

TALLER 1**Jesús Alberto Solano Gómez 201212411****Andrea Pilar Sissa Velandia 201424456****Punto 1 (35 puntos) – LAS CASAS EN HARRY POTTER**

Harry Potter es una serie de novelas de fantasía escrita por la autora J.K Rowling y publicada en 1997. El personaje principal es un huérfano que vive con su tío, tía y primo quienes durante toda su vida lo han maltratado. Cuando Harry cumple 11 años, Albus Dumbledore, el rector de la escuela Hogwarts de magia y hechicería lo recluta para que aprenda magia. En Hogwarts, Harry se siente en casa y hace muchos amigos, como Ronald Weasley y Hermione Granger (Pottermore, s.f.).

Ustedes tendrán el reto de modelar y analizar las redes sociales de la serie en el primer y quinto libro asumiendo que cada nodo representa un personaje y que existe relación dirigida entre dos nodos, cuando un personaje considera que el otro es su amigo.

Adjunto encontrará tres archivos .csv con la información de la red (Bossaert & Meidert, 2013):

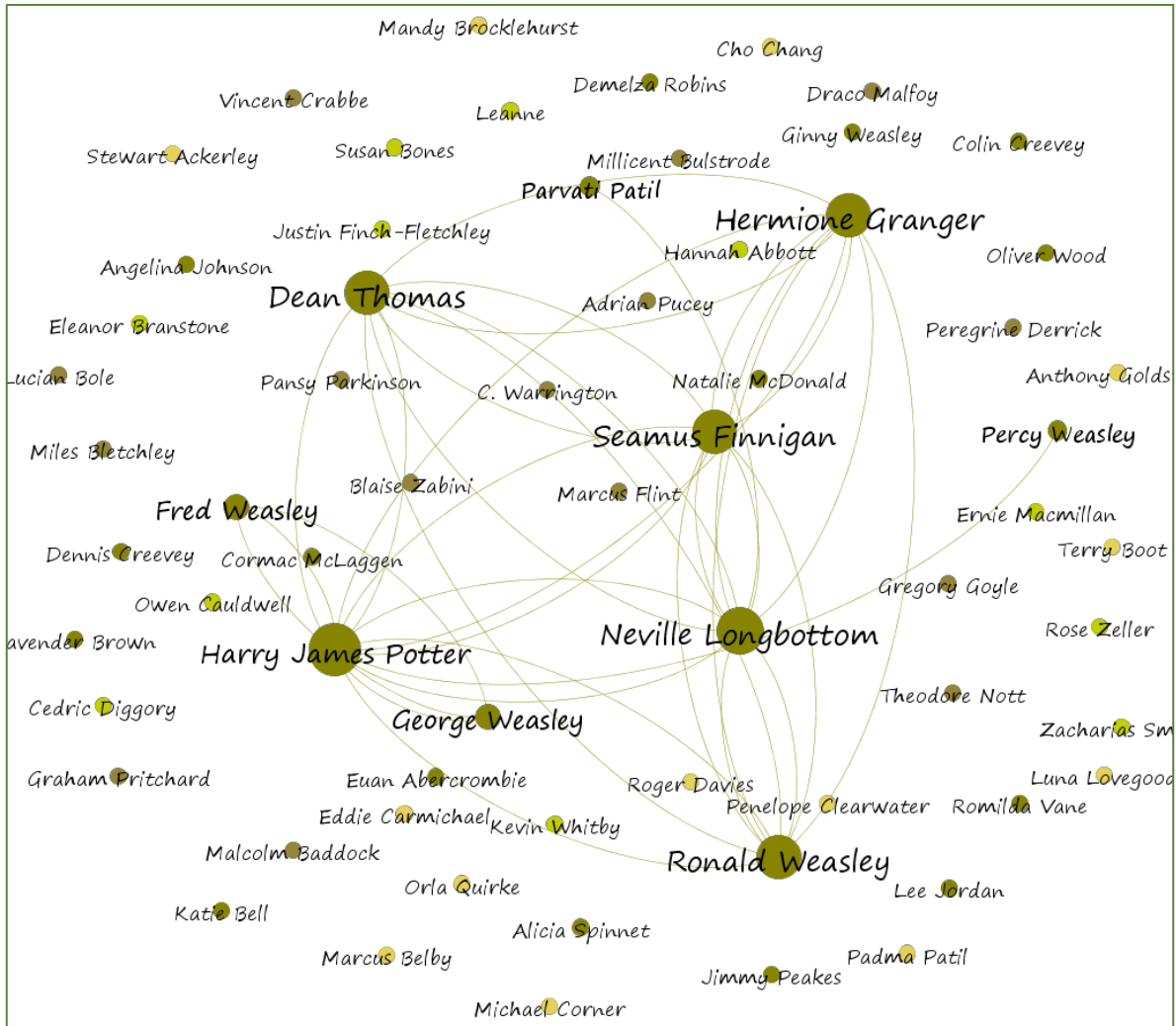
- **Nodos_HarryP.csv** contiene la información de los personajes de la serie:
 - **Id:** Id del personaje.
 - **Name:** Nombre del personaje.
 - **Schoolyear:** Año en el que el personaje ingresó a Hogwarts.
 - **Gender:** Género del personaje. (1: hombre 2: mujer)
 - **House:** Casa a la cual pertenece el personaje. (1: Gryffindor, 2: Hufflepuff, 3: Ravenclaw, 4: Slytherin)
 - **House Name:** Nombre de la casa a la cual pertenece el personaje.
- **Edges_Libro1.csv** contiene la información de las conexiones entre personajes en el primer libro (Harry Potter y la piedra filosofal):
 - **Source:** Personaje de origen de la conexión.
 - **Target:** Personaje de destino de la conexión.
 - **Type:** 1: Si el personaje que representa el nodo de origen considera que el otro personaje (nodo destino) es su amigo. 0: de lo contrario.
- **Edges_Libro5.csv** contiene la información de las conexiones entre personajes en el quinto libro (Harry Potter y la orden del Fénix):
 - **Source:** Personaje de origen de la conexión.
 - **Target:** Personaje de destino de la conexión.
 - **Type:** 1: Si el personaje que representa el nodo de origen considera que el otro personaje (nodo destino) es su amigo. 0: de lo contrario.

Teniendo en cuenta lo anterior, desarrollen los siguientes numerales:

1. **(15 puntos)** Su primera tarea es visualizar la red de cada libro en Gephi y así empezar a comprender cada sistema. Para esto, se recomienda seguir los pasos a continuación y concluir con base a lo observado acerca de la estructura de las relaciones de cada red.

Visualicen cada una de las redes (libro 1 y libro 2), muestren una imagen de cada red y realicen conclusiones acerca de las características que pudieron observar sobre la estructura de cada red y compárenlas:

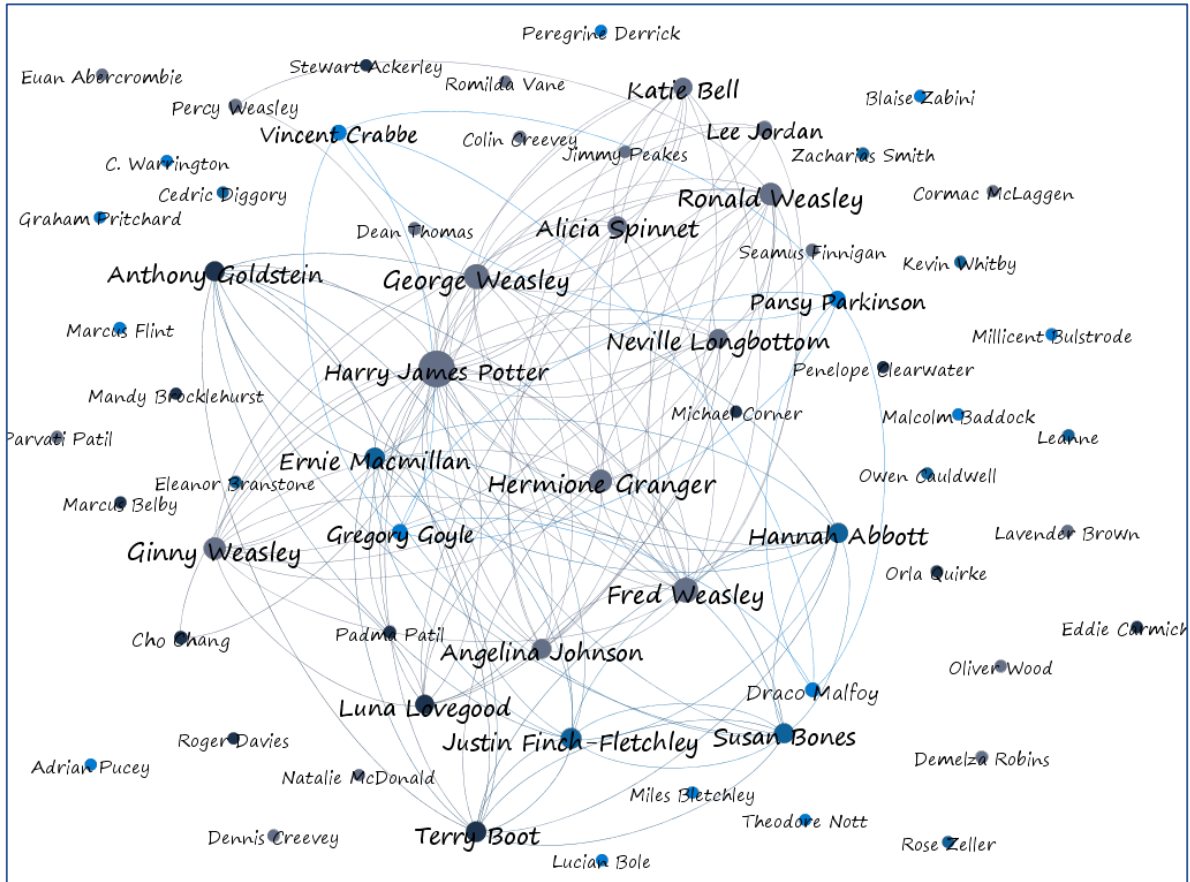
Red libro 1



Red Harry Potter Libro 1, elaborada en Gephi

Al observar la red del libro 1, se puede concluir que en este punto de la historia los personajes no están tan conectados entre ellos, ya que el número de arcos de la red es muy pequeño. De igual forma, se puede ver cómo se van formando relaciones entre los personajes y cómo unos tienen más conexiones que otros en especial para la casa Gryffindor (verde oscuro). De igual forma, las otras casas no tienen ningún tipo de conexión entre ellas y Harry James Potter desde ya empieza a ser un personaje foco importante en el comportamiento de esta red.

