

Confianza, Influencia, y Redes

Laboratorio 6 – Identificación de subgrupos

Los científicos sociales generalmente asumen que "la interacción social es la base de la solidaridad, las normas compartidas, la identidad y el comportamiento colectivo, por lo que es probable que las personas que interactúan intensamente se consideren parte de un grupo social". Por lo tanto, una herramienta principal del análisis de redes sociales es identificar grupos densos de actores "entre los cuales hay vínculos relativamente fuertes, directos, intensos y / o positivos". Por lo general, se denominan "subgrupos cohesivos", "subredes" o "subgrupos". Una forma de agrupar a los actores se basa en atributos compartidos (por ejemplo, raza, género, etc.). Otro es utilizar el patrón de vínculos entre actores.

En un mundo ideal, habría un solo algoritmo para identificar subgrupos cohesivos, pero este no es un mundo ideal, por lo que los analistas de redes sociales han desarrollado una variedad de algoritmos para identificar subredes. No los consideraremos todos aquí; simplemente hay demasiados. En cambio, nos centraremos en algunos. En Gephi consideraremos componentes, k-cores y algoritmos de detección de comunidades; y en snExplorer veremos brevemente los recuentos de componentes, k-núcleos y camarillas.

Para este laboratorio vamos a necesitar el modulo *Newman-Girvan clustering* antes de comenzar.

Parte I – Identificación de subgrupos en Gephi

[Gephi]
Archivo>Abrir

1. Abra Gephi y lea el archivo "Alive Trust Network.gephi". Comenzaremos explorando componentes, estos son subgrupos de actores que están conectados (ya sea directa o indirectamente) entre sí. En las redes dirigidas, puede identificar dos tipos de componentes: fuertes y débiles. Los componentes fuertes toman en cuenta la dirección de los lazos, mientras que los componentes débiles no. Con redes no dirigidas, solo puede identificar componentes débiles.

[Vista general]
Estadísticas
>Componentes conexos
>Ejecutar

2. Para identificar componentes en Gephi, haga clic en "Ejecutar" junto a la función "Componentes conexos" en la pestaña "Estadísticas". Aparecerá un cuadro de diálogo titulado "Parámetros de Componentes Conectados" (no se muestra) que ofrece la opción de indicar si nuestros datos son "Dirigidos" o "No dirigidos". Aquí, estamos trabajando con datos no dirigidos, así que elija esta opción y haga clic en "Aceptar". Ahora deberá aparecer el informe "Connected Components Report" (Reporte de Componentes conectados) (Figura 1), que informa los parámetros de la red, el número de componentes "Results" (Resultados) y un gráfico que indica el número y tamaño de cada uno de los componentes. Como puede ver, Gephi encontró 8 componentes, pero al comparar esta cifra con el gráfico de red, podemos ver que 7 de estos son nodos aislados. Igualmente podemos llegar a esta conclusión examinando el gráfico en el informe. Si observa verá que hay 7 componentes de tamaño 1 y un componente de un poco más grande que 60 (62 en realidad). Cierre la ventana del informe cuando haya terminado de observar el resultado.

