

REDES- Revista hispana para el análisis de redes sociales
Vol. 24,#2, Diciembre 2013
<http://revista-redes.rediris.es>

*El problema del actor clave*¹

*Stephen P. Borgatti*²

*University of Kentucky*³

Resumen

Se describe un procedimiento para identificar actores clave en una red social. Un supuesto básico es que la selección óptima depende de los fines de dicha selección. De acuerdo con ello, se articulan dos metas genéricas, referidas al problema del actor clave en términos positivos y negativos. Primero se propone un procedimiento para identificar actores clave con el objetivo de difundir algo de manera óptima en la red, valiéndose de los actores clave como semillas. El segundo procedimiento identifica actores clave con el objetivo de perturbar o fragmentar la red eliminando algunos de sus nodos. Los indicadores de centralidad habituales no son óptimos para este propósito, por lo que se proponen nuevos indicadores.

Palabras clave: Redes sociales – Centralidad – Cohesión.

Abstract

A procedure is described for finding sets of key players in a social network. A key assumption is that the optimal selection of key players depends on what they are needed for. Accordingly, two generic goals are articulated, referring to key player problem in positive (KPP-1) and negative (KPP-2) terms. KPP-1 is defined as the identification of key players for the purpose of optimally diffusing something through the network by using the key players as seeds. KPP-2 is defined as the identification of key players for the purpose of disrupting or fragmenting the network by removing the key nodes. It is found that off-the-shelf centrality measures are not optimal for solving either generic problem, and therefore new measures are presented.

Key words: Social networks – Centrality – Cohesion.

Introducción

El problema del actor clave (PAC) consta de dos sub-problemas diferentes, que pueden describirse de forma general como sigue:

¹ Publicado originalmente en: Borgatti, S. P. (2003). The Key Player Problem. In Breiger, R., K. Carley and P. Pattison (eds.) *Dynamic Social Network Modelling and Analysis: Workshop Summary and Papers*, pp. 241-252. Washington, D.C.: National Academy Press. Traducción de Isidro Maya Jariego <isidromj@us.es>.

² Agradecimientos: Este trabajo fue apoyado por la Oficina de Investigación Naval con la beca número N000140211032. Gracias a Scott Clair por acercarme a este problema, a Mark Newman por sugerirme la distancias recíprocas, a Kathleen Carley por las útiles discusiones que mantuve con ella, y a Valdis Krebs por proporcionarme datos ilustrativos.

³ En el momento de la publicación el autor era miembro del Department of Organization Studies, Carroll School of Management, Boston College, Chestnut Hill, MA 02467.

1. (PAC-1) Dada una red social, identifica un conjunto de nodos k (llamado *conjunto-kp*⁴ de orden k) que, si se retira perturbarían al máximo la comunicación entre los nodos restantes.
2. (PAC-2) Dada una red social, identifica un conjunto-kp de orden k que está conectado al máximo con todos los demás nodos.

Por supuesto, estas definiciones introductorias dejan de lado qué significa exactamente "perturbación máxima de la comunicación" y "conectado al máximo". Parte del proceso de resolución de estos problemas consiste en proporcionar definiciones de estos conceptos que conduzcan a soluciones factibles y resultados útiles. Sin embargo, parece claro que el PAC-1 implica fragmentar una red en componentes, o, en su defecto, hacer que las distancias entre los nodos sean tan amplias que estén prácticamente desconectados. Por el contrario, el PAC-2 consiste en encontrar los nodos que pueden llegar a la mayor cantidad de nodos restantes posible, ya sea a través de enlaces directos o de caminos cortos.

El primer problema, PAC-1, surge en una serie de contextos. Un buen ejemplo en el contexto de la salud pública es el problema de la inmunización o cuarentena. Dada una enfermedad infecciosa que se transmite de persona a persona, y dado que no es factible inmunizar y/o poner en cuarentena a toda una población, ¿qué subconjunto de los miembros deberían ser vacunados o puestos en cuarentena con el fin de dificultar al máximo la propagación de la infección? Un ejemplo en el contexto militar es la selección de objetivos. Dada una red de terroristas que deben coordinarse para poner en marcha ataques efectivos, y dado que sólo se puede intervenir sobre pequeño número de ellos (por ejemplo, mediante la detención o el descrédito), ¿cuáles deberían ser elegidos con el fin de perturbar al máximo la red?

El segundo problema, PAC-2, surge en el contexto de la salud pública cuando una organización tiene que seleccionar un pequeño grupo de miembros de la población para utilizarlos como semillas en la difusión de prácticas o actitudes que promueven la salud, tales como el uso de lejía para desinfectar las agujas. En el contexto de la gestión de organizaciones, el problema se produce cuando la directiva quiere poner en marcha una iniciativa de cambio y tiene que conseguir que se sume primero un pequeño grupo de líderes informales, por ejemplo con una intervención de fin de semana con ellos. En el contexto militar, el PAC-2 se traduce en la localización de

⁴ Mantenemos en este caso las iniciales originales del inglés para actores clave (KP, *Key Player*) [Nota del Traductor]

un conjunto eficiente de enemigos a los que vigilar, desinformar o convertir en agentes-doble.