Red libro 5



Red Harry Potter Libro 5, elaborada en Gephi

En esta red se puede observar en comparación con el libro 1, cómo las conexiones entre los personajes aumentaron de manera considerable. De igual forma, las conexiones entre casas y entre los personajes de cada casa aumentaron para este libro, aunque persiste la existencia de personajes que no tienen ningún tipo de conexión con la red. A su vez, persiste la existencia de Harry James Potter como un personaje con gran cantidad de conexiones al igual que otros personajes de su misma casa o casas diferentes que generaron conexiones del libro 1 al libro 5.

Por lo anterior, al comparar las redes se puede ver que del libro 1 al 5 se generaron más conexiones entre todos los personajes y entre casas y que, más personajes aumentaron su protagonismo en toda la red de personajes de Hogwarts.

2. **(20 puntos)** Su segunda tarea es caracterizar la red de Harry Potter. Para esto, se requiere evaluar las principales medidas de centralidad de la red.

Utilizando Python y Networkx, importen los archivos .csv de nodos y arcos como una red. Para cada literal especifiquen la medida utilizada para la medición de acuerdo a lo visto en clase sobre teoría de grafos y respondan a la pregunta. Realicen el análisis utilizando los archivos de arcos de los libros 1 y 5, es decir, construyan las dos redes por separado.

- a. **(2 puntos)** ¿Cuántos estudiantes entraron a Hogwarts entre 1990 y 1995?

```
El numero de estudiantes que ingresaron entre los años 1990 y  
1995 a Hogwarts es:  
48
```

Output Python, literal a

A partir de lo encontrado en Python por medio de un condicional simple para el año en el que ingresó el personaje a Hogwarts, se puede ver que entre 1990 y 1995 entraron 48 estudiantes a la escuela de magia.

Para cada red solucionen los siguientes literales:

- b. **(2 puntos)** Encontrar a los 5 personajes más conectados de la red por casa ¿Qué se puede interpretar del rol de estos personajes en la red?

Libro 1

```
En la casa ----> Slytherin los 5 personajes más conectados son:  
* Adrian Pucey -- degree : 0  
* Lucian Bole -- degree : 0  
* Malcolm Baddock -- degree : 0  
* Blaise Zabini -- degree : 0  
* C. Warrington -- degree : 0  
  
En la casa ----> Gryffindor los 5 personajes más conectados son:  
* Harry James Potter -- degree : 13  
* Neville Longbottom -- degree : 11  
* Dean Thomas -- degree : 10  
* Hermione Granger -- degree : 10  
* Ronald Weasley -- degree : 10  
  
En la casa ----> Ravenclaw los 5 personajes más conectados son:  
* Luna Lovegood -- degree : 0  
* Anthony Goldstein -- degree : 0  
* Mandy Brocklehurst -- degree : 0  
* Marcus Belby -- degree : 0  
* Cho Chang -- degree : 0  
  
En la casa ----> Hufflepuff los 5 personajes más conectados son:  
* Leanne -- degree : 0  
* Zacharias Smith -- degree : 0  
* Cedric Diggory -- degree : 0  
* Susan Bones -- degree : 0  
* Owen Cauldwell -- degree : 0
```

Output Python, literal b Libro 1

Para encontrar a los 5 personajes más conectados de la red por casa se calculó la centralidad de grado (número total de conexiones que entran y salen de un nodo a sus nodos adyacentes) de cada personaje y, se escogieron los 5 personajes que tenían el valor más alto. Así, dados los resultados anteriores, es posible decir que en las casas Slytherin, Ravenclaw y Hufflepuff los personajes no están conectados entre ellos ya que la centralidad de grado es igual a cero para todos. De esta manera, teniendo en

cuenta que se estudia una red social, los personajes que tienen una centralidad de grado alta pueden tener más influencia, más acceso a la información o más prestigio que aquellos que tienen menos conexiones (Newman, 2009).

Por lo tanto, en estas casas no existe un personaje central que pueda tener más influencia o prestigio. Por otro lado, en la casa Gryffindor sí es posible encontrar los 5 personajes más conectados en la red que son, en orden descendente, Harry Potter, Neville Longbottom, Dean Thomas, Hermione Granger y Ronald Weasley. Esto indica que en esta red social sí hay conexiones y que la persona más conectada que cuenta con más influencia o prestigio es Harry Potter y sus amigos más cercanos. Por lo tanto, el rol de estos personajes en la red es ser focos de información y atención para toda la casa, ya que tienen el mayor número de conexiones y a través de estas pueden llegar a todos los personajes que pertenecen a Gryffindor.

Libro 5

```
En la casa ----> Slytherin los 5 personajes más conectados son:
* Pansy Parkinson -- degree : 6
* Gregory Goyle -- degree : 5
* Vincent Crabbe -- degree : 5
* Draco Malfoy -- degree : 4
* Adrian Pucey -- degree : 0

En la casa ----> Gryffindor los 5 personajes más conectados son:
* Harry James Potter -- degree : 22
* Fred Weasley -- degree : 18
* George Weasley -- degree : 18
* Ronald Weasley -- degree : 13
* Ginny Weasley -- degree : 12

En la casa ----> Ravenclaw los 5 personajes más conectados son:
* Anthony Goldstein -- degree : 2
* Terry Boot -- degree : 2
* Luna Lovegood -- degree : 0
* Mandy Brocklehurst -- degree : 0
* Marcus Belby -- degree : 0

En la casa ----> Hufflepuff los 5 personajes más conectados son:
* Susan Bones -- degree : 6
* Ernie Macmillan -- degree : 6
* Hannah Abbott -- degree : 6
* Justin Finch-Fletchley -- degree : 6
* Leanne -- degree : 0
```

Output Python, literal b Libro5

Para encontrar a los 5 personajes más conectados de la red por casa se calculó la centralidad de grado (número total de conexiones que entran y salen de un nodo a sus nodos adyacentes) de cada personaje y, se escogieron los 5 personajes que tenían el valor más alto. Así, dados los resultados anteriores, es posible decir que en la casa Slytherin los personajes más conectados son Pansy Parkinson, Gregory Goyle, Vincent Crabbe y Draco Malfoy. De igual forma, para la casa Ravenclaw los personajes que tienen más centralidad de grado son Anthony Goldstein y Terry Boot. Ahora bien, los personajes más conectados en la casa Hufflepuff son Susan Bones, Ernie Macmillan, Hannah Abbott y Justin Finch-

Fletchley con una centralidad de grado de 6 cada uno. Finalmente, en la casa Gryffindor sí se pueden encontrar los 5 personajes más conectados que son, en orden decendente, Harry James Potter, Fred Weasley, George Weasley, Ronald Weasley y Ginny Weasley.

De esta manera, teniendo en cuenta que se estudia una red social, los personajes que tienen una centralidad de grado alta pueden tener más influencia, más acceso a la información o más prestigio que aquellos que tienen menos conexiones (Newman, 2009). Esto indica que en esta red social las personas que cuentan con un mayor alto grado de centralidad tienen más influencia o prestigio. Por lo tanto, el rol de estos personajes en la red es ser focos de información y atención para toda la casa, ya que tienen el mayor número de conexiones y a través de estas pueden llegar a todos los personajes que pertenecen a cada casa en Hogwarts.

- c. **(2 puntos)** Encontrar la casa más densamente conectada (medida como la densidad del subgrafo inducido por cada casa). ¿Qué se puede interpretar acerca de la estructura de esta casa con respecto a las demás?

Libro 1

```
Las densidades de las casas en Hogwarts son:
```

```
Slytherin -- > 0
Gryffindor -- > 0.06
Ravenclaw -- > 0
Hufflepuff -- > 0
```

Output Python, literal c Libro1

La casa más densamente conectada se encontró como aquella que tenía el mayor valor asociado a la propiedad densidad característica de los grafos. Dicha propiedad es la proporción del número de lazos presentes en la red sobre el número máximo posible, lo cual indica qué tan cerca está la red de estar totalmente conectada. Así, observando los resultados anteriores, la casa más densamente conectada es Gryffindor ya que tiene una densidad de 0.06 en comparación con las otras casas cuya densidad es cero. De esta manera, la estructura de Gryffindor es más compacta y su número de conexiones se acerca más al número de conexiones totales que se tendría si todos los nodos estuvieran conectados. A pesar de ello, no es un valor significativo ya que es el 6% del total de conexiones que podría tener esta red. Sin embargo, es mejor que la conexión que tienen las otras casas con una densidad igual a cero.

