada nodo de la siguiente forma: se toman todos los nodos de grado 1, se les asigna el primer nivel de capa ($k=1$) y se los elimina de la red; luego, se les asigna el mismo nivel de capa a todos los nodos de grado menor o igual a dos y se los elimina de la red. Ahora, se repite el proceso para el segundo nivel ($k=2$) y así de forma sucesiva hasta acabar con todos los nodos.

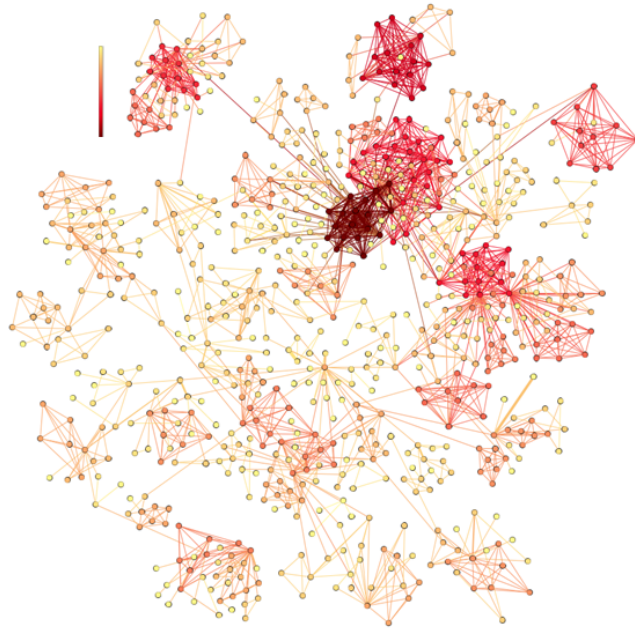


Figura 2:

Proyección de la red bipartita sobre las personas. En colores se encuentran pintadas las distintas capas que conforman la estructura k-shell de la red. Mientras más oscuro el color, más densa es la capa. Las capas que forman a nuestra red van desde $k=1$ hasta $k=17$.

3. Análisis de centralidad

Como se mencionó en la introducción, uno de los usos que se le da al análisis de redes sociales en redes criminales es encontrar a aquellos agentes que son más importantes en la red, dado que dependiendo del tipo de información contenida en la misma, encontrar a personas importantes puede implicar encontrar a los culpables de un crimen o a personas que sirvan de 'puente' entre criminales. En el caso de querer desarmar una red, como puede ser el caso de una red terrorista para frenar su accionar, es importante saber sobre qué personas se debe intervenir para garantizar el desarmado más eficiente. En nuestro caso, la idea de este análisis es encontrar aquellos nodos de los cuales depende la topología de la red y analizar su frecuencia de roles, bajo la hipótesis de que de una alta frecuencia de participación en los crímenes con rol sospechoso se infiere una mayor probabilidad de criminalidad.

Se estudió entonces la vulnerabilidad de la red en función del tipo de centralidad. En la bibliografía se puede leer que las medidas de grado, betweenness y coeficiente de clustering local son frecuentemente utilizadas en este tipo de análisis. Por lo tanto, se procedió a desarmar la red proyectada quitando de mayor a menor valor de centralidad los nodos, y a medida que se hacía se fue calculando en cada paso el tamaño de la componente gigante. En la figura 5 se muestra la proporción del tamaño de la componente gigante en función de la proporción de nodos quitados para las medidas de grado en la red proyectada (suma de los pesos de los enlaces para cada nodo), grado en la red bipartita, betweenness y coeficiente de clustering local. Por último, se comparó con

la variación en el tamaño de ésta al quitar la misma cantidad de nodos, pero tomados al azar, para tener una medida de control.