A primera vista, tanto el PAC-1 como el PAC-2 parecerían resolverse fácilmente empleando algunos conceptos de la teoría de grafos tales como los puntos de corte y los conjuntos de corte, o por medio de los métodos de análisis de redes sociales, tales como la medición de la centralidad de los nodos. Resulta, sin embargo, que ninguno de los métodos existentes son adecuados. Este documento explica por qué y presenta un nuevo enfoque diseñado específicamente para el problema del actor clave.

Enfoque de centralidad

La aproximación basada en la idea de centralidad consiste en medir la centralidad de cada nodo de la red para, a continuación, seleccionar los k nodos más centrales que conforman el conjunto- k_p . Puesto que existen muchas medidas de centralidad, una pregunta que se plantea es qué indicadores utilizar. Para el PAC-1, podemos esperar que las mejores medidas sean aquellas basadas en la intermediación (*betweenness*). Por ejemplo, la intermediación de Freeman suma la proporción de los caminos más cortos desde un nodo a otro que pasan a través de un nodo dado. Por lo tanto, un nodo con alta intermediación se encarga de conectar muchos pares de nodos a través de la mejor ruta, de forma que la eliminación de dicho nodo debería causar que muchos pares de nodos estuviesen más distantes entre sí (si es que no se quedan completamente desconectados).

Para el PAC-2, podemos esperar que las medidas basadas en la centralidad de grado (*degree*) y la cercanía (*closeness*) sean útiles. La centralidad de grado es simplemente el número de nodos a los que un nodo dado es adyacente. Por lo tanto, dependiendo de la relación social representada por el grafo y asumiendo que la adyacencia implica potencial de influencia, un nodo con alto grado puede influir potencialmente de forma directa a muchos otros nodos. La cercanía se define como la suma de las distancias geodésicas de un nodo dado a todos los demás, donde la distancia geodésica se refiere a la longitud de la ruta más corta entre dos puntos. Por lo tanto, un nodo con una puntuación de cercanía baja (muy central) debe ser capaz de influir, directa e indirectamente, a muchos otros.

Las medidas de centralidad son soluciones plausibles al PAC. Sin embargo, no son óptimas. Hay dos problemas básicos, a los que me refiero sucesivamente a

continuación como el problema de diseño y de selección de grupo. De los dos, el problema de la selección de grupo es el más grave.

El problema del diseño

El problema del diseño surge en última instancia por el hecho de que las medidas de centralidad no se diseñaron con el PAC-1 y el PAC-2 específicamente en mente. Empezando con el PAC-1, tomemos en consideración el grafo de la Figura 1.

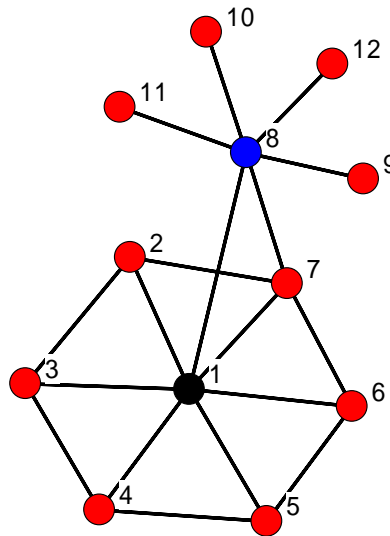


Figura 1.

El nodo 1 tiene la mayor centralidad en todas las medidas consideradas, incluyendo la intermediación. Sin embargo, la supresión del nodo 1 tiene un efecto relativamente pequeño en la red. Las distancias entre ciertos pares de nodos aumentan de hecho, pero es evidente que la comunicación entre todos los puntos sigue siendo posible, ya que no hay fragmentación. En contraste, la supresión del nodo 8, que no tiene el más alto grado de intermediación, es más eficaz. La eliminación del 8 divide el grafo en cinco fragmentos inconexos (componentes).

Para el PAC-2, el panorama es un poco más claro. Si formulamos el PAC-2 en términos de llegar directamente a la mayoría de los nodos, la centralidad de grado es óptima. Si se formula en términos de llegar a la mayor cantidad de nodos en hasta m pasos, entonces podemos definir fácilmente una nueva medida de centralidad ("centralidad de alcance- m ") que cuenta el número de nodos dentro de la distancia m de un nodo dado.

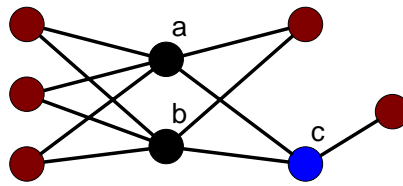


Figura 3.

Enfoques de teoría de grafos

Además de conceptos básicos, tales como los componentes y la distancia, una serie de conceptos de teoría de grafos son relevantes para el PAC. Para el PAC-1, los más evidentes son las nociones de punto de corte y puente, que son respectivamente los nodos y las líneas cuya eliminación aumentaría el número de componentes en el grafo. Estos conceptos no se refieren a la cuestión de la selección del grupo. Sin embargo, ambos tienen sus homólogos a nivel de conjuntos, en la forma de conjuntos de corte. Un conjunto de corte es un conjunto de nodos (o líneas) cuya eliminación aumentaría el número de componentes en el gráfico. La mayor parte del trabajo se ha centrado en conjuntos de corte de peso mínimo, que son los conjuntos más pequeños que tienen la propiedad de conjunto de corte. Hay tres dificultades con los conjuntos de corte en el contexto del PAC. En primer lugar, no se puede especificar el número de nodos en el conjunto y buscar a continuación el conjunto de ese tamaño que hace el mejor trabajo (más bien se establece el indicador de éxito y ya podemos buscar el conjunto más pequeño que logra ese nivel de éxito). En ese sentido, los enfoques de teoría de grafos resuelven el inverso del problema que tratamos de resolver. En segundo lugar, no se tienen en cuenta las distancias entre los nodos. En tercer lugar, todas las soluciones que aumentan el número de componentes son iguales, incluso si una solución crea sólo dos componentes mientras que otra crea diez.

