

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, un objetivo 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, estos grupos 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. En este ejercicio, no los consideraremos todos, nos centraremos en los más comunes y disponibles en Gephi.

Específicamente, consideraremos componentes, k-cores y algoritmos de detección de comunidades.

Para este laboratorio vamos a necesitar el módulo *Newman-Girvan clustering*, antes de comenzar asegúrese de instalar esta herramienta.

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 conexos (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 de clic en "Aceptar". Ahora deberá aparecer el informe "Connected Component Report" (Reporte de Componentes conectados) (Figura 1), que presenta 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 grande. Cierre la ventana del informe cuando haya terminado de observar los resultados.

Laboratorio 6 – Identificación de Subgrupos

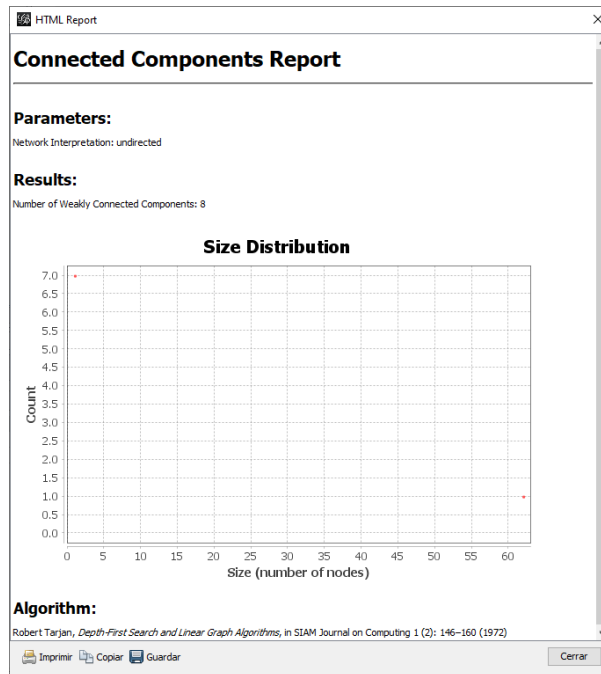


Figura 1: Reporte de Componentes Conectados en Gephi

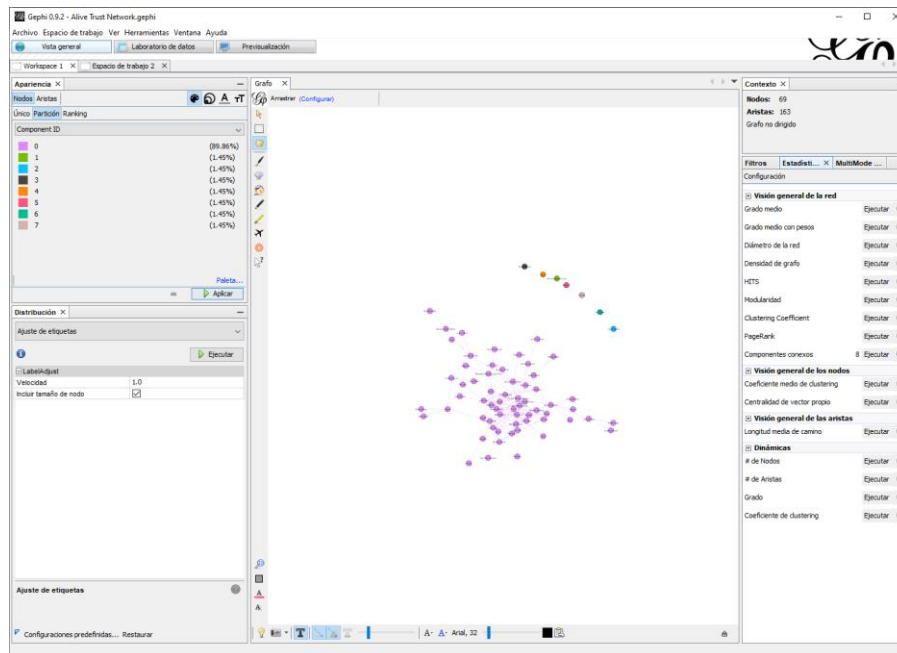


Figura 2: Red de confianza con actores vivos, color por componente

[Laboratorio de datos]
Tabla de datos>Nodos

3. También puede ver los resultados navegando al "Laboratorio de datos" y localizar 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. Podemos usar esta variable categórica para ajustar la visualización de la red. Retorne a "Vista General", seleccione las opciones *Nodos> Color> Partición* en la pestaña "Apariencia", y luego con el menú desplegable "--Escoge un atributo", seleccione "Component ID".