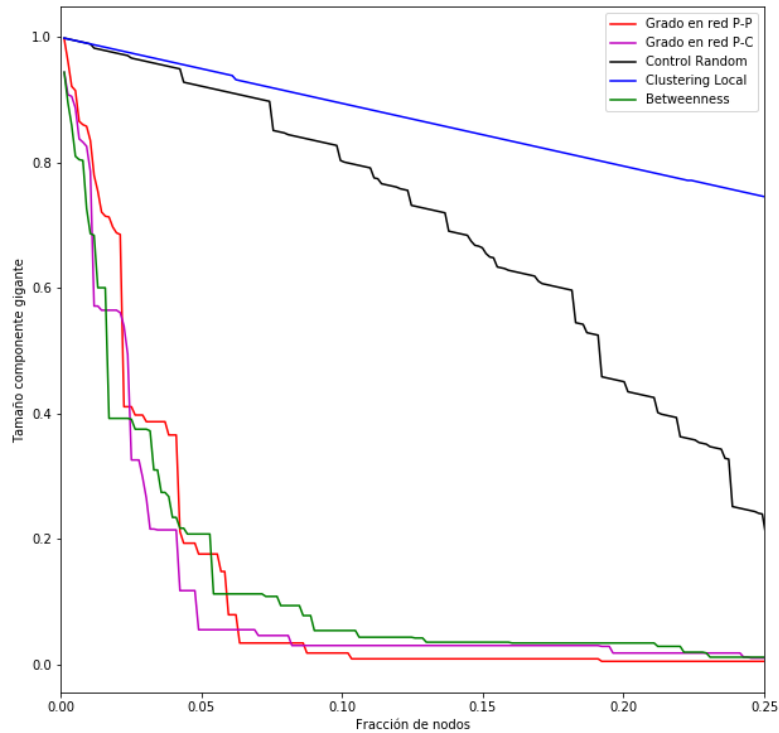


Figura 3: Impacto de los distintos valores de centralidad en el tamaño de la componente gigante.

Como puede observarse, el impacto de Betweenness, el grado en la red proyectada y el grado en la red bipartita son muy similares, y mucho mayores que el impacto de extraer los nodos con mayor coeficiente de clustering o nodos al azar. De hecho, el impacto generado por extraer los nodos de clustering más alto fue menor al del control random.

En este punto, es importante interpretar los resultados obtenidos. De la figura 5, se puede notar que la conectividad de la red depende casi de igual manera de los nodos de mayor Betweenness o grado de ambas redes. Las personas con alto grado en la red bipartita representan a aquellas personas con mayor frecuencia en participación de crímenes, mientras que personas con alto grado en la red proyectada representan aquellas personas involucradas no solo en varios crímenes, sino también con varias personas. Asimismo, personas con mayor grado de Betweenness son aquellas que se ven atravesadas por un mayor número de caminos mínimos entre nodos, por lo que representan personas 'puente' en la red. Se requiere analizar cual de estos valores de centralidad representa mejor a personas con mayor frecuencia de rol sospechoso.

Por otro lado, una de las razones por las cuales personas de coeficiente de clustering local alto no resulten verdaderamente importantes a nivel topológico es que, en la red proyectada, los crímenes generan cliques para todo trío de personas involucrados en el crimen. En la figura 4 se ilustra un ejemplo sencillo. Por lo tanto, personas en crímenes con mucha población involucrada adquieren un clustering local alto, lo que no significa que sean importantes para la conectividad de la red. De hecho, quitar un nodo de estos clusters de cliques no afecta al cluster significativamente dado que todos están interconectados. Sin embargo, el coeficiente de clustering local podría tener otras interpretaciones importantes que quedan excluidas de este análisis.

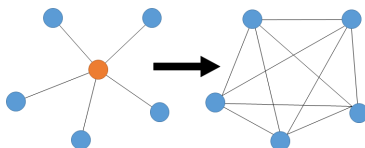


Figura 4

Descartando al coeficiente de clustering local, se realizó un ranking de los 15 nodos más centrales para cada medida (grado en red pesada y bipartita y betweenness) y se los comparó entre sí a partir de su solapamiento. En la tabla 3 se muestran los resultados del overlap entre rankings.

Overlap Rankings (15 personas)	Betweenness	Grado de red Pesada	Grado de red Bipartita
Betweenness	100 %	46,67 %	46,67 %
Grado de red pesada	—————	100 %	66,67 %
Grado de red Bipartita	—————	—————	100 %

Cuadro 3: Overlap entre los valores de centralidad.

donde, por ejemplo, el 46,67 % de las 15 personas con mayor grado también tuvieron un alto valor de Betweenness.