Libro 5

```
Las densidades de las casas en Hogwarts son:
```

```
Slytherin -- > 0.047619047619047616
Gryffindor -- > 0.11833333333333333
Ravenclaw -- > 0.01282051282051282
Hufflepuff -- > 0.10909090909090909
```

Output Python, literal c Libro5

La casa más densamente conectada se encontró como aquella que tenía el mayor valor asociado a la propiedad densidad característica de los grafos. Dicha propiedad es la proporción del número de lazos presentes en la red sobre el número máximo posible, lo cual indica qué tan cerca está la red de estar totalmente conectada. Así, observando los resultados anteriores, la casa más densamente conectada es Gryffindor ya que tiene una densidad de 0.12 en comparación con las otras casas. De esta manera, la estructura de Gryffindor es más compacta y su número de conexiones se acerca más al número de conexiones totales que se tendría si todos los nodos estuvieran conectados. A pesar de ello, no es un valor significativo ya que es el 11.83% del total de conexiones que podría tener esta red. Sin embargo, es mejor que la conexión que tienen las otras casas con una densidad menor al 10%.

- d. **(2 puntos)** Encontrar la casa con más conexiones con otras casas. ¿Qué se puede interpretar acerca del rol de esta casa en Hogwarts?

Libro 1

```
Las conexiones externas de cada casa son:
```

```
{'Gryffindor': 0, 'Hufflepuff': 0, 'Ravenclaw': 0, 'Slytherin': 0}
```

Output Python, literal d Libro 1

La casa con más conexiones con otras casas se calcula contando el número de vecinos de cada nodo que se conecta con otra casa. Así, se encontró que las casas no están conectadas entre ellas ya que el número de conexiones de cada casa con las otras es cero y, que, si hubiera alguna que tuviera más conexiones, sería central dentro de toda la red de Hogwarts. Es decir, sería un foco de información o prestigio para las otras casas. Dado lo anterior, es importante mencionar que este procedimiento es equivalente a encontrar la centralidad de grado de salida de cada casa, es decir, el número de conexiones que salen de cada nodo a los nodos de otras casas.

Libro 5

```
Las conexiones externas de cada casa son:
```

```
{'Gryffindor': 8, 'Hufflepuff': 12, 'Ravenclaw': 18, 'Slytherin': 0}
```

Output Python, literal d Libro 5

La casa con más conexiones con otras casas se calcula contando el número de vecinos de cada nodo que se conecta con otra casa. Así, se encontró que la casa más conectada a otras es Ravenclaw seguida de Hufflepuff y Gryffindor. Esto indica que Ravenclaw es central dentro de toda la red de Hogwarts ya que tiene mayor número de conexiones con Hufflepuff y Gryffindor. Es decir, es un foco de información o prestigio para las otras casas. Dado lo anterior, es importante mencionar que este procedimiento es equivalente a encontrar la centralidad de grado de salida de cada casa, es decir, el número de conexiones que salen de cada nodo a los nodos de otras casas. Finalmente, Slytherin es la única casa que no tiene conexiones con las otras y, por tanto, se puede considerar un nodo aislado dentro de la red de Hogwarts.

- e. (2 puntos) Encontrar el personaje de cada casa principal que en promedio se encuentra más cerca a los demás personajes de la red. ¿Cuál es una posible explicación de la posición de estos personajes?

Libro 1

```
El personaje más conectado de la casa es:  
  
Slytherin ---> Adrian Pucey con closeness centrality de 0.0  
Gryffindor ---> Harry James Potter con closeness centrality de 0.3375  
Ravenclaw ---> Luna Lovegood con closeness centrality de 0.0  
Hufflepuff ---> Leanne con closeness centrality de 0.0
```

Output Python, literal e Libro 1

El personaje de cada casa principal que en promedio se encuentra más cerca a los demás personajes de la red se encontró utilizando como medida de centralidad la cercanía o closeness centrality. La cual, es el promedio de las distancias más cortas desde un personaje hacia todos los demás. En este caso, el personaje que tiene mayor cercanía a todos los demás es Harry James Potter de la casa Gryffindor. La posición de este personaje indica que se encuentra en mejores condiciones para influir en toda la red más rápidamente, que se comporta como un buen emisor de información y que es más accesible a la red desde cualquier nodo (Disney A., 2014).

Libro 5

```
El personaje mas conectado de la casa es:  
  
Slytherin ---> Draco Malfoy con closeness centrality de 0.21428571428571427  
Gryffindor ---> Harry James Potter con closeness centrality de 0.5029761904761905  
Ravenclaw ---> Anthony Goldstein con closeness centrality de 0.08333333333333333  
Hufflepuff ---> Susan Bones con closeness centrality de 0.3
```

Output Python, literal e Libro 5

El personaje de cada casa principal que en promedio se encuentra más cerca a los demás personajes de la red se encontró utilizando como medida de centralidad la cercanía o closeness centrality. La cual, es el promedio de las distancias más cortas desde un personaje hacia todos los demás. En este caso, el personaje que tiene mayor cercanía a todos los demás en Gryffindor es Harry James Potter, en Slytherin Draco Malfoy, para Ravenclaw es Anthony Goldstein y, en Hufflepuff es Susan Bones. La posición de estos personajes indica que se encuentran en mejores condiciones para influir en toda la red más rápidamente, que se comportan como unos buenos emisores de información y que son más accesibles a la red desde cualquier nodo (Disney A., 2014).

- f. (2 puntos) 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 Voldemort asesinara a estos personajes?

Libro 1

```
El personaje más conectado de la casa es:  
  
Slytherin ---> Adrian Pucey con betweenness centrality de 0.0  
Gryffindor ---> Harry James Potter con betweenness centrality  
de 0.02717391304347826  
Ravenclaw ---> Luna Lovegood con betweenness centrality de 0.0  
Hufflepuff ---> Leanne con betweenness centrality de 0.0  
  
* Nota: Si no hay personajes conectados en la casa se toma el  
primer nodo asociado a la casa.
```

Output Python, literal f Libro 1

Los 5 personajes que se encuentran más veces dentro del camino más corto entre todos los nodos de la red se encontraron con la medida de centralidad intermediación o betweenness centrality. Esta, es la frecuencia con la que un nodo aparece en el camino más corto entre todos los pares de nodos de la red. Es decir, muestra qué nodos actúan como 'puentes' entre nodos en una red (Disney A., 2014). Así pues, con los resultados encontrados se puede decir que esta medida de centralidad es mayor para el personaje Harry James Potter de la casa Gryffindor, lo cual lo convierte en una persona que influye mucho en el flujo de información y unión en el sistema. Por tanto, si Voldemort asesinara a este personaje, la topología de la red entendida como el mapa físico o lógico de una red para intercambiar datos se vería muy afectada, ya que no está uno de los principales personajes más influyentes dentro de la red al ser el que tiene más acceso a todos los demás. De igual forma, la estructura de la red cambiaría considerablemente en cuanto a la velocidad del flujo de información y la accesibilidad a varios personajes de la red.

Libro 5

```
El personaje más conectado de la casa es:  
  
Slytherin ---> Pansy Parkinson con betweenness centrality de  
0.01098901098901099  
Gryffindor ---> Harry James Potter con betweenness centrality  
de 0.07744565217391304  
Ravenclaw ---> Luna Lovegood con betweenness centrality de 0.0  
Hufflepuff ---> Leanne con betweenness centrality de 0.0  
  
* Nota: Si no hay personajes conectados en la casa se toma el
```

Output Python, literal f Libro 5

Los 5 personajes que se encuentran más veces dentro del camino más corto entre todos los nodos de la red se encontraron con la medida de centralidad intermediación o betweenness centrality. Esta, es la frecuencia con la que un nodo aparece en el camino más corto entre todos los pares de nodos de la red. Es decir, muestra qué nodos actúan como 'puentes' entre nodos en una red (Disney A., 2014). Así pues, con los resultados encontrados se puede decir que esta medida de centralidad es mayor para los personajes Harry James Potter de la casa Gryffindor y Pansy Parkinson de la casa Slytherin, lo cual los convierte en personas que influyen mucho en el flujo de información y unión en el sistema. Por

tanto, si Voldemort asesinara a estos personajes, la topología de la red entendida como el mapa físico o lógico de una red para intercambiar datos se vería muy afectada, ya que no están los principales personajes más influyentes dentro de la red al ser los que tiene más acceso a todos los demás. De igual forma, la estructura de la red cambiaría considerablemente en cuanto a la velocidad del flujo de información y la accesibilidad a varios personajes de la red.

- g. **(2 puntos)** Encontrar la probabilidad de que los amigos de un personaje sean también amigos entre sí. ¿Qué puede inferir de las relaciones de amistad de los personajes de Harry Potter?