Confianza, Influencia, y Redes

Laboratorio 6 – Identificación de subgrupos

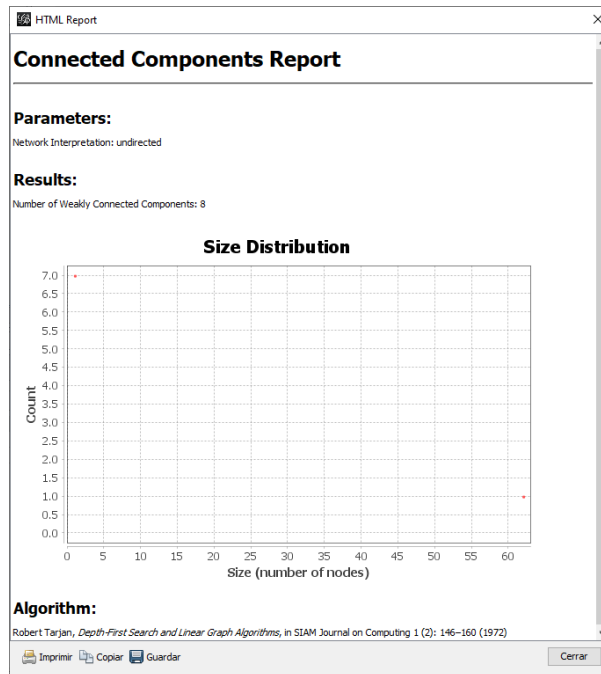


Figura 1: Reporte de Componentes Conectados en Gephi

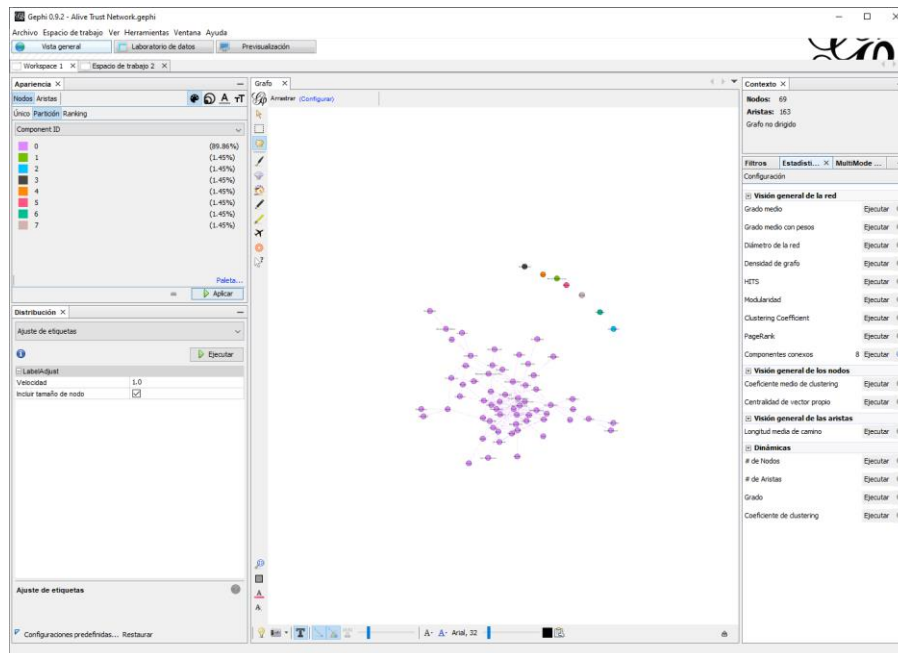


Figura 2: Red de confianza con actores vivos, color por componente

[Laboratorio de datos]
Tabla de datos>Nodos

- También puede ver los resultados navegando al "Laboratorio de datos" y buscando una nueva columna (partición) en la sección "Nodos" denominada "Component ID" (identificación del componente). Gephi genera una serie de números, comenzando con 0, que identifican los componentes respectivos de cada actor. También podemos usar esta variable categórica para visualizar la red, seleccionando primero la opción *Nodos> Color> Partición* en la pestaña "Apariencia", y luego con el menú desplegable "--Escoge un atributo", seleccionar "Component ID" y haga clic en

Confianza, Influencia, y Redes

Laboratorio 6 – Identificación de subgrupos

[Vista general]

Apariencia>Nodos
Color>Partición

--Escoja un atributo
>Component ID>Aplicar

[Vista general]

Filtros>Atributos>Partición

Queries>Partition>Filter
>Export filtered graph to a
new workspace

“Aplicar”. Sin embargo, el resultado no es muy interesante (Figura 2) ya que cada nodo aislado posee un color diferente, mientras que los actores conectados en el componente más grande (llamado componente principal) son todos del mismo color.