Se prosiguió a analizar la proporción de sospechosos dentro de cada ranking de las 15 personas con mayor valor de centralidad, tomando como criterio que el rol correspondiente a una persona es aquel que tuvo la persona en el 50 % o más de los crímenes. De esta manera hubo personas que por haber tenido presencia de dos o tres roles pero siempre menor al 50 % no se las sumó a ningún rol. También, para los cuales se los consideró como sospechosos dado que todos habían participado en al menos otro crimen como sospechosos. El porcentaje de sospechosos correspondiente a cada ranking, en base a este criterio, se lo tomó como el valor esperado de sospechosos para el ranking. Se comparó si la proporción de veces que una persona fue sospechosa se correspondía o no con una distribución al azar de los roles. Este contraste con una hipótesis nula se debe a que haber participado en un crimen con rol de sospechoso no implica ser un criminal, y la asignación de un cierto rol puede deberse a una diversa cantidad de factores (como por ejemplo, haber encontrado a una persona inocente en la escena del crimen). Para ello, se distribuyó sucesivas veces en la

red bipartita los roles, y se calculó para cada distribución al azar la proporción de veces que cada persona fue sospechosa, elaborándose un histograma para cada una de las primeras 15 personas con mayor valor de centralidad.

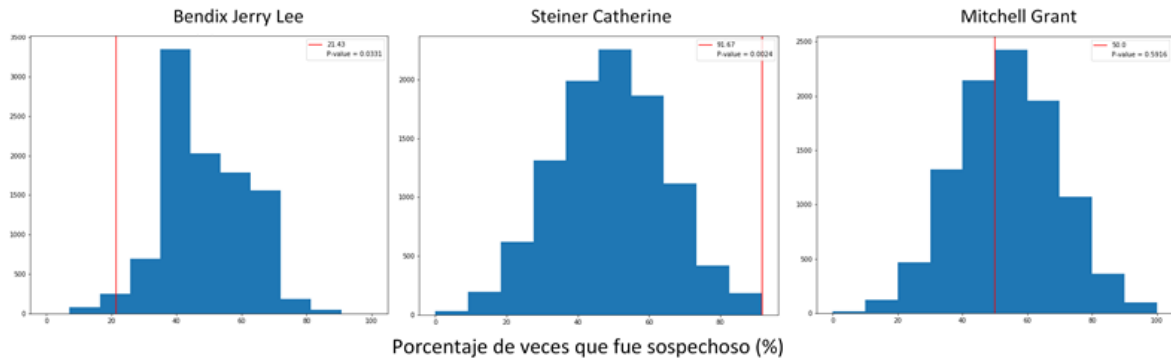


Figura 5: Histograma donde la línea roja representa el porcentaje real de sospechoso de la persona. A la izquierda, un caso donde la frecuencia observada en veces que fue sospechoso es menor a la media de la distribución al azar, con un $p\text{-value}=0,0331$. En el centro, el caso de un sujeto que fue más veces sospechoso que la media de la distribución al azar con un $p\text{-value}=0,0024$. A la derecha, el caso de un sujeto cuya frecuencia de rol sospechoso es igual a la esperada en una distribución al azar de roles, con un $p\text{-value}=0,5916$.

Sospechosos	Betweenness	Grado de red Pesada	Grado de red Bipartita
Esperado	70 %	75 %	70 %
Real	47 %	53 %	67 %

Cuadro 4: Proporción de sospechosos para cada valor de centralidad.

Donde el valor esperado representa el porcentaje de personas cuya participación en los crímenes fue en un 50 % o más como sospechosos, y el valor real la proporción de personas que además de cumplir la condición anterior se alejaron de la media del histograma y dieron un P-Value menor a 0,05.

Para todas aquellas personas cuyo rol fue considerado sospechoso, por ejemplo como puede observarse en la tabla (4) el 47 % de las personas con mayor Betweenness, se descartó la hipótesis nula de que fueron sospechosas al azar. Se obtuvo que el ranking de Betweenness tiene el menor porcentaje de sospechosos, mientras que el grado en la red bipartita no solo contiene el mayor porcentaje de sospechosos sino también una mayor correspondencia con el valor esperado.

4. Cluster de sospechosos

Con el fin de estudiar la existencia de grupos de personas que trabajen juntas a la hora de cometer crímenes, se rankeó a los enlaces en función de su peso en la red proyectada. El criterio que se

utilizó fue considerar como enlaces pesados aquellos que tuvieron un peso mayor a 3, representando estos el 0,85 % del total de enlaces en la componente gigante. A estos enlaces pesados se los clasificó en dos subgrupos (dentro de los que entraron casi todos los enlaces pesados): 8 de estos enlaces (el 38,00 %) unían pares de personas donde una de ellas había sido víctima más de la mitad de las veces en las que compartieron crímenes y la otra, sospechosa más de la mitad de las veces; a su vez, otra fracción de los enlaces (también representada por el 38,00 %) unían pares de personas donde ambas habían sido sospechosas en más de la mitad de los crímenes compartidos. Se observó que en estos últimos enlaces participaba un grupo de 5 nodos, interconectados por los mismos. En la Fig. (6) se muestran los enlaces pintados de este cluster de personas sospechosas.