Libro 1

```
Adrian Pucey -- > 0
Alicia Spinnet -- > 0
Angelina Johnson -- > 0
Anthony Goldstein -- > 0
.
.
Theodore Nott -- > 0
Vincent Crabbe -- > 0
Zacharias Smith -- > 0
```

Output Python, literal g Libro 1

Para encontrar la probabilidad de que los amigos de un personaje sean también amigos entre sí, se calculó la medida de transitividad de cada personaje, es decir, la fracción de todos los triángulos (cuando dos nodos que conocen a un tercero se conocen entre ellos) posibles presentes en la red. Así, los resultados anteriores muestran que, para Dean Thomas, Hermione Granger, Ronald Weasley y Seamus Finnigan, la probabilidad de que sus amigos sean amigos entre ellos es 100%, es decir, tienen una transitividad alta. Por otro lado, Neville Longbottom tiene una transitividad menor con una probabilidad de 80%. Después, en orden descendente están Harry James Potter, Fred Weasley y George Weasley con una probabilidad no menor a 50%. Así, se puede inferir que las relaciones de amistad de los personajes de Harry Potter son muy cerradas, ya que la probabilidad de que los amigos de un personaje sean también amigos entre sí es muy alta, lo cual indica que hay muchos triángulos en la red, al menos en las personas que están conectadas entre sí. Ya que los personajes que no tienen conexión alguna con otros tienen una transitividad igual a cero.

Libro 5

```
Adrian Pucey -- > 0
Alicia Spinnet -- > 1.0
Angelina Johnson -- > 1.0
.
.
Theodore Nott -- > 0
Vincent Crabbe -- > 0.875
Zacharias Smith -- > 0
```

Output Python, literal g Libro 5

Para encontrar la probabilidad de que los amigos de un personaje sean también amigos entre sí, se calculó la medida de transitividad de cada personaje, es decir, la fracción de todos los triángulos (cuando dos nodos que conocen a un tercero se conocen entre ellos) posibles presentes en la red. Así, los resultados anteriores muestran que, para 8 personajes la probabilidad de que sus amigos sean amigos entre ellos es 100%, es decir, tienen una transitividad alta. Por otro lado, la mayoría de los personajes tienen una transitividad mayor al 56%. Y, finalmente, Harry James Potter es el personaje que tiene un menor nivel de transitividad con una probabilidad igual al 23%. Así, se puede inferir que las relaciones de amistad de los personajes de Hogwarts son muy cerradas, ya que la probabilidad de que los amigos de un personaje sean también amigos entre sí es muy alta, lo cual indica que hay muchos triángulos en la red, al menos en las personas que están conectadas entre sí. Ya que los personajes que no tienen conexión alguna con otros tienen una transitividad igual a cero.

- h. **(2 puntos)** 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?

Libro 1

```
Los cinco personajes con mayor eigenvector centrality en Harry Potter son:

Harry James Potter -- > 0.4202
Neville Longbottom -- > 0.4070
Dean Thomas -- > 0.4054
Hermione Granger -- > 0.4054
Ronald Weasley -- > 0.4054
```

Output Python, literal h Libro 1

Los 5 personajes más conectados y cuyos vecinos tienen mayor grado se encontraron por medio de la medida de centralidad de vector propio o eigenvector centrality, la cual es una medida de centralidad que toma en cuenta también la importancia de sus vecinos. Es decir, mide la influencia de un nodo en función del número de enlaces que tiene a otros nodos dentro de la red. Así, dados los resultados anteriores, Harry James Potter, Neville Longbottom, Dean Thomas, Hermione Granger y Ronald Weasley son los que tienen un mayor nivel de centralidad de vector propio, esto quiere decir que están conectados a muchos nodos que a su vez están bien conectados, y, por lo tanto, son buenos candidatos para difundir información. Los personajes más centrales en este sentido corresponden a

centros de grandes grupos cohesivos. Lo cual, quiere decir que tienen una gran influencia dentro de toda la red de personajes de Hogwarts, ya que son nodos que se encuentran conectados a personajes con altos grados de centralidad y, por tanto, cualquier tipo de información puede ser dispersada fácil y rápidamente por ellos.

Libro 5

Los cinco personajes con mayor eigenvector centrality en Harry Potter son:

```
Harry James Potter -- > 0.4361
Fred Weasley -- > 0.3469
George Weasley -- > 0.3469
Hermione Granger -- > 0.3072
Ginny Weasley -- > 0.3013
```

Output Python, literal h Libro 5

Los 5 personajes más conectados y cuyos vecinos tienen mayor grado se encontraron por medio de la medida de centralidad de vector propio o eigenvector centrality, la cual es una medida de centralidad que toma en cuenta también la importancia de sus vecinos. Es decir, mide la influencia de un nodo en función del número de enlaces que tiene a otros nodos dentro de la red. Así, dados los resultados anteriores, Harry James Potter, Fred Weasley, George Weasley, Hermione Granger y Ginny Weasley son los que tienen un mayor nivel de centralidad de vector propio, esto quiere decir que están conectados a muchos nodos que a su vez están bien conectados, y, por lo tanto, son buenos candidatos para difundir información. Los personajes más centrales en este sentido corresponden a centros de grandes grupos cohesivos. Lo cual, quiere decir que tienen una gran influencia dentro de toda la red de personajes de Hogwarts, ya que son nodos que se encuentran conectados a personajes con altos grados de centralidad y, por tanto, cualquier tipo de información puede ser dispersada fácil y rápidamente por ellos.

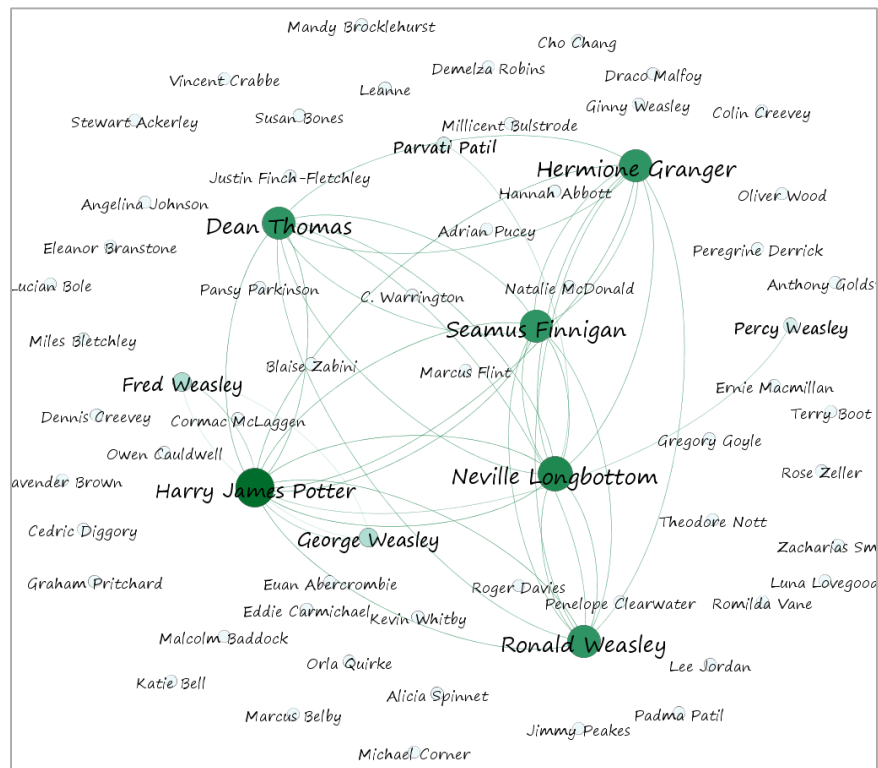
- i. **(2 puntos)** Visualicen en Gephi las redes del libro 1 y 5, resaltando a los personajes encontrados en los literales anteriores con distintos colores según la medida de centralidad (teniendo en cuenta cada medida calculada en los literales anteriores). Especifique los casos en que un personaje aparezca con mayor valor en varias medidas.

A continuación, se muestran los personajes encontrados en los literales anteriores con distintos tamaños según la medida de centralidad, ya que las casas se diferencian según sus colores insignia:

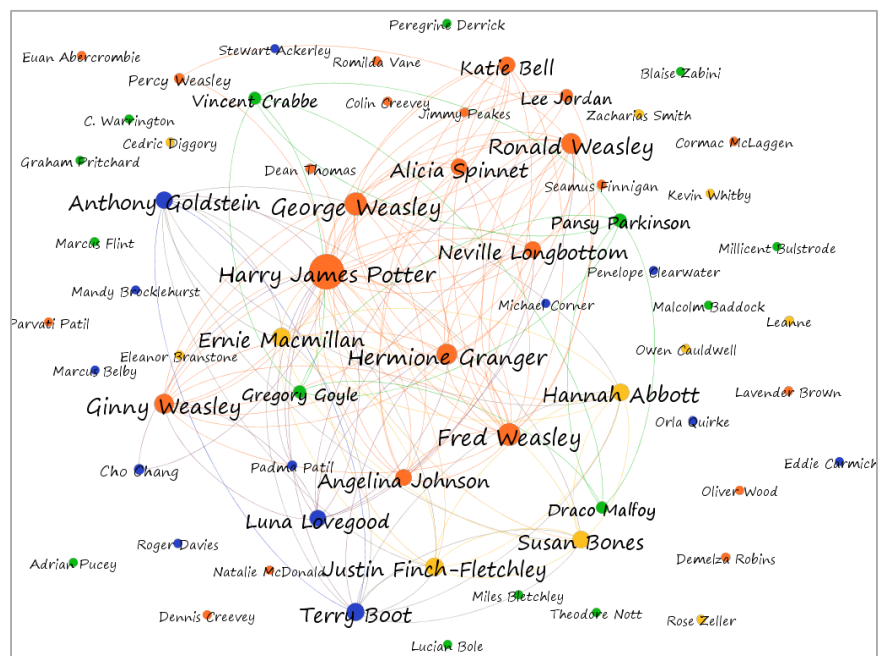
Casa	Color asociado
Gryffindor	Naranja
Slytherin	Verde
Ravenclaw	Azul
Hufflepuff	Amarillo

