

TALLER 1

Instrucciones:

Este taller es de carácter individual. El taller se enviará el jueves 31 de agosto y tiene como fecha de entrega el jueves 28 de septiembre antes de las 5 p.m. El informe escrito, los scripts de Python y el archivo de Gephi deberán estar comprimidos y deberán ser enviados por SicuaPlus. Este será el único medio de entrega del taller, no se aceptarán talleres fuera de la hora de entrega o por otros medios. Para cualquier tipo de reclamo, en los horarios de atención previstos podrán revisar las calificaciones de sus talleres.

Punto 1 – LAS FAMILIAS EN GAME OF THRONES

Game of Thrones es una serie de drama y fantasía americana producida por la cadena HBO. La serie está basada en la saga A Song of Ice and Fire, escrita por George R.R. Martin. Debido a las intrincadas relaciones que existen dentro del universo GOT usted tendrá el reto de modelar y analizar las redes sociales familiares de la serie. Adjunto encontrará dos archivos .csv con la información de la red:

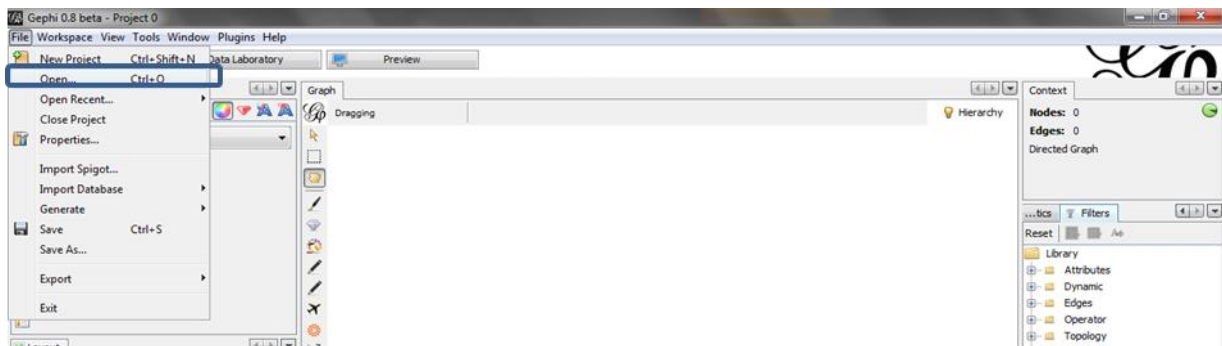
- Nodes_GOT.csv contiene la información de los personajes de la serie: Id (Nombre), Sexo, Cultura, Casa, Popularidad, Segunda casa.
- Edges_GOT.csv contiene la información de las conexiones entre personajes: Source (Origen), Target (Destino) y Tipo de relación (Padre, madre o esposo).

Todos los términos propios del universo GOT con los que no estén familiarizados se encuentran al final del taller.

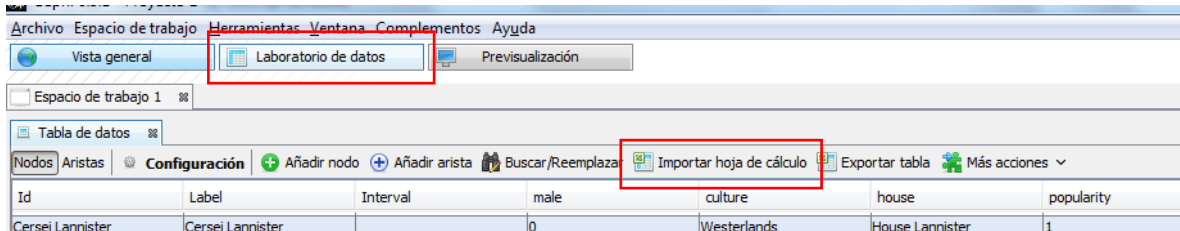
1. Su primera tarea es visualizar la red en Gephi para entender con qué información cuenta, debe seguir los siguientes pasos y concluir con base en los resultados:

Deben abrir un nuevo proyecto en *Gephi 0.9.1* y desde allí abrir la red:

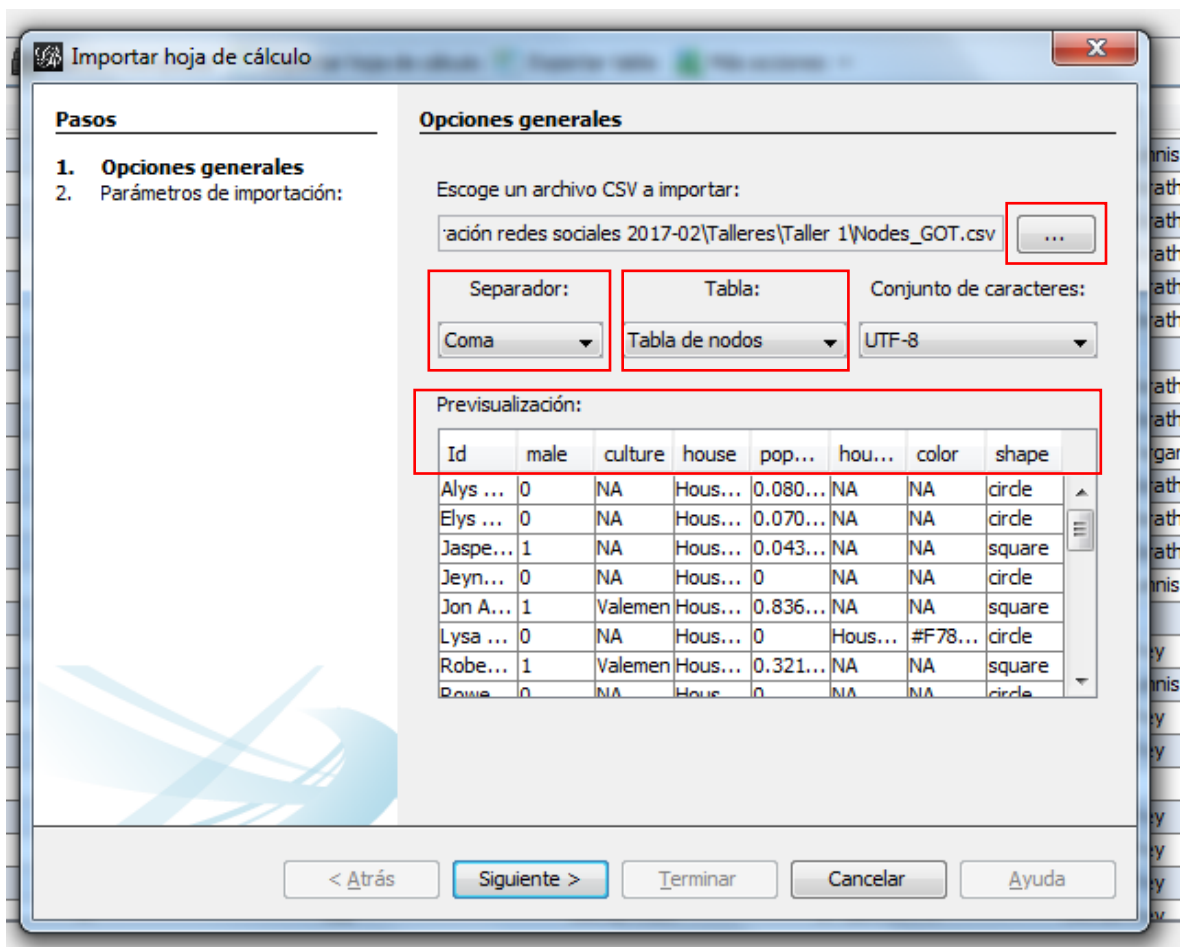
I. “File” → “Open”:



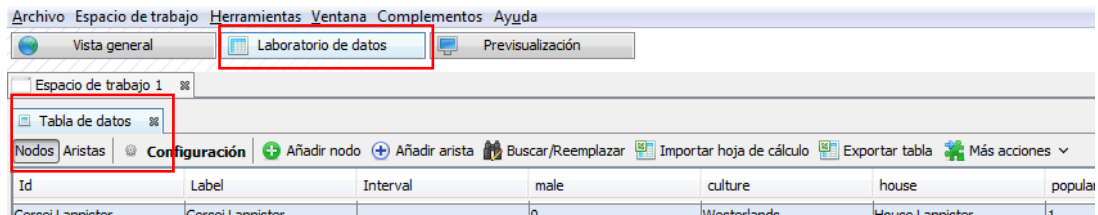
II. Ir a la opción “Data Laboratory” → “Importar hoja de cálculo”



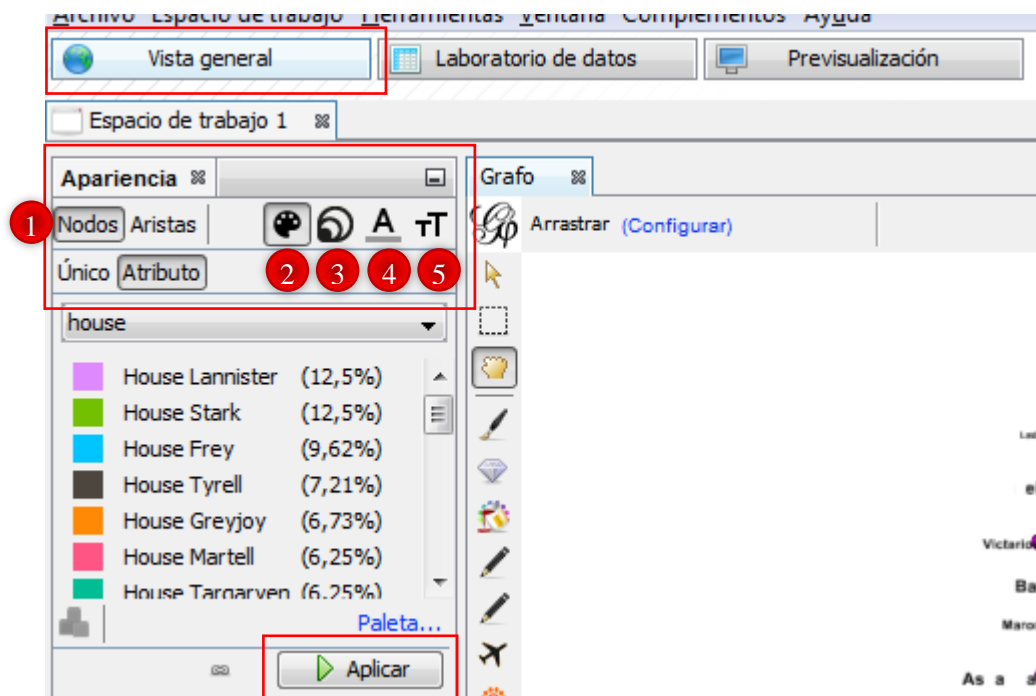
III. Importar por separado la hoja de cálculo “Edges_GOT.csv” y “Nodes_GOT.csv”; en “Tabla:” verificar que la esté seleccionada la opción adecuada y en “Separador:” que la opción seleccionada muestre la información en columnas:



IV. En laboratorio se encontrará la información de los nodos y las conexiones:



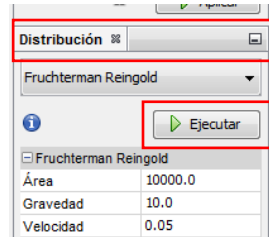
V. En vista general se puede modificar la apariencia de la red:



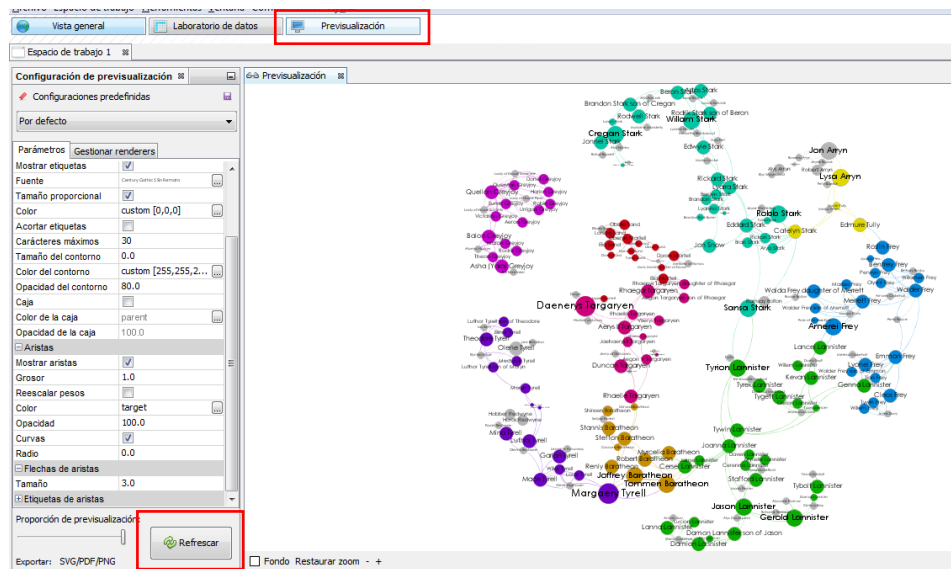
1. Modificar la apariencia de los nodos o las aristas:
2. Modificar el color del nodo/arista, con valores únicos o atributos.
3. Modifica el tamaño del nodo/arista, con valores únicos o atributos.
4. Modifica el la fuente de la etiqueta del nodo/arista, con valores únicos o atributos.
5. Modifica el tamaño de la fuente de la etiqueta del nodo/arista, con valores únicos o atributos.

Es necesario “Aplicar” todos los cambios

- VI. Para modificar la distribución de los nodos escoger entre los algoritmos que Gephi ofrece y modificar los parámetros:



- VII. Para visualizar la red ingresar a “Previsualización” modificar los parámetros necesarios y refrescar la imagen para ver la red:



2. Su segunda tarea es caracterizar la red de familiares de GOT. Para esto, se requiere evaluar las principales medidas de centralidad de la red.

Utilizando Python y Networkx, importe los archivos .csv de nodos y arcos como una red. Para cada literal especifique la medida utilizada para la medición de acuerdo a lo visto en clase sobre teoría de grafos y de solución al literal:

- Encontrar a los 5 personajes más conectados de la red por descendencia y matrimonios. ¿Qué se puede interpretar del rol de estos personajes en la red?
- Encontrar la casa más densamente conectada (medida como la densidad de del subgrafo que representa a cada casa). ¿Qué se puede interpretar acerca de la estructura de esta casa con respecto a las demás?
- Encontrar la casa con más conexiones con otras casas. ¿Qué se puede interpretar acerca del rol de esta casa en el sistema social de GOT?
- Encontrar el personaje de cada casa principal que en promedio se encuentra más cerca a los personajes de la red. ¿Cuál es una posible explicación de la posición de estos personajes?

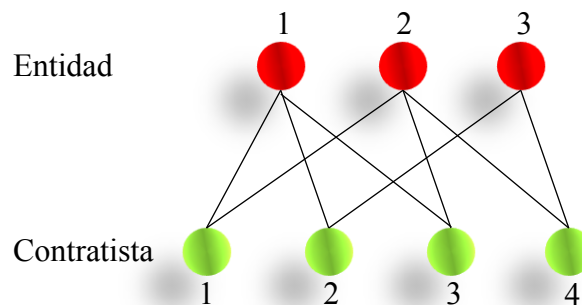
- e. Encontrar a los 5 personajes que se encuentran más veces dentro del camino más corto entre todos los nodos de la red. ¿Qué puede suceder con la topología de la red si estos personajes son asesinados?
- f. Encontrar la probabilidad de que los padres de un personaje sean esposos. ¿Qué puede inferir de las relaciones familiares de los personajes de GOT?
- g. Encontrar a los 5 personajes más conectados y cuyos vecinos tienen mayor grado. ¿Qué puede inferir de la influencia de estos personajes en la red?
- h. Visualice la red resaltando a los personajes encontrados en los literales anteriores con distintos colores según la medida de centralidad. Especifique los casos en que un personaje aparezca en varias medidas.