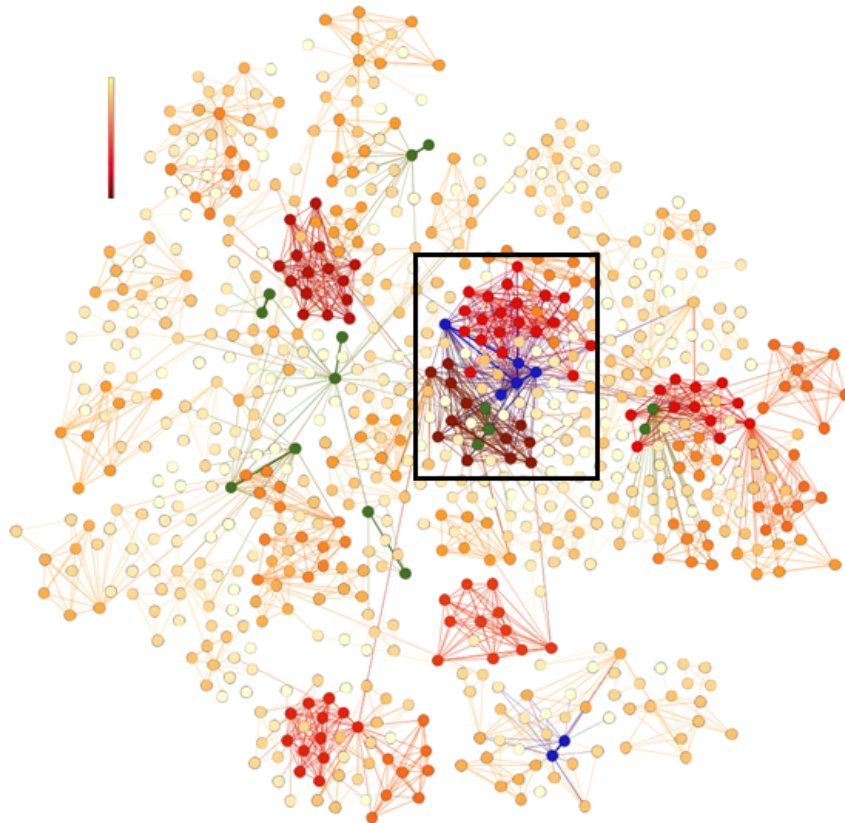


Figura 6: Red de criminales con los nodos pintados de acuerdo a su capa k-shell. En particular, los nodos pintados de verde representan enlaces donde en 3 o más crímenes compartidos, más del 50 % de las veces una persona fue víctima y la otra sospechosa. En azul, los nodos que compartieron más de 2 crímenes en los que ambos fueron más de un 50 % de las veces sospechosos. Con el cuadro se remarca el núcleo de la red.

Se puede observar que estos 5 nodos se encuentran dentro del núcleo de la red. Además, éstos se encontraban dentro de los 15 de mayor grado en la red bipartita (es decir, dentro de las 15 personas que se vieron involucradas en más crímenes). Tres de estos mismos nodos también formaban parte del ranking de grado en la red pesada (es decir, dentro de las 15 personas con mayor cantidad de

vínculos), y dos de estas últimas tres se encontraban dentro del ranking de Betweenness. Además, todas estas personas rechazaron la hipótesis nula en sus debidos histogramas.

Debido a la fuerte correspondencia entre la pertenencia de algunos de los miembros del cluster a los rankings de centralidad, junto con el hecho de que en sus roles mayoritarios fueron sospechosos, se quiso verificar la importancia topológica de este cluster en comparación con otro grupo de sospechosos, es decir, personas que siendo consideradas sospechosas estuvieran enlazadas entre sí en la red pesada.

Analizar el impacto que tiene quitar a este grupo de sospechosos junto con los crímenes que cometieron podría decirnos que tan importante es este grupo. Para ello, analizamos cómo variaba el tamaño de la componente gigante, tanto en la red bipartita como en la red pesada, al sacar a todas estas personas con los crímenes en los que estuvieron involucradas. Luego, lo comparamos con la hipótesis nula: sacar al azar grupos formados por la misma cantidad de personas, vinculadas entre sí por crímenes en los que fueron simultáneamente sospechosas, con sus respectivos crímenes en los que participaron.