Para el PAC-2, los conceptos aplicables incluyen las cubiertas de vértices y los conjuntos dominantes. Una cubierta vértice es un conjunto de nodos cuyos miembros son incidentes sobre cada arista en el grafo. Un conjunto dominante es un conjunto de nodos cuyos miembros son adyacentes a todos los demás nodos del grafo⁵. Para nuestros propósitos, éstos son equivalentes y fallan por exactamente las mismas razones que los conjuntos de corte.

⁵ Los teóricos de los grafos difieren en si estos conjuntos se entienden como mínimos o no.

Medición del éxito

Con el fin de resolver de forma óptima el PAC, debemos tener una definición clara de éxito. Propongo que para el PAC-1 hay dos propiedades de la red que queremos maximizar quitando el conjunto-kp: la fragmentación y la distancia. Para el PAC-2, queremos medir el alcance del conjunto-kp hacia la red que le rodea, basándonos en la distancia. Por lo tanto, necesitamos medidas para cada uno de estos conceptos.

Fragmentación

Tal vez la medida más obvia de fragmentación de la red es un recuento del número de componentes. Si el recuento es 1, no hay fragmentación. La fragmentación máxima se produce cuando cada nodo está aislado, dando lugar a tantos componentes como nodos. Por conveniencia, se normaliza el recuento dividiendo por el número de nodos.

$$C = \frac{K}{n}$$

Ecuación 1

El problema de esta medida es que no tiene en cuenta el tamaño de los componentes. Por ejemplo, en la Figura 2, la supresión del nodo m rompería la red en dos componentes, pero la gran mayoría de los nodos permanecen conectados. Por el contrario, la eliminación del nodo i también daría lugar a sólo dos componentes, pero son más los pares de nodos que estarían desconectados.

Esto sugiere otra medida de fragmentación que simplemente cuenta el número de pares de nodos que están desconectados entre sí. Dada una matriz R en la que $r_{ij} = 1$ si i puede alcanzar j y $r_{ij} = 0$ en caso contrario, podemos definir la nueva medida de la siguiente manera:

$$F = 1 - \frac{2 \sum_i \sum_{j < i} r_{ij}}{n(n-1)}$$

Ecuación 2.

Dado que, por definición, los nodos dentro de un componente son mutuamente alcanzables, la medida F se puede reescribir de manera más económica en términos de los tamaños (S_k) de cada componente (indexados por k):

$$F = 1 - \frac{\sum_k s_k (s_k - 1)}{n(n-1)}$$

Ecuación 3.

La medida F es notablemente similar a una medida de diversidad conocida como heterogeneidad, coeficiente de concentración, índice de Hirschman-Herfindahl, o el índice de Herfindahl. Aplicado al contexto actual, se define como sigue:

$$H = 1 - \sum_k \left(\frac{s_k}{n} \right)^2$$

Ecuación 4.

Una diferencia entre F y H es que mientras ambas alcanzan valores mínimos de 0 cuando la red está formada por un solo componente, cuando la red está máximamente fragmentada (y todos los nodos están aislados) H sólo puede alcanzar $1-1/n$. Si normalizamos H dividiendo por $1-1/n$, obtenemos la medida F (y ver F como una normalización de H nos acerca a una fórmula un poco más rápida de computación).

Un enfoque alternativo es la entropía de información. Aplicado a este contexto, la medida se define como

$$E = - \sum_k \frac{s_k}{n} \ln \left(\frac{s_k}{n} \right)$$

Ecuación 5.

La medida está limitada por abajo a cero, pero no está acotada por arriba. Podemos hacerlo dividiéndola por su valor cuando todos los nodos están aislados:

$$E = \frac{\sum_k \frac{s_k}{n} \ln \left(\frac{s_k}{n} \right)}{\sum_k \ln \left(\frac{s_k}{n} \right)}$$

Ecuación 6.

Distancia

Mientras que la medida de fragmentación F es muy satisfactorio por lo que hace, no tiene en cuenta la forma -la estructura interna- de los componentes. Una red que se divide en dos componentes de tamaño 5 en la que cada componente es una camarilla (Figura 4a) se considera igual de fragmentada que una red dividida en dos componentes de tamaño 5 en la que cada componente es una línea (Figura 4b). Sin embargo, las distancias y por lo tanto los tiempos de transmisión son mucho más altos en la segunda red. Además, si sólo somos capaces de eliminar tantos nodos, puede que no haya una selección de nodos cuya eliminación desconecte el grafo. En tales casos, aún quisiéramos tener alguna manera de evaluar qué sistemas son mejores que otros.

Una solución obvia sería medir la distancia total entre todos los pares de nodos de la red. Sin embargo, esto sólo funciona en caso de que el gráfico permanezca conectado. De lo contrario, hay que sumar distancias infinitas. Una alternativa práctica consiste en basar la medida en la suma de los recíprocos de las distancias, asumiendo la convención de que el recíproco de infinito es cero. En ese caso, podemos crear una versión de F , basado en la ecuación 2, que pondera en función de las distancias recíprocas.

$${}^D F = 1 - \frac{2 \sum_{i>j} \frac{1}{d_{ij}}}{n(n-1)}$$

Ecuación 7.