- No obstante, puede que algunas ocasiones sea útil identificar el componente principal. De hecho, no es raro que los investigadores extraigan el componente principal y lo analicen por separado. Por lo tanto, exportaremos el componente principal a un espacio de trabajo separado donde lo analizaremos por separado. Para hacer esto, en la pestaña "Filtros" a la derecha de la ventana "Vista general", primero seleccione "Atributos" y luego "Partición". A continuación, arrastre la partición "Component ID" a "Consultas". En las opciones abajo, haga clic en el cuadro que tiene el mayor porcentaje de nodos (probablemente "0"), haga clic en "Filtrar" y luego use el botón "Exportar el grafo filtrado en un nuevo espacio de trabajo " (Figura 3). **¿Cuál es el tamaño del componente principal? ¿Cuál es su densidad? ¿Grado medio?**

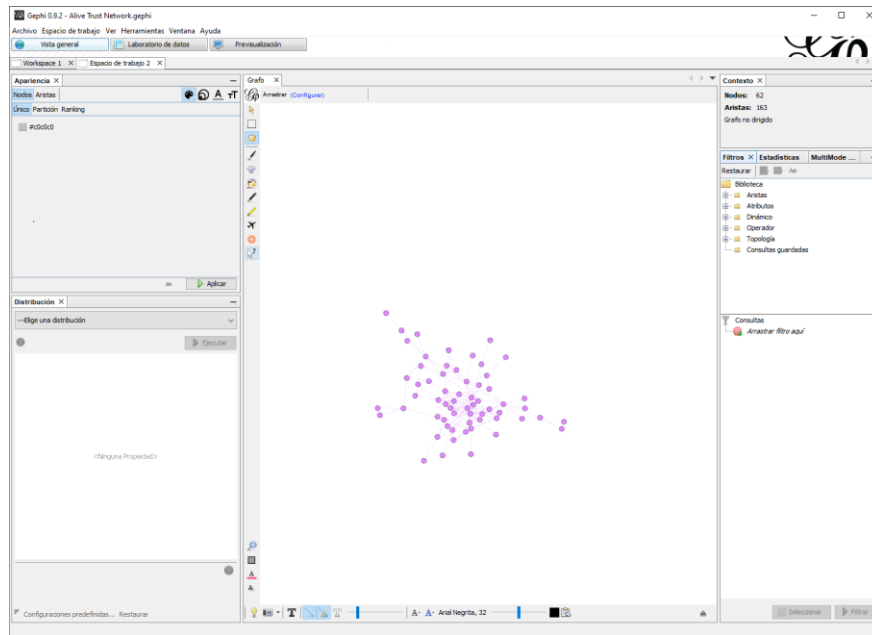


Figura 3: Red de confianza con actores vivos, componente principal

- Veamos si identificar k -cores es una mejor estrategia para identificar subgrupos cohesivos para esta red en particular. La "k" en k-core indica el número mínimo de vínculos de cada actor dentro del núcleo (“core”); por ejemplo, un núcleo 2 incluye a todos los actores que tienen dos o más conexiones con otros actores dentro del núcleo, y un núcleo 3 incluye todos los actores que tienen tres o más vínculos con otros actores dentro de un núcleo. Es importante tener en cuenta que el k-core más alto de una red no corresponde necesariamente al puntaje de centralidad de grado más alto obtenido por un actor en la red. Más bien, es el subgrupo en el que cada actor comparte el mayor número de vínculos. Imagínese una red en la que un actor tiene

Confianza, Influencia, y Redes

Laboratorio 6 – Identificación de subgrupos

seis vínculos, pero nadie más tiene más de cuatro. El k-core más alto no sería un 6-core porque todos los demás en el subgrupo también necesitarían tener seis enlaces.