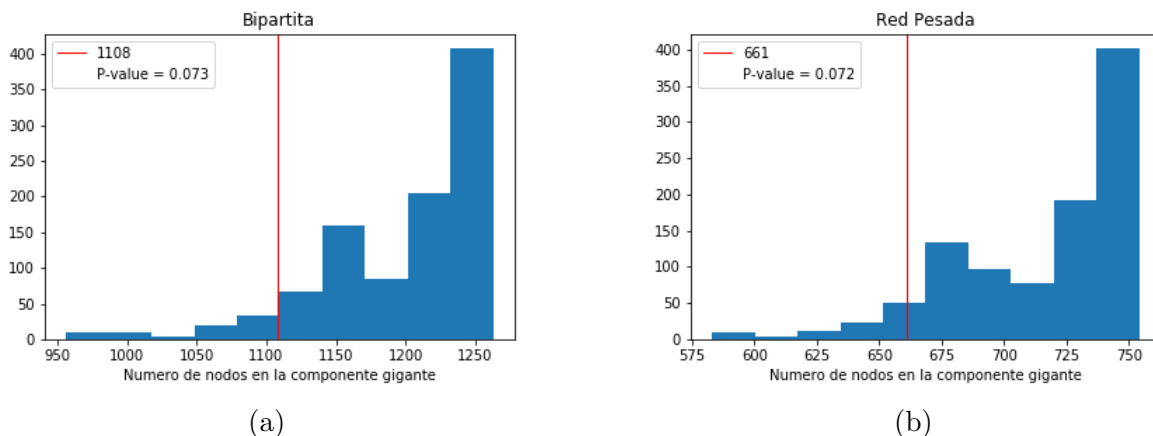


Figura 7: Impacto en el tamaño de la componente gigante del cluster de sospechosos.

Como puede observarse en la Fig. (7) el impacto en el tamaño de ambas componentes gigantes, es decir, en la red bipartita y en la red pesada, fue mayor al quitar las personas del cluster de sospechosos hallado que eliminar azarosamente grupos de personas que hubiesen participado simultáneamente en crímenes como sospechosas. Los p-valores obtenidos para ambos histogramas fueron bajos, si bien no menores a 0,05.

A su vez, el hecho de que el análisis previo compare grupos de sospechosos, es decir, personas cuyo rol simultáneo en el crimen que comparten fue sospechoso, nos permite formular la pregunta de si existe alguna acción conjunta dentro de nuestra red.

5. Análisis de similaridad

Bajo la hipótesis de que un alto número de crímenes compartidos podría ser característica de un grupo de criminales, analizamos la similaridad estructural de cada par de nodos enlazados en la red proyectada. Para cada par de nodos, calculamos el número de crímenes compartidos y les restamos el número de crímenes que compartirían en la red si se mantuviese su grado pero se asignaran aleatoriamente los crímenes en los que participaron. Para el cálculo realizado, consideramos que el rol de testigo no aporta información a la similaridad entre nodos, por lo que no tuvimos en cuenta estos enlaces en el análisis. En la ecuación 1 mostramos la expresión de este coeficiente de similaridad. Se puede notar que es el mismo que aparece en el denominador de la correlación de Pearson

$$S_{ij} = \sum_k^N A_{ik}A_{jk} - \frac{k_i k_j}{n} \quad (1)$$

donde A_{ik} es la matriz de adyacencia, cuyos valores son 1 cuando la persona i comparte enlace con el crimen j , y 0 cuando no. De allí se tiene que $\sum A_{ik}A_{jk}$ es el número de crímenes compartidos, n es el número de crímenes en la red y $\frac{k_i k_j}{n}$ es la cantidad de vecinos que compartirían en una red aleatoria (siendo k_i el grado de cada nodo en la red bipartita). No normalizamos, como lo hace la correlación de Pearson, dado que así es más fácil ubicar un umbral.

Una vez calculadas las similaridades para todo par de personas que comparten crímenes, elegimos distintos umbrales de similaridad (en particular, 0, 1, 2, 3 y 4) y graficamos el número de pares por encima del umbral, en función del umbral. En la figura 8 se muestra dicho gráfico. Se tomó como umbral de referencia $S = 2$ dado que para $S = 1$ había más de 30 pares de nodos (número alto) y para $S = 3$ había 3 nodos (número muy bajo). El valor medio de similaridad entre los pares de nodos de la red es de 1,03. Para un umbral mayor a 2, solo hay 7 pares de nodos que lo sobrepasan, es decir que la probabilidad de encontrar un enlace con una similaridad mayor o igual a la umbral es del 0,83%, muy baja.