La medida ${}^D F$ es idéntica a F cuando todos los componentes son completos (es decir, cada componente es también una camarilla). Sin embargo, cuando las distancias dentro de los componentes son mayores que 1, este indicador captura la cohesión relativa de los componentes. Por ejemplo, el grafo de la Figura 4a tiene dos componentes de tamaño 5 y ${}^D F$ es 0.556. El grafo de la figura 4b, que es menos cohesivo, también tiene dos componentes de tamaño 5, pero ${}^D F$ es igual a 0,715, lo que indica mucho menos cohesión. Al igual que la medida F , ${}^D F$ alcanza su valor máximo de 1,0 cuando el grafo está compuesto en su totalidad por nodos aislados.

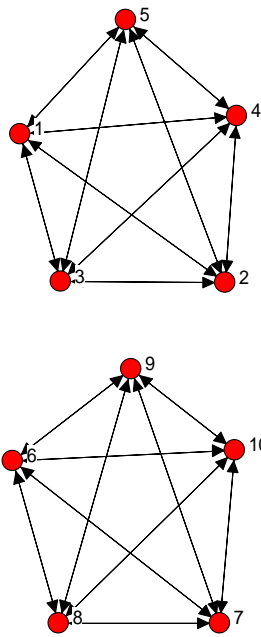


Figura 4a. $D^pF = 0.556$

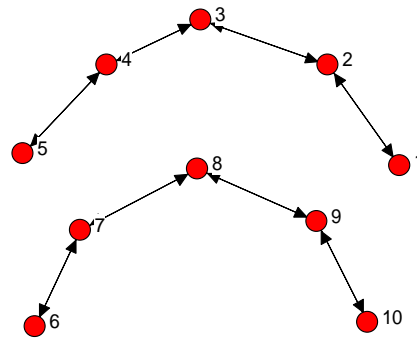


Figura 4b. $D^pF = 0.715$

Alcance basado en la distancia

Las medidas que hemos discutido para el PAC-1 son indicadores a nivel de grafo que aplicamos a un conjunto-k mediante la eliminación del conjunto y la medición de la fragmentación en el grafo restante. Para el PAC-2, buscamos un conjunto de nodos que, como grupo, se conecta al máximo a todos los demás nodos. Por lo tanto, necesitamos una medida directa de la centralidad del conjunto-kp como grupo. El concepto de centralidad de grupo ya ha sido elaborado por Everett y Borgatti (1999), pero sólo se han discutido indicadores de centralidad de grado, cercanía, intermediación y vector propio (*eigenvector*). Como se señaló anteriormente, estas medidas no son óptimas para el problema PAC-2. Por lo tanto, debemos desarrollar otras nuevas basadas en el concepto de alcance.

El indicador de alcance de grupo más sencillo, denominado *m*-alcance, es un recuento del número de nodos únicos alcanzados por cualquier miembro del conjunto-kp en *m* vínculos o menos. La ventaja de esta medida es su validez aparente. La desventaja es que asume que todos los caminos de longitud *m* o menos tienen la misma importancia (cuando en realidad un camino de longitud 1 es probable que sea más importante que un camino de longitud 2) y que todas las rutas de más de *m* son totalmente irrelevantes.

Una medida más sensible, llamada alcance de distancia ponderada, se puede definir como la suma de los recíprocos de las distancias desde el conjunto-kp S a todos los nodos (ver Ecuación 8). Como señalaron Everett y Borgatti (1999), la distancia desde un conjunto a un nodo fuera del conjunto puede ser definida de manera útil de diferentes modos, tales como tomar la distancia máxima desde cualquier miembro del conjunto al nodo exterior, tomar la distancia media, o tomar la distancia mínima. Para el PAC-2, la distancia mínima es apropiada, ya que el hecho de que la distancia a un nodo externo pudiera ser grande para un miembro determinado del conjunto normalmente será irrelevante.

$${}^D R = \frac{\sum_j \frac{1}{d_{sj}}}{n}$$

Ecuación 8.

En la ecuación, la distancia desde el conjunto-kp S al nodo j se indica por d_{sj} . Además, cabe señalar que la suma es a través de todos los nodos y la distancia desde el conjunto a un nodo dentro del conjunto se define como 1. Como antes, el recíproco de una distancia infinita se define como 0. Tomándonos una licencia interpretativa, podemos entender ${}^D R$ como la proporción de todos los nodos alcanzados por el conjunto, donde los nodos son ponderados por su distancia desde el conjunto y sólo a los nodos a distancia 1 se les da el peso completo. Por lo tanto, ${}^D R$ alcanza un valor máximo de 1 cuando cada nodo exterior es adyacente a al menos un miembro del conjunto-kp (es decir, el conjunto-kp es un conjunto dominante). El valor mínimo de 0 se logra cuando cada nodo está aislado.

Selección de un conjunto-KP

Para los conjuntos-kp de tamaño 1, las medidas presentadas anteriormente pueden utilizarse sin rodeos para seleccionar actores clave con sólo elegir el que tenga la puntuación más alta en cualquier medida dada. Por lo tanto, pueden ser consideradas como nuevas medidas de centralidad de los nodos que se optimizan para el problema del actor clave.