Laboratorio 6 – Identificación de Subgrupos

[Vista general]

Apariencia>Nodos
Color>Partición

--Escoge un atributo
>Component ID>Aplicar

[Vista general]

Filtros>Atributos>Partición

Consultas> Partición
>Filtrar
>Exportar el grafo filtrado
en un nuevo espacio de
trabajo

Concluya dando clic en “Aplicar”. El resultado (Figura 2) consta de nodos aislados de diferentes colores, mientras que los actores conectados en el componente más grande (llamado componente principal) tendrán el mismo color.

- En algunas ocasiones será útil identificar y aislar 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, seleccione el cuadro que corresponde a el mayor porcentaje de nodos (probablemente "0"), presione "Filtrar" y luego use el botón "Exportar el grafo filtrado en un nuevo espacio de trabajo" (Figura 3). Seleccione el nuevo espacio de trabajo generado y responda, **¿Cuál es el tamaño del componente principal? ¿Cuál es su densidad? ¿Grado medio?**

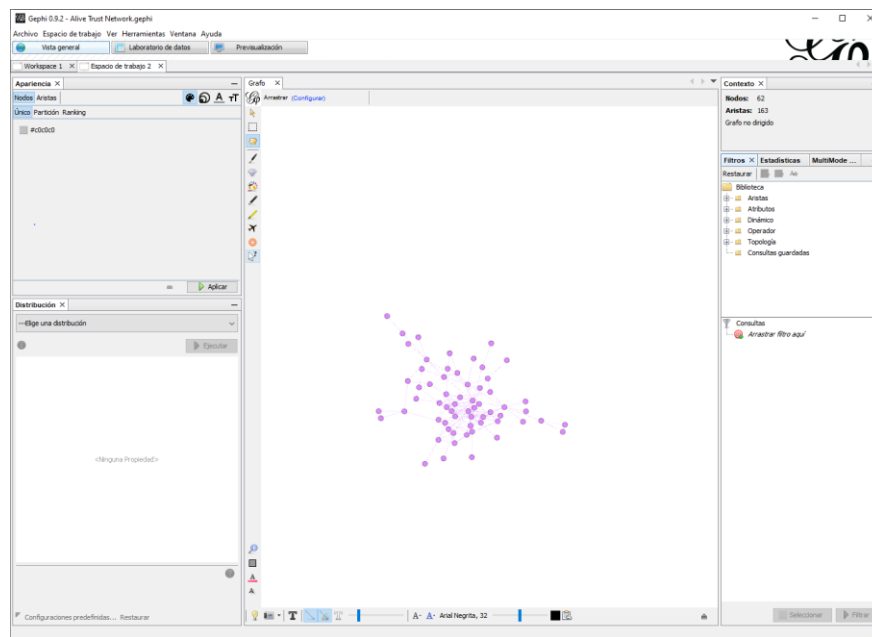


Figura 3: Red de confianza con actores vivos, componente principal

Filtros>Topología
>K-core

- Veamos si identificar *k-cores* es una mejor estrategia para identificar subgrupos cohesivos para esta red en particular. Recuerde que "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

Laboratorio 6 – Identificación de Subgrupos

alto de una red no corresponde necesariamente al puntaje de centralidad de grado más alto obtenido por un actor en la misma. Imagínese una red en la que un actor tiene 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. Regrese a el primer espacio de trabajo. Una vez más usaremos un filtro para encontrar k-cores. Remueva el filtro que utilizamos previamente. Ahora, 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", deberá ver un cuadro titulado "K-core Configuración" con el valor predeterminado de "1". Haga clic en el botón "Filtrar" y los aislados deberán desaparecer. Aumente la configuración a "2" desaparecerán los pendientes. Si sigue aumentando el valor del filtro, todos los nodos desaparecen al llegar a ocho, lo que significa que el núcleo k más alto de la red es el séptimo núcleo (7-core). Vea la Figura 4. Exporte el 7-core a un nuevo espacio de trabajo y responda **¿Qué actores parecen ser los más centrales en la red del séptimo núcleo (k = 7)?**