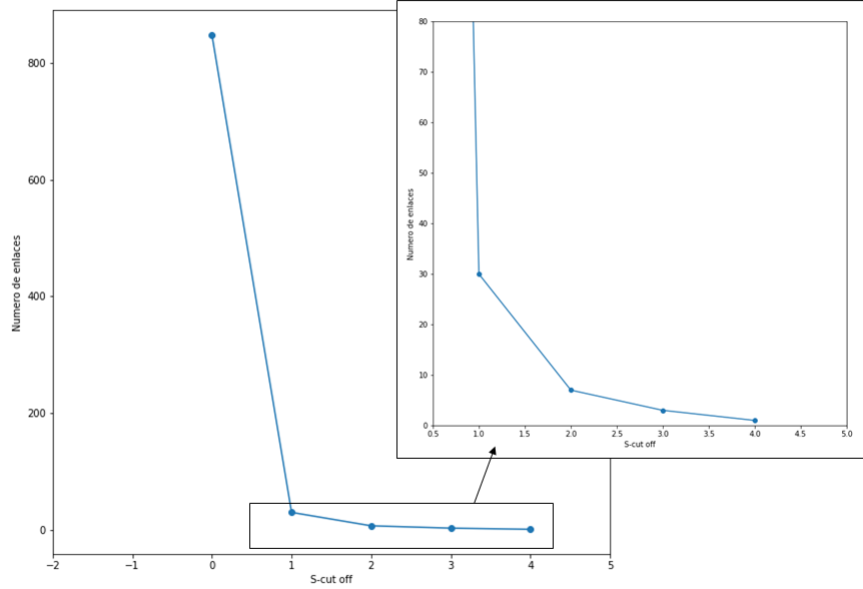


Figura 8: Numero de pares de nodos en función del umbral con similaridad mayor a este.

Ahora bien, habiendo encontrado la similaridad entre pares de nodos y habiendo fijado un umbral, aun no es posible generar conclusiones acerca de las características que puedan compartir estas personas. Un par de nodos con alta similaridad podría corresponderse con un par de personas donde una fue atacada por la otra en varios crímenes, y en términos de similaridad estructural, es indistinguible de dos personas que coparticiparon como criminales en más de un crimen. A su vez, una frecuente participación en crímenes en conjunto sin una notable correlación entre sus roles tampoco podría implicar una coparticipación criminal.

Como método de clasificación entre estos pares, volvimos a calcular su similaridad, haciendo uso de la información contenida en los enlaces. Ahora, le asignamos a la matriz de adyacencia 3 valores posibles,

$$A_{ij} = \begin{cases} 1 & \text{si participó en crimen como sospechoso} \\ 0 & \text{si no participó en crimen} \\ -1 & \text{si participó en crimen como víctima} \end{cases}$$

La ecuación de similaridad puede expresarse como la suma de tres términos, donde dos de ellos denotan la similaridad de rol (S_{vij} y S_{sij}) y el otro denota la disimilaridad de rol [ver Apéndice].

$$S_{ij} = S_{sij} + S_{vij} - D_{ij} \quad (2)$$

De esta manera, la similaridad calculada para los pares de nodos nos dará negativa en caso de que los roles cumplidos en los crímenes compartidos hayan sido contrarios, indicando una característica de disimilaridad entre personas. Mientras que una similaridad positiva estará hablando de una coincidencia en roles dentro de los crímenes compartidos. A su vez, valores cercanos al cero se corresponderán con una cancelación entre términos e indicarán que no es posible establecer una similaridad entre roles utilizando la información disponible.

En la figura 9 se muestra el número de nodos para cada valor de similaridad $S \pm 0,5$. Se obtuvo una polarización de similaridades, donde un grupo de 4 pares de nodos obtuvo un valor menor a -2, mientras que un grupo de 3 pares obtuvo similaridad mayor a 2.