Para los conjuntos-kp de tamaño $k > 1$, sin embargo, no existe un procedimiento simple para la elección de un conjunto óptimo. Algunos procedimientos heurísticos pueden ser de valor. Por ejemplo, para el PAC-2, empezamos seleccionando el nodo con la mayor puntuación ${}^D R$. A continuación, para cada una de selecciones $k-1$ restantes, elegimos el nodo con la puntuación más alta que no sea adyacente a

ninguno de los nodos ya seleccionados. Este algoritmo, una variante del primer algoritmo de *bin-packing*⁶, es rápido y fácil, pero a menudo da soluciones que están lejos de ser óptimas.

Se pueden construir otras aproximaciones heurísticas específicas para el PAC, pero el hecho de disponer de claras medidas de éxito que se calculan fácilmente recomienda un algoritmo de optimización combinatoria genérico, como búsqueda-tabu (Glover, 1986), KL (Kernighan y Lin, 1970), recocido simulado (Metropolis et al, 1953) o algoritmos genéticos (Holland, 1975). Los experimentos iniciales sugieren que todos ellos hacen un trabajo excelente para el PAC, de modo que sólo presento un algoritmo voraz simple. La Figura 5 describe el método, que se repite normalmente utilizando docenas de conjuntos de partida aleatorios.

1. Seleccione k nodos k al azar para poblar el conjunto S
2. Establezca el ajuste F = utilizando una métrica PAC adecuada
3. Para cada nodo u en S y cada nodo v no en S
 - a. DELTAF = mejora en ajuste si u y v fueron intercambiadas
4. Seleccione el par con mayor DELTAF
 - a. Si DELTAF <= entonces terminar
 - b. Si no, intercambie por el par con la mayor mejora en el ajuste y establezca
 $F = F + DELTAF$
5. Vaya al paso 3

⁶ Se trata de un Problema en el que objetos de diferentes volúmenes deben ser embalados en un número finito de recipientes o contenedores, de manera que se minimicen el número de contenedores utilizados [Nota del Traductor].

Ensayos empíricos

El funcionamiento del algoritmo se ilustra usando dos bases de datos del ámbito de la salud pública (el SIDA) y del contexto militar (el terrorismo). Ambos casos son abordados desde el punto de vista del PAC-1 y del PAC-2.

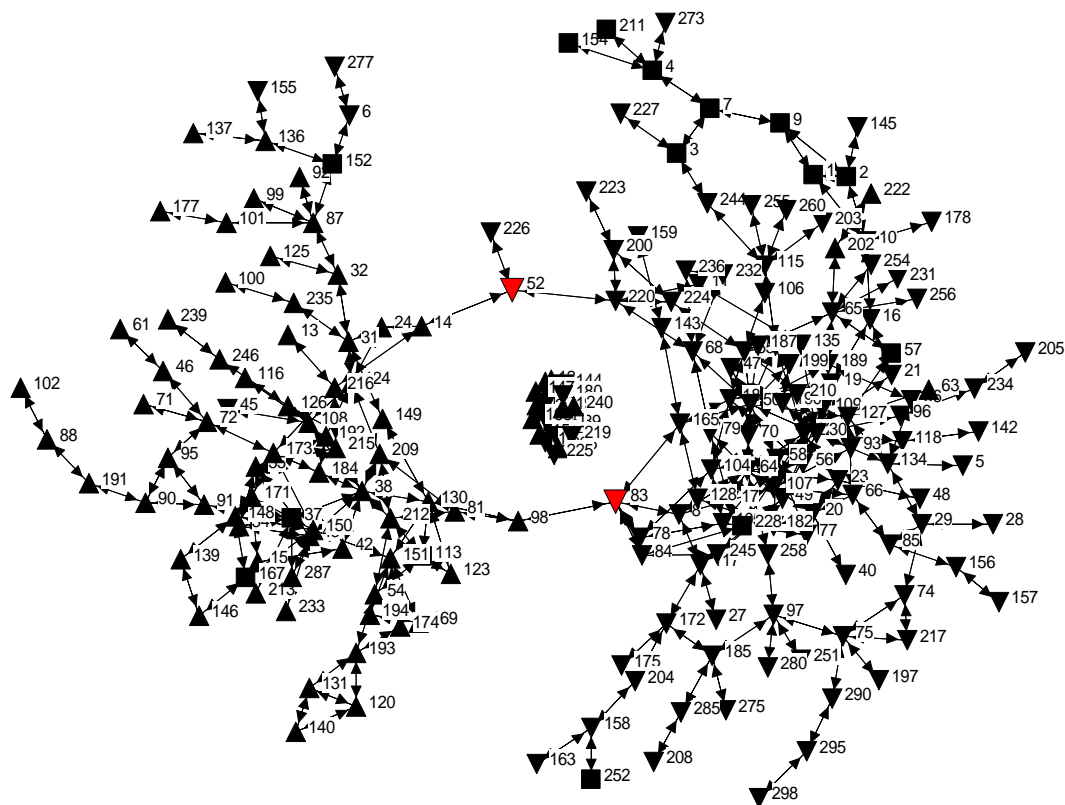


Figura 6. Red de conocidos. Los triángulos ascendentes indican a los afroamericanos, los triángulos invertidos señalan a los puertorriqueños, y con los cuadrados se identifican todos los demás.