Tabla 1. Casa según color

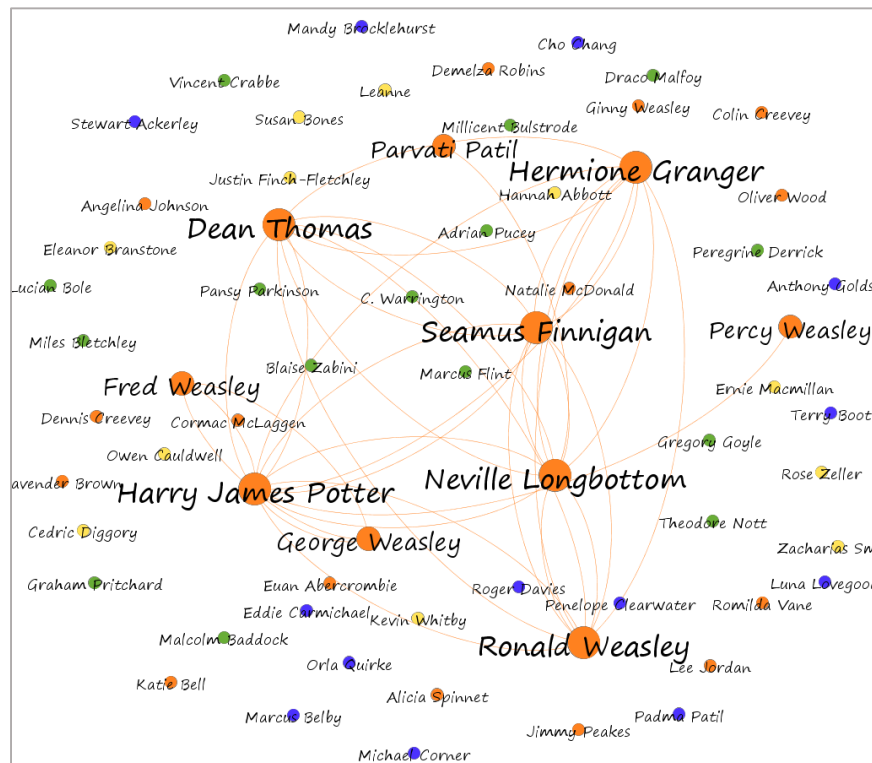
Personajes con mayor centralidad de grado libro 1



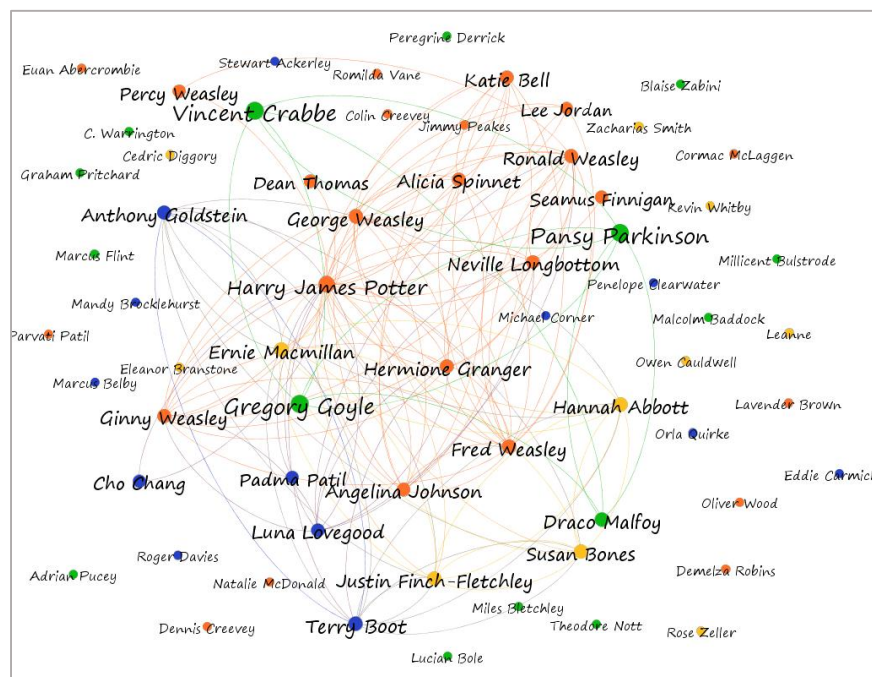
Personajes con mayor centralidad de grado libro 5 por casa



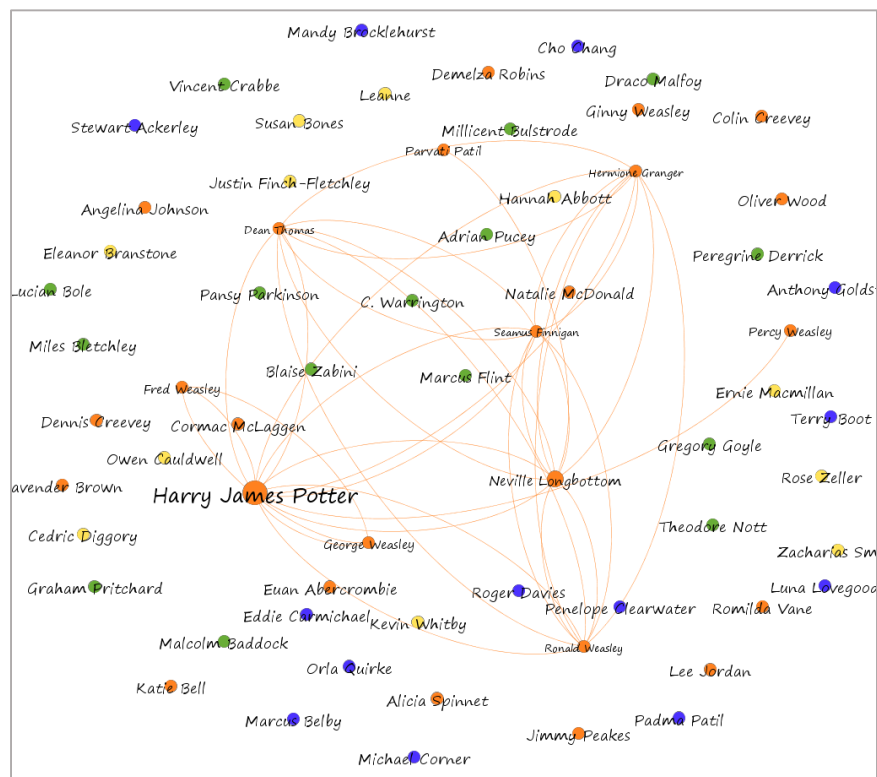
Personajes con mayor closeness centrality Libro 1



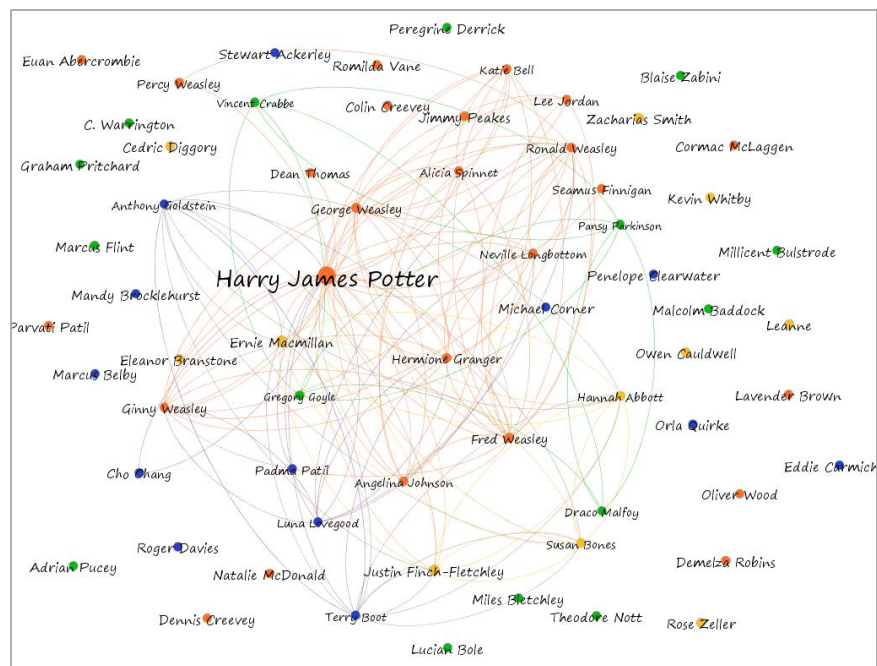
Personajes con mayor closeness centrality Libro 5