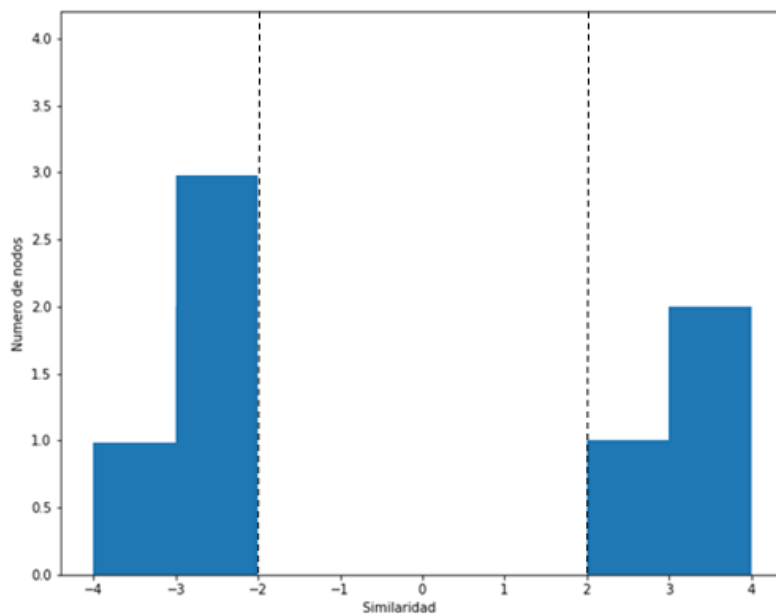


Figura 9: Numero de pares de nodos en función del umbral con similaridad mayor a este.

Dentro de estos 3 pares de nodos, hay presentes 4 nodos distintos, de los cuales, todos pertenecen al cluster de sospechosos hallado previamente. Por lo tanto, todas estas personas se encuentran en el ranking de las 15 personas con mayor número de crímenes involucrados. Asimismo, se contrastaron con la hipótesis nula de rol, es decir que tuvieron un p-valor menor a 0,05 en la distribución de rol al azar. Por lo tanto, valores de similaridad mayores al umbral establecido podrían permitirnos inferir que existe alguna posibilidad de que dichos nodos actúen de manera conjunta como potenciales criminales.

6. Conclusiones y discusiones

Se encontró que la red es igual de vulnerable a la hora de ser desarmada comenzando por los nodos con centralidad más alta en función del betweenness, grado de la red bipartita y grado ponderado con pesos de la red proyectada. A diferencia del coeficiente de clustering, que no permitió distinguir entre nodos importantes en la topología de la red. Dentro de los rankings de las 15 personas más centrales realizados para cada valor de centralidad, el ranking de grado de la red bipartita tuvo mayor correspondencia y cantidad de personas que se esperaba que fuesen sospechosas (en contraste con una distribución de rol al azar). Por otro lado, los rankings de centralidad tuvieron un solapamiento de al menos 47 %, lo que significa que comparten casi la mitad de los nodos. Recordamos que se trabajó sobre la hipótesis de que una alta frecuencia de rol sospechoso aumenta las probabilidades de que esa persona sea efectivamente un criminal. Sin embargo, los resultados no pudieron ser respaldados por información verdadera debido a la falta de la misma. Recordamos también que con postular esta hipótesis, no se está tratando de ejercer juicio sobre las personas involucradas, sino asumir un cierto conocimiento sobre las personas para poder trabajar con los datos.

A partir de un ranking de pesos de los enlaces en la red proyectada, se encontró un cluster de sospechosos preteneciente al núcleo de la red. El mismo estaba formado por 5 nodos, de los cuales, 5 estaban entre los 15 de grado más alto en la red bipartita, 3 de ellos también estaban rankeados en los 15 de grado más alto en la red proyectada, y dos de ellos, rankeados en los 15 de betweenness más alto. Además, para todos ellos, su frecuencia de rol sospechoso había sido contrastada con la hipótesis nula de rol, y los 5 tuvieron una frecuencia mayor a la media aleatoria, con p-value menor a 0,05. Además, se encontró que este grupo de personas, al ser eliminado junto con los crímenes en los que participó cada miembro, la componente gigante de ambas redes (bipartita y proyectada) se desarmaba más que al eliminar al azar cualquier grupo de personas que hayan compartido crímenes con rol sospechoso. Esto permitiría obtener un indicio de que este cluster de personas tiene cierta funcionalidad en la red, y que debido a su densidad en conexiones de rol sospechoso, sea factible que se trate de un grupo que comete crímenes en conjunto. Poder identificar este tipo de estructuras en redes criminales podría permitir predecir la existencia de crímenes en el futuro a partir de colaboraciones entre personas.