Ejemplo sobre el SIDA

La base de datos del SIDA consiste en una red de relaciones entre los 293 usuarios de drogas inyectables en las calles de Hartford, CT. Los datos se describen en Weeks et al (2002). La red consta de un gran componente principal (193 nodos), y muchos componentes muy pequeños. Como se muestra en la Figura 6, el principal componente de la red tiene una estructura muy clara. Se compone de dos grupos, uno afroamericano (con una mayor proporción de VIH+), y el otro en gran parte puertorriqueño (con menor proporción de VIH+). La conexión entre los dos grupos se limita a unos pocos conocidos, conformando un cuello de botella que ayuda a mantener una tasa más baja de VIH+ en la parte puertorriqueña de la red. Ya sea a

través de la inmunización (si es que es posible) o la cuarentena, está claro que una cosa que querríamos hacer desde el principio es aislar a los dos grupos entre sí, tanto porque queremos mantener baja la tasa de VIH+ en el grupo puertorriqueño, como por el hecho de que el riesgo es una función del tamaño del componente del que un nodo forma parte. Por lo tanto, tenemos un caso de PAC-1.

La red proporciona una primera prueba útil del algoritmo de optimización de actores clave, por dos razones. En primer lugar, la estructura de la red hace que sea muy vulnerable a la desconexión, y resulta fácil comprobar los resultados visualmente. Si el algoritmo no pasa esta prueba, es evidente que no es el adecuado. En segundo lugar, la red ya está fragmentada, aportando un ruido que podría confundir a algunos algoritmos.

Basándonos en la inspección visual, está claro que la inmunización o la cuarentena de sólo dos nodos separaría al componente principal en dos mitades casi iguales. Así que para la primera ejecución del algoritmo buscamos un conjunto-kp de tamaño 2. El índice de fragmentación de partida para el gráfico es 0.567. El algoritmo seleccionó dos nodos, identificados en negro en la figura 6, que, si se suprimen, aumentarían la fragmentación hasta 0.658 (un aumento proporcional en la fragmentación del 16%). Los nodos seleccionados se ajustan a nuestra intuición y dividen el componente principal en dos grandes bloques.

Pasando al PAC-2, también estamos interesados en la selección de un pequeño grupo de nodos para que sean objeto de intervención -en concreto, para ser capacitados como educadores de iguales (conocidos como "agentes de salud") para transmitir y enseñar las prácticas de prevención del VIH. Weeks et al (2002) hicieron esto a mano, seleccionando laboriosamente el grupo más pequeño que podría alcanzar a más de 50% del componente principal de la red. El conjunto ganador contenía 14 nodos. Aplicando el algoritmo *KeyPlayer* a los mismos datos, se obtiene el mismo resultado, como se muestra en la Tabla 1.

Como era de esperar, el número de personas alcanzadas aumenta en relación con la potencia fraccionaria del tamaño del grupo, pero una consideración más completa de este fenómeno está fuera del alcance de este documento.

Tamaño del grupo	Número alcanzado	Porcentaje alcanzado
1	16	8.3
2	27	14.0
3	36	18.7
4	43	22.3
5	49	25.4
6	55	28.5
7	61	31.6
8	67	34.7
9	72	37.3
10	77	39.9
11	82	42.5
12	87	45.1
13	92	47.7
14	97	50.3

Tabla 1.

Ejemplo sobre el terrorismo

La base de datos sobre terrorismo, elaborada por Krebs (2001), consiste en una red de supuesto conocimiento entre 74 sospechosos de terrorismo. Para los fines de este análisis, se utiliza sólo el componente principal, que consiste en 63 individuos.