Filtros > Topología
> K-core

6. Regrese a la pestaña *Alive Trust Network*. Una vez más usaremos un filtro para encontrar k-cores. Esta vez, arrastre la opción "K-core" que se encuentra sección de "Topología" en la pestaña "Filtro" a "Consultas". Hacia la parte inferior de la pestaña "Filtros", debería ver un cuadro "K-core Configuración" que estará predeterminado en "1". Haga clic en el botón "Filtrar" y los aislados deberán desaparecer. Aumente la configuración a "2" y no debería haber ningún cambio. Eso es porque Gephi aparentemente cuenta los lazos dos veces, para reflejar las dos direcciones de un enlace aunque, en este caso, Gephi sabe (o debería saber) que se trata de una red no dirigida. Por lo tanto, el filtro de 1 y 2 capturan el primer núcleo ($k = 1$), 3 y 4 capturan el segundo núcleo ($k = 2$), y así sucesivamente. Si sigue aumentando la configuración, todos los nodos desaparecen al llegar a quince, lo que significa que el núcleo k más alto de la red es el séptimo núcleo ($14/2$). Vea la Figura 4. **Exporte el 7-core a un nuevo espacio de trabajo. ¿Qué actores parecen ser los más centrales en la red del séptimo núcleo ($k = 7$)?**

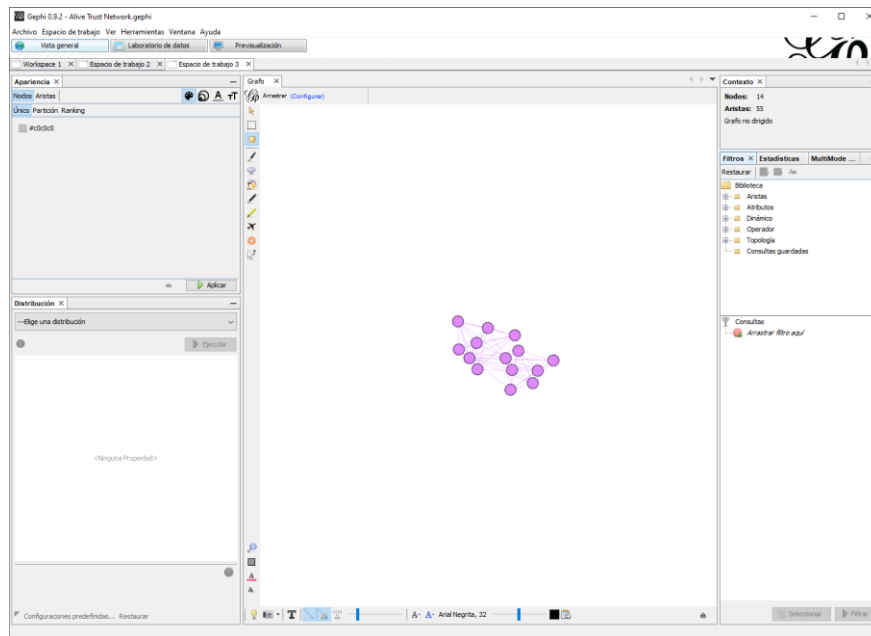


Figura 4: Red de confianza con actores vivos, 7-core

7. Ahora consideremos otro método para identificar subgrupos: algoritmos de detección de comunidades. Consideraremos dos: Girvan-Newman y Louvain. El segundo está implementado en Gephi, pero Girvan-Newman no, por lo que deberá descargar e instalar el módulo Newman-Girvan Clustering, si aún no lo ha hecho. En la pestaña "Estadísticas" de *Alive Trust Network*, haga clic en el botón "Ejecutar" a un lado de función "Girvan-Newman Clustering". Esto abrirá un cuadro de diálogo (no se muestra). Acepte sus valores predeterminados y haga clic en "Aceptar". Esto