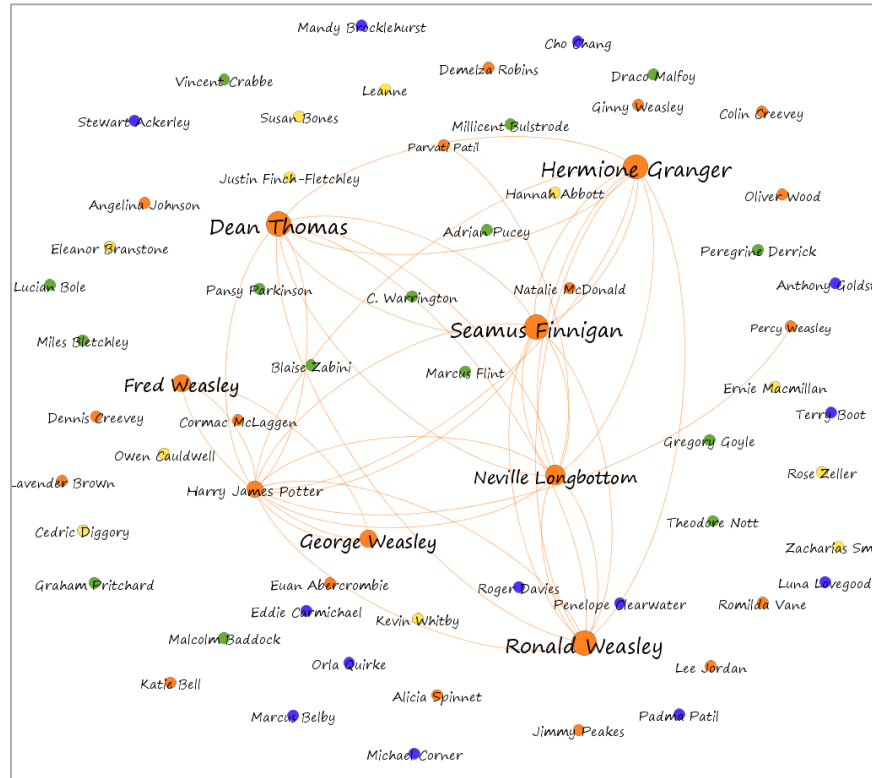
Personajes con mayor betweenness centrality Libro 1



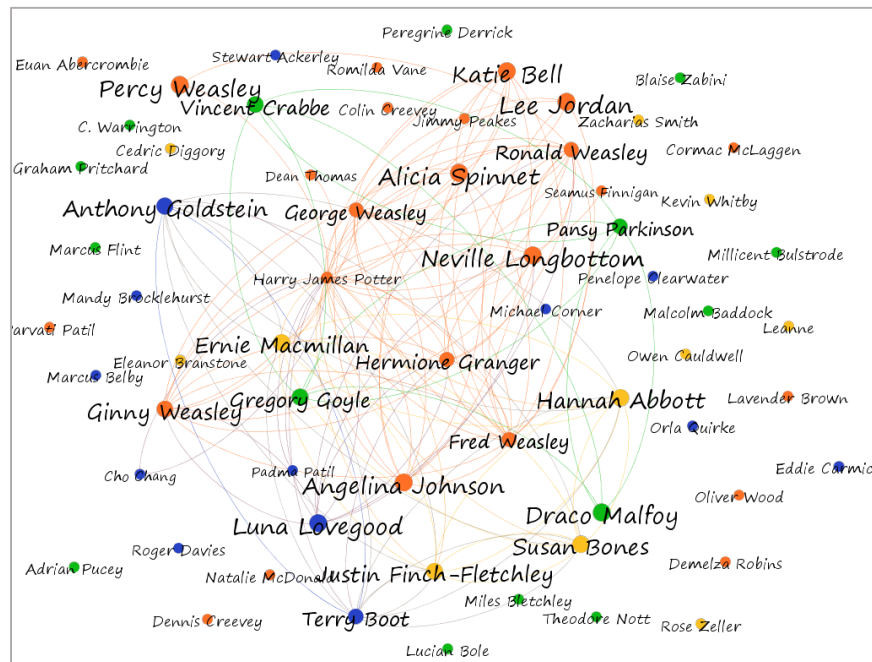
Personajes con mayor betweenness centrality Libro 5



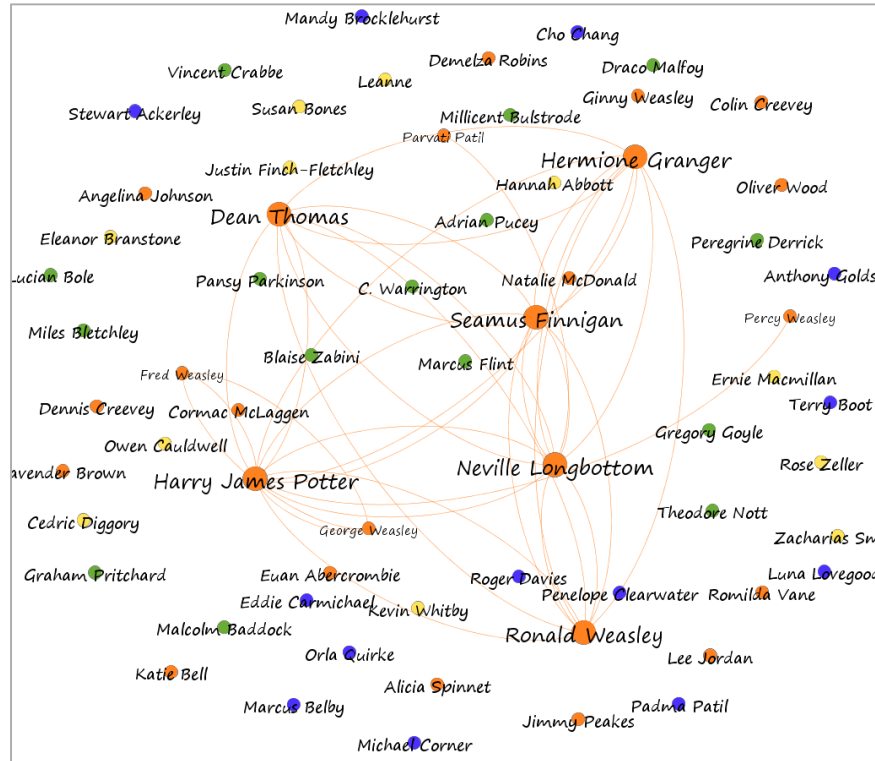
Personajes con mayor transitividad Libro 1



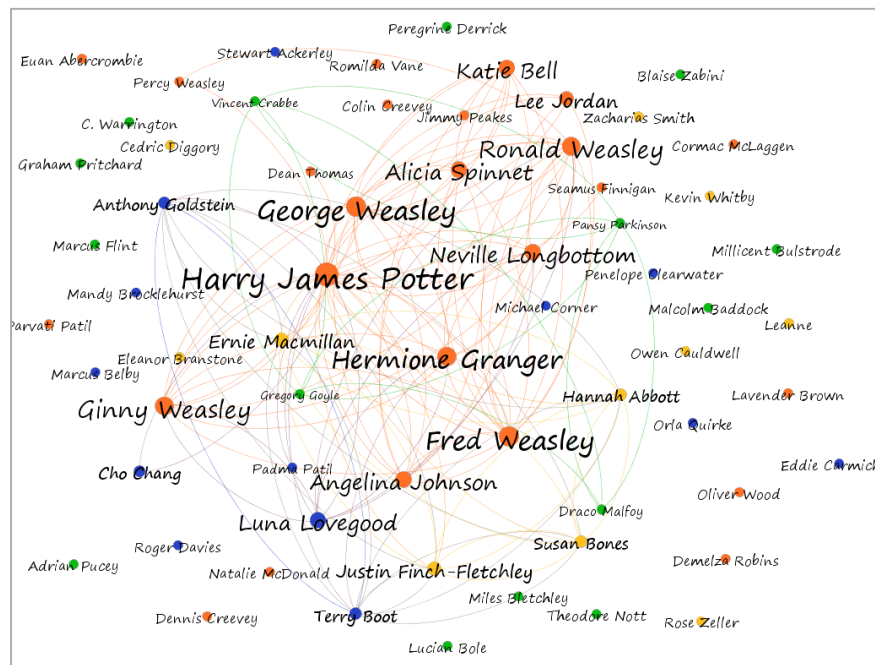
Personajes con mayor transitividad Libro 5



Personajes con mayor eigenvector centrality Libro 1



Personajes con mayor eigenvector centrality Libro 5



A l observar las gráficas anteriores, se puede concluir que Harry James Potter es un personaje muy importante para la red así como sus amigos cercanos.

- j. (2 puntos) Para todos los literales anteriores, compare la diferencia entre la red social del libro 1 y el libro 5, comentando las medidas de centralidad que considere más importantes. Concluya sobre la evolución de la estructura de la red.

A partir de las medidas de centralidad encontradas en los literales anteriores, es posible decir que de acuerdo con la centralidad de grado en cada casa desde el libro 1 al libro 5 se han venido construyendo nuevas conexiones entre los personajes, ya que para el libro 1 únicamente había centralidad de grado diferente de cero para los personajes de la casa Gryffindor. Sin embargo, al final del libro 5 todas las casas tenían al menos dos personajes muy conectados a la red. Por otro lado, la casa más densamente conectada tanto en el libro 1 como en el 5 es Gryffindor. Ahora bien, la evolución de las conexiones entre casas aumentó considerablemente entre el libro 1 y 5 y la casa Ravenclaw es aquella que se conecta más con el resto de la red. Por otro lado, Harry Potter se destaca por ser el personaje más cercano a los demás durante los dos libros, lo cual indica que es un personaje central en la red. Sin embargo, este personaje no se encuentra dentro del camino más corto entre nodos de toda la red. Finalmente, la transitividad en la red evolucionó bastante del libro 1 al libro 2 haciendo que los amigos de los personajes estén más conectados entre ellos. En conclusión, la red de los personajes de Hogwarts ha evolucionado de manera que las conexiones entre los personajes han aumentado entre casas y entre personajes, y se han instituido personajes importantes para la conexión de toda la red.