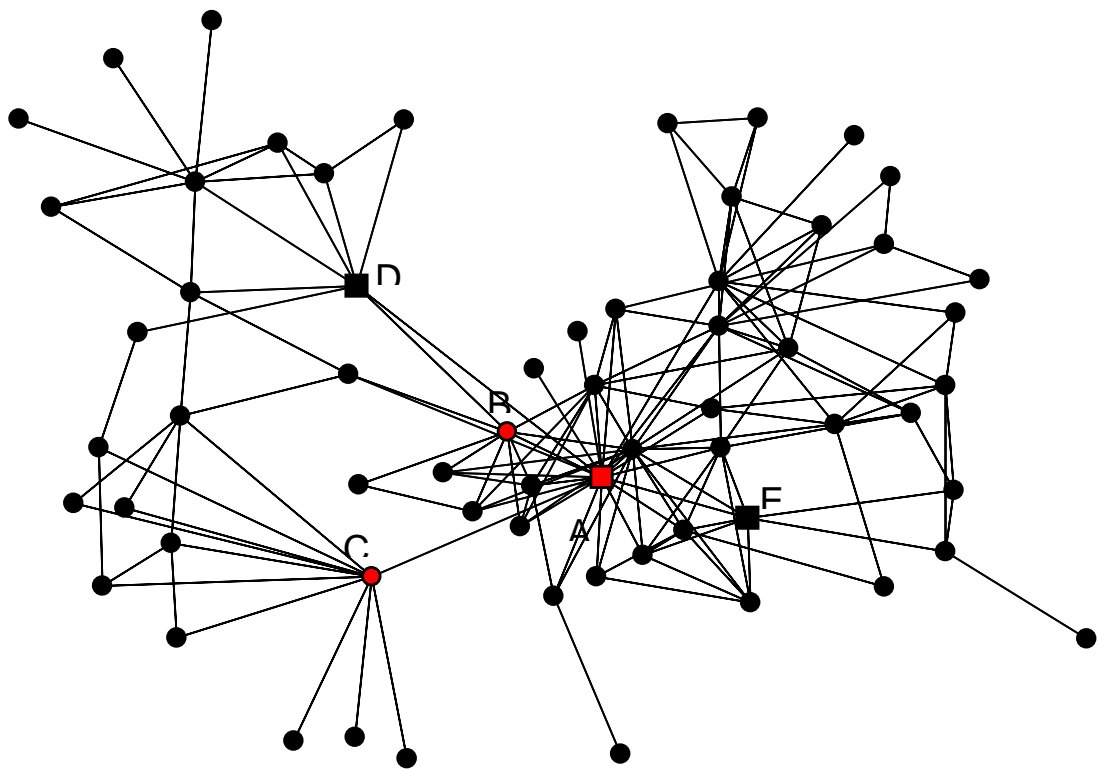


Figura 7. Red de terroristas elaborada por Krebs (2002).

La primera pregunta que nos hacemos es qué personas deben ser aisladas con el fin de perturbar al máximo la red. Supongamos que sólo podemos aislar tres personas. La ejecución del programa *KeyPlayer* selecciona los tres nodos identificados en rojo en la figura 7 (los nodos A, B y C). La eliminación de estos nodos arroja un indicador de fragmentación de 0,59, y rompe el gráfico en 7 componentes (incluyendo dos grandes que comprenden las mitades izquierda y derecha de la gráfica).

La segunda pregunta que nos hacemos es, teniendo en cuenta que nos gustaría difundir cierta información, qué nodos querríamos que estuviesen expuestos a la información con el fin de alcanzar potencialmente a todos los demás nodos rápidamente y con seguridad. Supongamos que la información que viaja a través de más de dos enlaces tiende a degradarse o a ser visto con sospecha. Por lo tanto queremos que el conjunto más pequeño de nodos alcance a los demás dentro de dos eslabones o menos. El algoritmo *KeyPlayer* encuentra que un conjunto de tres nodos (los nodos cuadrados en la Figura 7, con la etiqueta A, C y D) alcanza el 100% de la red (incluidos ellos mismos).

Discusión

En este trabajo se ha definido el problema *KeyPlayer* y demostramos por qué un enfoque ingenuo basado en la centralidad, al igual que otras aproximaciones más sofisticadas de la teoría de grafos, no resuelven el problema. Hemos introducido nuevas métricas para medir el éxito, e implementado un algoritmo de optimización combinatoria para maximizar estas cantidades. Mostramos las aplicaciones en el ámbito de la salud y en el contexto militar.

Las métricas para medir el éxito en el PAC-1 son esencialmente indicadores de la cohesión del grafo, que pueden ser útiles de manera descriptiva en diversos contextos, además de para el problema del actor clave. Las aplicaciones típicas pueden ser la comparación entre organizaciones, o el uso de la cohesión como un predictor de rendimiento del grupo.

El hecho de que el PAC-1 y el PAC-2 tengan diferentes soluciones es interesante. Para muchos, la centralidad es o bien un concepto unitario, con muchas medidas altamente correlacionadas, o un concepto totalmente multidimensional en el que cada medida indica un tipo diferente de centralidad. Me permito sugerir una visión intermedia que divide las nociones de importancia nodal en dos tipos básicos, que corresponden a los problemas PAC-1 y PAC-2. Las medidas del PAC-1 cuantifican la medida en que la cohesión de un grafo se reduce por la eliminación de ese nodo, mientras que los indicadores del PAC-2 cuantifican la medida en la que los lazos de un nodo llegan a la red. El segundo tipo es el que está más directamente relacionado con la centralidad: un nodo que está bien conectado. El primer tipo no trata tanto de la buena conectividad del nodo como de la conexión del resto del grafo en ausencia del nodo. Desde el punto de vista de los grafos, existe cierta analogía con la distinción en el condicionamiento operante entre el refuerzo positivo (PAC-2) y el refuerzo negativo (PAC-1): el refuerzo positivo proporciona un beneficio (la conectividad), aportando directamente un don (las conexiones del nodo), mientras que el refuerzo negativo proporciona beneficios al aliviar un factor de estrés (el nodo mantiene unido un grafo que de otro modo estaría fragmentado). Desde el punto de vista de los nodos, un nodo alcanza sus valores más altos en los indicadores del PAC-2 cuando el grafo es altamente cohesivo. Por el contrario, los valores altos en las medidas del PAC-1 sólo ocurren normalmente cuando el grafo no es muy cohesivo. Actores con puntuaciones altas en las medidas de PAC-2 contribuyen a maximizar la utilización de los recursos que fluyen a través de la red, mientras que los actores que puntúan alto en las medidas de PAC-1

contribuyen a la maximización de los beneficios de la intermediación, el control de acceso y la oposición entre actores.