Confianza, Influencia, y Redes

Laboratorio 6 – Identificación de subgrupos

- [Vista general]
Estadísticas>Girvan-Newman
Clustering>Ejecutar
- Apariencia>Nodos
Color>Partición
--Escoge un atributo
>Cluter-ID>Aplicar
- Estadísticas>Modularidad>Ejecutar
- Apariencia>Nodos
Color>Partición
--Escoge un atributo
>Modularity Class>Aplicar
- Generate groups by partition
- Archivo>Guardar como...
- Archivo>Exportar
>Archivo de grafo...
- producirá un informe (no se muestra) que indica el número de comunidades que detectó y la modularidad. **¿Cuántas comunidades encontró el algoritmo?**
8. Asimismo, colorea los nodos según la partición creada por el algoritmo (Cluter-ID). Note que el algoritmo Girvan-Newman ha asignado cada uno de los aislados a una comunidad separada. No todos los algoritmos de detección de comunidades tratan los aislamientos de esta manera, por lo que cuando compare los resultados de dos o más de ellos, tenga esto en cuenta.
9. Ahora, repitamos el proceso con el algoritmo de Louvain, que se implementa mediante la función "Modularidad" en la pestaña "Estadísticas". Esto abrirá un cuadro de diálogo (no se muestra). Tenga en cuenta que incluye un parámetro de ajuste ("Resolución"), que le permite ajustar si desea obtener menos o más comunidades. Por ahora, haga clic en "Aceptar". Esto produce un informe (no se muestra). **¿Cuántas comunidades encontró el algoritmo? ¿Qué es la modularidad? ¿Cómo se compara esto con los resultados de Girvan-Newman?**
10. Colorea los nodos usando la partición creada por el algoritmo de Louvain ("Clase de modularidad"). Debido a que hay más de ~8 comunidades, deberá ajustar la paleta de colores como hicimos en laboratorios anteriores. **¿Louvain asigna a cada aislado en una comunidad separada?**
11. Ahora colapse la red usando la herramienta "Generate groups by partition" (Generar grupos por partición). Asegúrese de decirle a Gephi que cree un nuevo espacio de trabajo. Recuerde que los atributos son aspectos no relacionales de las redes. Pueden ser previamente dados (por ejemplo, rol, género, nacionalidad) o pueden ser generados por nuestro análisis de una red. Los algoritmos de agrupación, como los núcleos k y los de detección de comunicación, clasifican actores en varios subgrupos, y la partición de "pertenencia" del subgrupo es un atributo de actor.
12. Visualice la red reducida. **¿Aparece algún subgrupo (o subgrupos) más central que otros?** Proporcione una captura de pantalla.
13. Antes de pasar a la última parte de este ejercicio, necesitamos guardar la red *Alive Trust Network* en archivo Pajek. Para hacer esto, debe seleccionar la pestaña que desea guardar y seguir los siguientes comandos *Archivo>Exportar>Archivo de grafo...* (en la siguiente ventana asegúrese de guardar el archivo como tipo "Archivos NET (Pajek) (*.net)"). Asegúrese de guardar el archivo con un nombre obvio, por ejemplo, la red de amistad debe ser nombrada "red_de_confianza_vivos". Al terminar de guardar ambos archivos, pase al siguiente paso utilizando snExplorer.

Confianza, Influencia, y Redes
Laboratorio 6 – Identificación de subgrupos

Parte II – Identificación de subgrupos en snExplorer

*[snExplorer]
Select an input format
>Use external data

Select an input format
>Pajek>
Undirected>Dismiss

[Network Measures]
Group Level Metrics*

1. Abra snExplorer (https://corelab.nps.edu/apps/sn_explorer/) e inmediatamente seleccione la opción para importar datos de Pajek no dirigidos de un archivo externo (*Select an importing option > Use external data > Select an input format > Pajek (*.net) > Undirected > Browse*). Termine haciendo clic en el botón “Dismiss”.
2. Cambie a la pantalla "Network Measures" (medidas de red). Haga clic en el signo "+" a la derecha de las " Group Level Metrics " y aparecerá una serie de métricas.
¿Cuántos componentes detecta snExplorer? ¿Cuántas camarillas? ¿Cuál es el k-core más alto? ¿Cómo se comparan estos resultados con los de Gephi?