Se encontró además, que definiendo una similaridad estructural de tipo correlación de Pearson sin normalizar, existe un porcentaje de nodos muy bajo en la red (0,83 %) que supere el umbral de similaridad $S=2$. Dentro de estos nodos más similares, se les calculó de vuelta la similaridad, pero teniendo en cuenta que haber cumplido un mismo rol en el crimen, suma de forma positiva a la similaridad, mientras que haber cumplido distinto rol, suma de forma negativa. De esta manera, nodos de similaridad positiva y mayor a cierto umbral indican una similaridad funcional, además de estructural. Se obtuvo que solo 4 nodos tienen una similaridad positiva mayor a 2. Además, todos ellos pertenecen al cluster de sospechosos hallado previamente. Por lo tanto, suponiendo que efectivamente el cluster encontrado se corresponde con un grupo de personas que trabajan en conjunto en crímenes, la similaridad funcional definida podría ser un buen indicador para hallar personas que trabajen en conjunto.

7. Apéndice

7.1. Similaridad

Usando que

$$A_{ij} = \begin{cases} 1 & \text{si participó en crimen como sospechoso} \\ 0 & \text{si no participó en crimen} \\ -1 & \text{si participó en crimen como víctima} \end{cases}$$

$$k_i k_j = (k_{si} + k_{vi})(k_{sj} + k_{vj}) \quad (3)$$

Se puede desarrollar la similaridad de la siguiente manera:

$$S_{ij} = \sum_k^N A_{sik} A_{sjk} - \frac{k_{si} k_{sj}}{n} + \sum_k^N A_{vik} A_{vjk} - \frac{k_{vi} k_{vj}}{n} + \sum_k^N A_{sik} A_{vjk} - \frac{k_{si} k_{vj}}{n} + \sum_k^N A_{vik} A_{sjk} - \frac{k_{vi} k_{sj}}{n} \quad (4)$$

Donde $\sum A_{sik} A_{sjk}$ es el número de crímenes donde ambos fueron sospechosos; $\sum A_{vik} A_{vjk}$ es el número de crímenes donde ambos fueron víctimas; y $\sum A_{sik} A_{vjk}$ es el número de crímenes donde i fue sospechoso y j fue víctima. Éste último término es negativo, mientras que los dos primeros son positivos por cómo se define A_{ij} . Los grados k_{vi} y k_{si} representan el número de veces que i fue víctima y sospechoso, respectivamente. Por último, n es el número de crímenes en la red.

Si corregimos el signo en los términos $\frac{k_{vi} k_{sj}}{n}$ para restarle al número de crímenes en común con rol contrario, el número de crímenes que compartirían con roles opuestos en una red aleatoria, podemos expresar a la similaridad como,

$$S_{aij} = \sum_k^N A_{aik} A_{ajk} - \frac{k_{ai} k_{aj}}{n}$$

$$D_{ij} = - \sum_k^N (A_{sik} A_{vjk} + A_{vik} A_{sjk}) - \frac{k_{si} k_{vj} + k_{vi} k_{sj}}{n}$$

(5)

con a=s ó v.

8. Referencias

- [1] Berlusconi G, Calderoni F, Parolini N, Verani M, Piccardi C,(2016) Link Prediction in Criminal Networks: A Tool for Criminal Intelligence Analysis PLoS One 11(4): e0154244.
- [2] Rostami A, Mondani H (2015) The Complexity of Crime Network Data: A Case Study of Its Consequences for Crime Control and the Study of Networks. PLoS ONE 10(3): e0119309. doi:10.1371/journal.pone.0119309
- [3] Emra Budur, Seungmin Lee, Vein S Kong. Structural Analysis of Criminal Network and Predicting Hidden Links using Machine Learning. [<https://arxiv.org/pdf/1507.05739.pdf>]
- [4] Jennifer Xu, Hsinchin Chen. Criminal Network Analysis And Visualization. June 2005/Vol. 48, No. 6 COMMUNICATIONS OF THE ACM
- [5] Barta Daniel (16/03/2018). Fight Crime with Social Network Analysis. Hackermoon: <https://hackermoon.com/fight-crime-with-social-network-analysis-7a879d4a65ea>
- [6] Liben-Nowell D, Kleinberg J. The link-prediction problem for social networks. Journal of the American Society for Information Science and Technology. 2007;58(7):1019–1031.
- [7] Guimera R, Sales-Pardo M. Missing and spurious interactions and the reconstruction of complex networks. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(52):22073–22078. pmid:20018705
- [8] Crime network dataset – KONECT, April 2017.
http://konect.uni-koblenz.de/networks/moreno_crime
- [9] David Liben-Nowell, Jon Kleinberg (08/01/2004). The Link Prediction Problem for Social Networks. [<https://es.scribd.com/document/112560130/link-prediction-in-social-networks>]