Limitaciones y próximos pasos

Hay dimensiones significativas del problema que he ignorado. Tal vez la más importante es la cuestión de la calidad de datos. Si este enfoque pretende producir una herramienta práctica, no podemos asumir sin más datos perfectos. Más bien, el método debe ser robusto al enfrentarse a errores en los datos. Se estudiarán dos enfoques. En primer lugar, tenemos la noción de no optimizar demasiado cerca de los datos observados. Si se sabe que los datos varían en una magnitud dada (por ejemplo, el 10% de las relaciones observadas no existen en realidad y el 10% de las parejas no adyacentes observadas son de hecho adyacentes), entonces podemos variar aleatoriamente los datos en esta magnitud y realizar la optimización a través de un conjunto de conjuntos de datos "adyacentes" obtenidos de esta manera. El resultado es un conjunto-kp que no es necesariamente óptimo para el conjunto de datos observados, pero representará una solución de alta calidad para el barrio del grafo en su conjunto.

Un enfoque alternativo consiste en tratar el conocimiento de los vínculos como probabilístico, modificando las métricas de *KeyPlayer* en consecuencia. Por ejemplo, supongamos que sabíamos, para cada par de nodos, la probabilidad independiente de que exista un lazo entre ellos. Entonces, en principio, podríamos trabajar con la distancia esperada (incluyendo el infinito) entre los nodos a través de todas las redes posibles⁷. Los indicadores de PAC basados en la distancia y la accesibilidad podrían ser computados sustituyendo la distancia esperada por la distancia observada. El desafío práctico consiste en encontrar fórmulas de acceso directo para la distancia esperada y la conectividad que permitan el cálculo rápido.

Otra cuestión se refiere a la utilización de la distancia geodésica. Como señalamos en otra ocasión (Borgatti, 2002), diferentes tipos de procesos de flujos tienen trayectorias características. Por ejemplo, las infecciones que inmunizan (o matan) al anfitrión no vuelven a los nodos que han infectado previamente. Por lo tanto, se

⁷ Téngase en cuenta que el problema que se aborda es la seguridad en los valores observados, no la existencia de lazos. Se supone en este enfoque que los lazos son fijos y no emergen de manera probabilística como una función de los atributos del nodo o de otros lazos. La naturaleza dinámica de las relaciones es un fenómeno diferente, que requiere modelos propios.

desplazan a lo largo de los caminos teóricos de los grafos. Por el contrario, los cotilleos que se transmiten a través del correo electrónico pueden llegar al mismo nodo varias veces, pero en general no desde las mismas fuentes (es decir, alguien no suele contar la misma historia confidencial a la misma persona más de una vez). Por lo tanto, las historias viajan a través de los senderos. Ni las infecciones ni las historias viajan necesariamente a través de los caminos más cortos a cada nodo. En consecuencia, para esos procesos, la distancia esperada recorrida por algo que fluye a través de la red no es igual a la distancia geodésica, y tendría sentido sustituir esta otra distancia esperada en las ecuaciones.

Para simplificar la exposición, este artículo se ha basado en grafos simples no dirigidos con aristas no valoradas. La extensión a gráficos dirigidos es sencillo, pero la extensión a aristas valoradas requerirá algo de desarrollo. Una excepción es la clase de medidas de fragmentación, que se generalizan muy bien a lo largo de las líneas de conjuntos de corte mínimos ponderados, perdiendo sólo los métodos abreviados de cálculo disponibles con datos no valorados. Las métricas basadas en la distancia requerirán diferentes generalizaciones en función de los tipos de valores, lo que podría ir desde distancias a las capacidades, y de las capacidades a las probabilidades de transmisión. Las relaciones múltiples, registradas en forma de gráficos separados con un conjunto de nodos comunes, pueden manejarse mediante la suma de los criterios de éxito en todas las relaciones.

Finalmente, puede ser de interés en el futuro incorporar los atributos del actor. En el contexto militar, la comunicación entre los actores con habilidades redundantes puede a veces ser menos importante que la comunicación entre los actores con habilidades complementarias. En el contexto de la salud pública, es útil reducir al mínimo la relación entre poblaciones diferentes para frenar las epidemias (por ejemplo, cuando las mujeres casadas están vinculadas a las prostitutas a través de sus maridos). Por lo tanto, un criterio adicional que habría que considerar en la fragmentación de una red es maximizar la separación de los actores con determinados atributos.

Referencias

Borgatti, S.P. 2002. Stopping terrorist networks: Can social network analysis really contribute? *Sunbelt International Social Networks Conference*. 13-17 February. New Orleans.

- Everett, M. G., & Borgatti, S. P. 1999. The centrality of groups and classes. *Journal of Mathematical Sociology*. 23(3): 181-201
- Glover F., 1986, Future paths for integer programming and links to artificial intelligence, *Computers and Operations Research*, 5: 533-549.
- Holland, J. 1975. *Adaptation in Natural and Artificial Systems*. University of Michigan Press.
- Kernighan, B.W., and Lin, S. 1970. Efficient heuristic procedure for partitioning graphs. *Bell Systems Technical Journal*. 49(2): 291-297.
- Krebs, V. 2002. Uncloaking terrorist networks. *First Monday* 7(4): April. http://www.firstmonday.dk/issues/issue7_4/krebs/index.html
- Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, E. Teller. 1953. Equation of state calculations by fast computing machines", *J. Chem. Phys.*, 21, 6, 1087-1092.
- Weeks, M.R., Clair, S., Borgatti, S.P., Radda, K., and Schensul, J.J. 2002. Social networks of drug users in high risk sites: Finding the connections. *AIDS and Behavior* 6(2): 193-206.