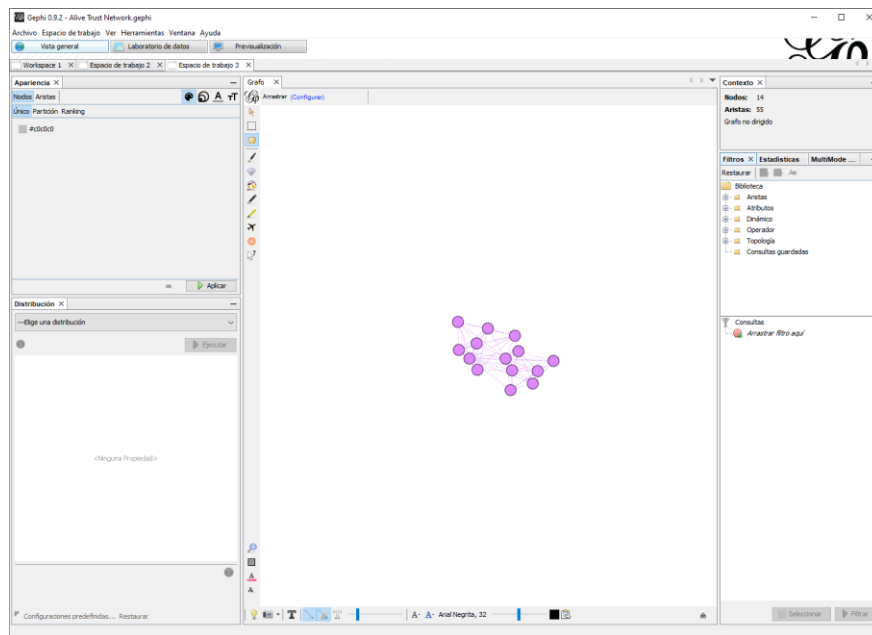


Figura 4: Red de confianza con actores vivos, 7-core

- [Vista general]
- Estadísticas>
Girvan-Newman
Clustering>
Ejecutar
6. Ahora consideremos otro método para identificar subgrupos: algoritmos de detección de comunidades. Consideraremos dos: Girvan-Newman y Louvain. Deberá descargar e instalar el módulo Newman-Girvan Clustering, si no lo ha hecho. Retorne al espacio de trabajo *Alive Trust Network*. En la pestaña "Estadísticas" de 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 de clic en "Aceptar". Esto producirá un informe (no se muestra) que indica el número de comunidades que detectó y la modularidad. **¿Cuántas comunidades encontró el algoritmo?**

Laboratorio 6 – Identificación de Subgrupos

Apariencia>Nodos
Color>Partición
--Escoge un atributo
>Cluter-ID>Aplicar

7. Asimismo, coloree 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 aislados de esta manera.

Estadísticas>Modularidad>
Ejecutar

8. Ahora, repitamos el proceso con el algoritmo de Louvain, que se implementa mediante la función "Modularidad" en la pestaña "Estadísticas". Al seleccionar "Ejecutar" se abrirá un cuadro de diálogo (no se muestra), en este puede modificar la *Resolución*, que le permite ajustar si desea obtener menos o más comunidades. Por ahora, no modifique nada y de clic en "Aceptar". Esto producirá un informe (no se muestra), responda **¿Cuántas comunidades encontró el algoritmo? ¿Qué es la modularidad? ¿Cómo se compara esto con los resultados de Girvan-Newman?**

Apariencia>Nodos
Color>Partición
--Escoge un atributo
>Modularity Class>Aplicar

9. Coloree los nodos usando la partición creada por el algoritmo de Louvain ("Modularity Class"). 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?**