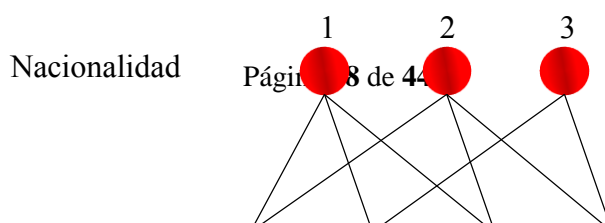
Punto 2 (30 puntos): *Fédération Internationale de Football Association (FIFA)*

La *Fédération Internationale de Football Association (FIFA)* es la organización encargada de regular el Fútbol a nivel mundial. Nació el 21 de mayo de 1904 y hoy en día cuenta con 204 asociaciones afiliadas y más de 200 millones de miembros activos. Por esto, es una de las federaciones deportivas más importantes en el mundo (FIFA, s.f.).

Adjunto encontrará un archivo .csv llamado “Fifa 2018.csv” (delimitado por punto y coma). Este contiene la información correspondiente a los clubes y sus jugadores durante el año 2018 a nivel mundial. Específicamente se tiene la siguiente información (Gupta, 2018):

- **Club:** nombre del club de Fútbol miembro de la FIFA.
- **Nacionalidad:** nacionalidad de los jugadores que componen cada club.
- **Jugadores:** cantidad de jugadores que un club tiene de una nacionalidad específica.
- **Promedio de edad:** promedio de edad de los jugadores que un club tiene de una nacionalidad específica.
- **Promedio Rating:** promedio del rating de los jugadores que un club tiene de una nacionalidad específica.
- **Promedio Altura:** promedio de la altura de los jugadores que un club tiene de una nacionalidad específica.
- **Promedio Peso:** promedio del peso de los jugadores que un club tiene de una nacionalidad específica.

La anterior información corresponde a una red bipartita (Newman, 2009) en la cual solo existen dos tipos de nodos: el primer tipo de nodos representa a los clubes y el segundo tipo representa a las nacionalidades de sus jugadores. En este tipo de redes solo hay conexiones entre los nodos de diferente tipo.



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 clubes y las columnas a las nacionalidades de los jugadores. En este caso el valor del elemento b_{ij} es:

$$b_{ij} = \begin{cases} 1 & \text{si el club } i \text{ tiene al menos un jugador de la nacionalidad } j \\ 0 & \text{de lo contrario} \end{cases}$$

Por ejemplo, para una matriz de 4 clubes y 3 nacionalidades la matriz de incidencia es:

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

De acuerdo a lo anterior, respondan el siguiente numeral:

1. **(20 puntos)** Su primera tarea es Crear la red bipartita de nacionalidades y clubes y responder a las siguientes preguntas:

Nota: Para sus respuestas indiquen cuál medida de centralidad utilizaron, muestren los valores de la medida y muestren gráficamente su resultado en la red (denle tamaño a los nodos/arcos de acuerdo a la medida de centralidad seleccionada).

- **(3 puntos)** ¿Cuáles son las 10 nacionalidades más contratadas por los clubes?

Las diez naciones más contratadas por los clubes son:

- 1 . Brazil
- 2 . France
- 3 . Argentina
- 4 . Spain
- 5 . England
- 6 . Portugal
- 7 . Uruguay
- 8 . Republic of Ireland
- 9 . Serbia
- 10 . Ghana

El valor de la centralidad de grado es:

- 1 . Brazil : 232
- 2 . France : 190
- 3 . Argentina : 170
- 4 . Spain : 138
- 5 . England : 137
- 6 . Portugal : 115
- 7 . Uruguay : 106
- 8 . Republic of Ireland : 105
- 9 . Serbia : 102
- 10 . Ghana : 101

Output Python, literal a FIFA

Para encontrar las 10 nacionalidades más contratadas por los clubes se calculó la centralidad de grado (número total de clubes que contratan a cada nacionalidad) de cada país, organizándolas de manera descendente y escogiendo las 10 primeras. Los resultados encontrados muestran que las 10 nacionalidades más contratadas, en orden descendente son: Brasil, Francia, Argentina, España, Inglaterra, Portugal, Uruguay, República de Irlanda, Serbia y Ghana.

- **(3 puntos)** ¿Cuántos jugadores son contratados para cada nacionalidad del inciso a?

Los jugadores contratados para cada nacionalidad del inciso a:

```
Brazil : 921
France : 974
Argentina : 1097
Spain : 1008
England : 1618
Portugal : 360
Uruguay : 153
Republic of Ireland : 442
Serbia : 136
Ghana : 119
```

Output Python, literal b FIFA

La cantidad de jugadores que son contratados para cada nacionalidad de las 10 más contratadas por los clubes se encontró calculando la centralidad de grado (número total de clubes que contratan a cada nacionalidad) de cada país y calculando el peso de cada arco, entendido este como el número de jugadores que un club tiene de una nacionalidad específica. Finalmente, se suman los jugadores de cada nacionalidad para cada club y se obtiene el resultado mostrado anteriormente.

- **(3 puntos)** ¿Cuál es la estatura y el peso de los jugadores contratados para cada nacionalidad del inciso a?

```
Brazil :
  Peso Promedio = 76.24888478379857
  Altura Promedio = 181.0088792792456
France :
  Peso Promedio = 76.41835713461712
  Altura Promedio = 182.38717866958862
Argentina :
  Peso Promedio = 75.977213394775
  Altura Promedio = 180.004987867339
Spain :
  Peso Promedio = 74.37891133506396
  Altura Promedio = 180.5599531272976
England :
  Peso Promedio = 75.24215689743087
  Altura Promedio = 181.58516890197825
Portugal :
  Peso Promedio = 74.71628728191386
  Altura Promedio = 180.45646892749193
Uruguay :
  Peso Promedio = 75.0685534591195
  Altura Promedio = 180.18003144654088
Republic of Ireland :
  Peso Promedio = 75.74314100565682
  Altura Promedio = 181.83380989189237
Serbia :
  Peso Promedio = 78.86764705882354
  Altura Promedio = 185.05555555555557
Ghana :
  Peso Promedio = 72.52227722772277
  Altura Promedio = 178.17821782178217
```

Output Python, literal c FIFA

La estatura y el peso de los jugadores contratados para cada nacionalidad de las 10 más contratadas por los clubes se encontró calculando la centralidad de grado (número total de clubes que contratan a cada nacionalidad) de cada país y calculando el peso de cada arco, entendido este como el número de jugadores que un club tiene de una nacionalidad específica y sus atributos estatura y peso. Finalmente, se suma la altura y peso de los jugadores de cada nacionalidad para cada club y se divide en el número total de jugadores. Así, se obtiene el resultado mostrado anteriormente.

- **(3 puntos)** ¿Cuáles son los clubes que más contratan a esas 10 nacionalidades del inciso a?

Los 5 clubes que más contratan a las 10 nacionalidades del inciso a se encontraron preguntando a los vecinos de cada nacionalidad (todos los clubes que la contratan), el número de jugadores que contratan. Dado esto, se organizan de manera descendente el número de jugadores contratados por club y se obtienen los siguientes resultados:

```
Los clubes que mas contratan a Brazil son:
1 . A. Minas G. con 20 jugadores.
1 . Atl. Paranaense con 20 jugadores.
1 . Atlético Mineiro con 20 jugadores.
1 . Avaí con 20 jugadores.
1 . Botafogo con 20 jugadores.
Los clubes que mas contratan a France son:
1 . EA Guingamp con 23 jugadores.
1 . OGC Nice con 23 jugadores.
1 . AJ Auxerre con 22 jugadores.
1 . Bourg-Péronnas con 22 jugadores.
1 . Olym. Lyonnais con 22 jugadores.
Los clubes que mas contratan a Argentina son:
1 . Belgrano con 30 jugadores.
1 . Racing Club con 30 jugadores.
1 . Fl. Varela con 29 jugadores.
1 . Quilmes con 29 jugadores.
1 . Banfield con 28 jugadores.
Los clubes que mas contratan a Spain son:
1 . Athletic Bilbao con 27 jugadores.
1 . CA Osasuna con 26 jugadores.
1 . CD Mirandés con 26 jugadores.
1 . Levante UD con 26 jugadores.
1 . Real Oviedo con 26 jugadores.
Los clubes que mas contratan a England son:
1 . AFC Wimbledon con 23 jugadores.
1 . Bournemouth con 23 jugadores.
1 . Mansfield Town con 23 jugadores.
1 . Oxford United con 23 jugadores.
1 . Burton Albion con 22 jugadores.
Los clubes que mas contratan a Portugal son:
1 . Paços Ferreira con 18 jugadores.
1 . Rio Ave con 18 jugadores.
1 . Belenenses con 17 jugadores.
1 . V. Setúbal con 16 jugadores.
1 . Boavista con 15 jugadores.
Los clubes que mas contratan a Uruguay son:
1 . Puebla con 5 jugadores.
1 . Wanderers con 4 jugadores.
1 . Atlético Madrid con 3 jugadores.
1 . Fiorentina con 3 jugadores.
1 . Godoy Cruz con 3 jugadores.
Los clubes que mas contratan a Republic of Ireland son:
1 . Bray Wanderers con 24 jugadores.
1 . St. Pats con 24 jugadores.
1 . Bohemian FC con 23 jugadores.
1 . Shamrock Rovers con 23 jugadores.
1 . Longford Town con 22 jugadores.
Los clubes que mas contratan a Serbia son:
1 . Royal Mouscron con 4 jugadores.
1 . Grasshopper con 3 jugadores.
1 . KV Kortrijk con 3 jugadores.
1 . Lazio con 3 jugadores.
1 . SL Benfica con 3 jugadores.
Los clubes que mas contratan a Ghana son:
1 . Columbus Crew SC con 4 jugadores.
1 . Aytemiz Alanyaspor con 2 jugadores.
1 . Benevento con 2 jugadores.
1 . BK Häcken con 2 jugadores.
1 . Crystal Palace con 2 jugadores.
```

Output Python, literal d FIFA

- **(3 puntos)** ¿Cuáles son los clubes que contratan a los mejores jugadores?