Punto 2 - Sistema electrónico de contratación pública (SECOP):

El sistema electrónico de contratación pública (SECOP) es la plataforma en la cual las Entidades Estatales de Colombia deben publicar los documentos de los procesos de contratación, desde la planeación del contrato hasta su liquidación. También permite a las Entidades Estatales y al sector privado tener una comunicación abierta y reglada sobre los Procesos de Contratación.¹ Adjunto encontrará tres archivos .csv con la información correspondiente a los procesos de contratación del año 2017 a nivel nacional. A partir de largas búsquedas en el SECOP, el equipo de Estructuración e Redes les ha facilitado la construcción de los siguientes archivos para ser analizados en términos de redes sociales:

- “Nodos_entidad.csv” contiene la información de las entidades que tienen procesos de contratación: NIT de la entidad y nombre de la entidad.
- “Nodos_contratista.csv” contiene la información de los contratistas a los cuales les fue adjudicado cada proceso de contratación: Identificación del contratista, Nombre/Razón social del contratista y Departamento/Municipio del contratista.
- “Arcos_procesos.csv” contiene la información de cada proceso de contratación: NIT de la entidad (Origen), Identificación del contratista (Destino), Estado del proceso, Objeto a contratar, Origen de los Recursos y Valor contrato con Adiciones (USD).

1. La anterior información corresponde a una red bipartita en la cual solo existen dos tipos de nodos: el primer tipo de nodos representa a las entidades del estado y el segundo tipo representa a las empresas contratistas. En este tipo de redes solo hay conexiones entre los nodos de diferente tipo.



¹ <https://www.colombiacompra.gov.co/secop/secop-i>

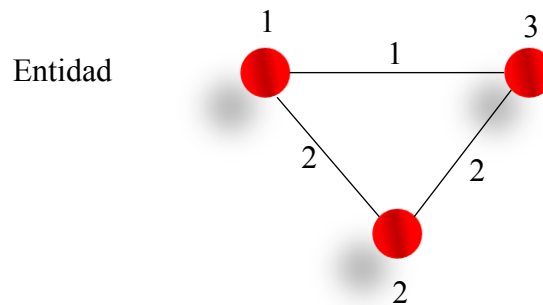
Las redes bipartitas se representan utilizando una *matriz de incidencia* B . Dicha matriz es comparable con la matriz de adyacencia en el caso de redes monopartitas (cuyos nodos son de un mismo tipo). En la *matriz de incidencia* B , las filas representan a los contratistas y las columnas a las entidades del estado. En este caso el valor del elemento b_{ij} es:

$$b_{ij} = \begin{cases} 1 & \text{si la empresa contratista } i \text{ fue contratada por la entidad } j \\ 0 & \text{de lo contrario} \end{cases}$$

Por ejemplo para una matriz de 4 contratistas y 3 entidades la matriz de incidencia es:

$$B = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

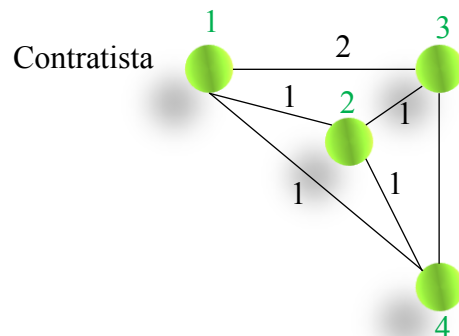
Aunque la red bipartita puede ser la representación más completa de una red particular, a menudo es más conveniente trabajar con conexiones directas entre nodos de un solo tipo. Para esto, podemos crear una *proyección de un tipo de nodo* desde la forma de *dos tipos de nodo* de la red bipartita. Como ejemplo podríamos decir que dos entidades están conectadas si tienen un contratista en común, entonces la proyección sería:



Por otro lado se podría entender la red desde el punto de vista de los contratistas, donde cada conexión entre contratista representa la existencia de una o varias entidades en común:

$$b_{ij}^T = \begin{cases} 1 & \text{si la entidad } i \text{ contrató a la empresa contratista } j \\ 0 & \text{de lo contrario} \end{cases}$$

$$B^T = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$



En las redes anteriores cada arco entre entidades/contratistas tiene un peso asociado a la cantidad de contratistas/entidades que tenían en común en la red bipartita. Para hacer la proyección desde la matriz de incidencia a una matriz de adyacencia debemos:

- I. Calcular la matriz $P = B^T B$
 - II. Igualar los valores de la diagonal de la matriz P a cero.
-
1. Su primera tarea es importar los archivos .csv con los dos tipos de nodos y los arcos en Python, crear la red bipartita con la librería networkx y hacer dos proyecciones:
 - a. Proyección desde la red bipartita a una red de entidades con pesos
 - b. Proyección desde la red bipartita a una red de contratistas con pesos.
 2. Para ambas redes calcular las siguientes medidas y concluir con respecto a los resultados:
 - a. Calcule la cantidad promedio de conexiones.
 - b. La transitividad promedio de la red.
 - c. Fracción de conexiones reales sobre conexiones posibles.
 - d. Longitud promedio del camino entre todo par de nodos
 3. Para ambas redes encontrar a las 5 entidades que tengan el valor más alto de las siguientes medidas de centralidad y concluya con respecto a sus resultados:
 - a. Grado
 - b. Betweenness
 - c. Coeficiente de clustering
 - d. Closeness
 - e. Page Rank
 4. Haga una matriz de correlación entre las medidas y concluya cuales son las entidades o empresas contratistas con mayor incidencia en la contratación estatal.
 5. Bono: Grafique ambas redes en Gephi de tal manera que pueda inferir información relevante acerca de los procesos de contratación nacionales del año 2017.

Punto 3: Análisis de homicidios

Usted ha sido contratado por el departamento de justicia de los Estados Unidos para analizar el comportamiento de las redes sociales creadas entorno a los homicidios cometidos en el 2017 en todo el país. Para esto usted cuenta con la base de datos de asesinatos del reporte suplementario de homicidios del FBI y los homicidios registrados por el FOIA que no se encuentran reportados por el departamento de justicia. En el archivo adjunto “Edges_homicides.csv” encontrarán la información correspondiente a un grafo bipartita que contiene:

- La ciudad donde se cometió el homicidio.
- Un patrón del homicidio: Tipo del crimen, sexo de la víctima, raza de la víctima, sexo del homicida y raza del homicida.

- A continuación se presentan las siglas utilizadas para describir cada patrón:

Crime Type	Manslaughter by Negligence	MN
	Murder or Manslaughter	MM

Gender victim/perpetrator	Male	M
	Female	F

Race victim/perpetrator	White	W
	Black	B
	Asian/Pacific Islander	A
	Native American/Alaska Native	N
	Unknown	U

1. Su primera tarea es realizar las proyecciones en Python de la red bipartita (basándose en el proceso realizado en el Punto 2) donde se evidencien:
 - a. Los homicidios ocurridos en las diferentes ciudades, teniendo en cuenta patrón con el que se cometieron: entiéndase al patrón del homicidio como nodo original y a la ciudad como grupo.
 - b. Los homicidios según los diferentes patrones, teniendo en cuenta la ciudad donde se cometieron: entiéndase la ciudad como nodo original y al patrón del homicidio como grupo.
2. Su segunda tarea es analizar la topología de cada red del literal anterior. Para esto usted debe:
 - a. Encontrar las medidas de centralidad: Grado promedio, coeficiente de clustering promedio, diámetro y densidad.
 - b. Encontrar la distribución del grado de cada red.
 - c. Asociar cada red con alguno de los modelos de redes aleatorias vistos en clase: Ęrdos-Rényi, Wats-Strogatz y Barabasi-Albert. Para esto usted debe soportar su decisión con los resultados de los literales a y b.
3. ¿Podría inferir la existencia de asesinos en serie con la información suministrada?
4. Según el resultado del literal anterior, recomiende una estrategia de destrucción de la red de homicidios en Estados Unidos identificando el modus operandi más relevante dentro de cada red.
5. Grafique en Gephi cada proyección de la red bipartita donde se evidencie la topología de la red y los nodos más influyentes. Concluya.