Los clubes que contratan a los mejores jugadores se encontraron calculando la centralidad de grado de cada club (nacionalidades de sus jugadores contratados) y adicionalmente este valor con el rating promedio de los jugadores de cada nacionalidad contratados por ese club. Así, se encontró que los clubes que contratan a los mejores jugadores son:

Los diez clubes que contratan a los mejores jugadores son:

- 1 . FC Bayern
- 2 . Juventus
- 3 . Free agent
- 4 . FC Barcelona
- 5 . PSG
- 6 . Athletic Bilbao
- 7 . Sevilla FC
- 8 . Napoli
- 9 . Real Madrid
- 10 . Arsenal

Número de jugadores contratados por club:

- 1 . FC Bayern : 9
- 2 . Juventus : 12
- 3 . Free agent : 1
- 4 . FC Barcelona : 11
- 5 . PSG : 10
- 6 . Athletic Bilbao : 2
- 7 . Sevilla FC : 8
- 8 . Napoli : 13
- 9 . Real Madrid : 14
- 10 . Arsenal : 14

Output Python, literal e FIFA

- **(3 puntos)** ¿De cuáles nacionalidades son los mejores jugadores?

La nacionalidad de los mejores jugadores se encuentran calculando la centralidad de grado de cada club (nacionalidades de sus jugadores contratados) y adicionalmente este valor con el rating promedio de los jugadores de cada nacionalidad. Posteriormente, se ordenan de forma descendente las nacionalidades según el rating promedio de sus jugadores y se seleccionan las primeras 10. Por tanto, las nacionalidades de los mejores jugadores son:

Las nacionalidades que tienen EN PROMEDIO a los mejores jugadores son:

- 1 . Chad
- 2 . São Tomé & Príncipe
- 3 . Dominican Republic
- 4 . Cuba
- 5 . Algeria
- 6 . Ecuador
- 7 . Portugal
- 8 . Argentina
- 9 . Togo
- 10 . Oman

Número de mejores jugadores por nacionalidad:

- 1 . São Tomé & Príncipe : 1
- 2 . Chad : 1
- 3 . Dominican Republic : 2
- 4 . Cuba : 2
- 5 . Algeria : 42
- 6 . Ecuador : 22
- 7 . Portugal : 115
- 8 . Argentina : 170
- 9 . Oman : 1
- 10 . Eritrea : 1

Output Python, literal f FIFA

- **(2 puntos)** Analicen los resultados de los incisos anteriores y concluyan al respecto de las características de los procesos de contratación de los clubes de la FIFA durante el 2018.

En primer lugar, es claro que los clubes con mejores resultados en sus respectivas ligas son aquellos que contratan a los mejores jugadores ('clasificados por rating'). No obstante, en los clubes grandes y exitosos se puede observar que no existe una preferencia de nacionalidad. Ahora bien, al observar los clubes pequeños de la red se observa una preferencia de conexión con la nacionalidad de la liga en la que participan. Finalmente, los mejores jugadores por nacionalidad están sesgados por pequeños outliers que tienen ratings muy altos pero no son comunes en la nacionalidad.

El 29 de noviembre de 2010, se enfrentaron FC Barcelona y Real Madrid en el Camp Nou para la Liga BBVA. El resultado del partido fue 5 – 0 a favor del equipo blaugrana quien jugaba al mando de Pep Guardiola para ese entonces (Goal, s.f). En anexos podrán encontrar las matrices de adyacencia para cada uno de los equipos durante el partido. En estas matrices, las filas y columnas corresponden a los jugadores de cada equipo y los cruces entre ellas contienen la cantidad de pases que el jugador de una fila le hizo al jugador de una columna y los que el jugador de la fila i convirtió en el gol j (Ver Matrices Barca 5 RM 0 (Liga 2010 - 2011).xlsx).

De acuerdo a lo anterior, realicen el siguiente numeral:

2. (10 puntos) Se requiere analizar el estilo de juego y los principales jugadores de cada uno de los equipos durante el partido. De acuerdo a esto, ustedes deberán realizar la red para cada equipo donde:

- Los nodos representan a los jugadores de cada equipo.
- Los arcos se forman cuando un jugador le hace al menos un pase a otro jugador. El peso asociado a estos arcos corresponde a la cantidad de pases que un jugador le hizo al otro.

Teniendo en cuenta la información dada, respondan:

- a. (2 puntos) Para cada red: ¿La red es dirigida o no dirigida? ¿Por qué?

Las redes de pase son dirigidas ya que un pase tiene un emisor y un receptor. De esta manera, cada pase entre jugadores es intencional por parte del emisor y debe ser recibido por otro jugador dependiendo de su posición. Así, las redes de los pases en Barcelona y Real Madrid fueron construidas a partir de un dataframe en pandas usando la clase digraph de networkx.

- b. (2 puntos) Para cada red: ¿Quiénes fueron los futbolistas más importantes del equipo y por qué? Utilicen las medidas que consideren adecuadas para justificar su respuesta.

Los futbolistas más importantes pueden entenderse desde 2 puntos de vista: (1) Los que más generan futbol y (2) los que más generan opciones de gol porque más reciben pases. Para ambos equipos se elaboró un ranking para ambas medidas usando el *out* y el *in degree* de cada uno de los jugadores.

Real Madrid

Las jugadores que mas hacen pases en el Real Madrid son:

- 1 . Xabi con 36 pases.
- 2 . Sergio Ramos con 32 pases.
- 3 . Pepe con 27 pases.

Output Python, literal b Pases Out

Las jugadores que mas reciben pases en el Real Madrid son:

- 1 . C. Ronaldo con 35 pases.
- 2 . Xabi con 30 pases.
- 3 . Di María con 26 pases.

Output Python, literal b Pases In

Barcelona

Las jugadores que mas hacen pases en el Barcelona son:

- 1 . Xavi Hernandez con 114 pases.
- 2 . Sergio Busquets con 84 pases.
- 3 . A. Iniesta con 72 pases.

Output Python, literal b Pases

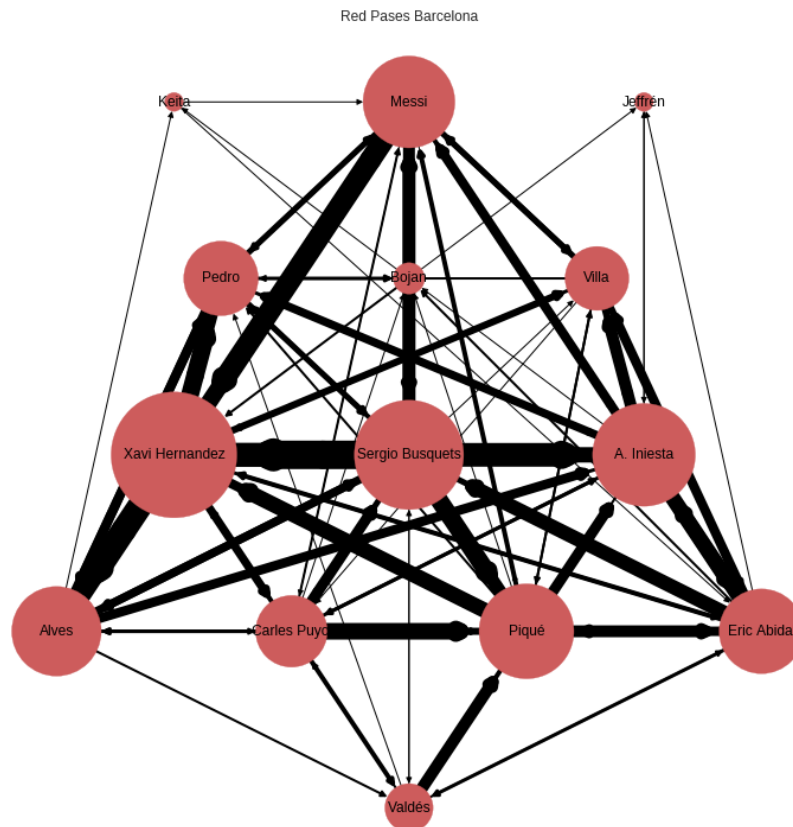
Las jugadores que mas reciben pases en el Barcelona son:

- 1 . Xavi Hernandez con 110 pases.
- 2 . Sergio Busquets con 85 pases.
- 3 . A. Iniesta con 77 pases.

Output Python, literal b Pases

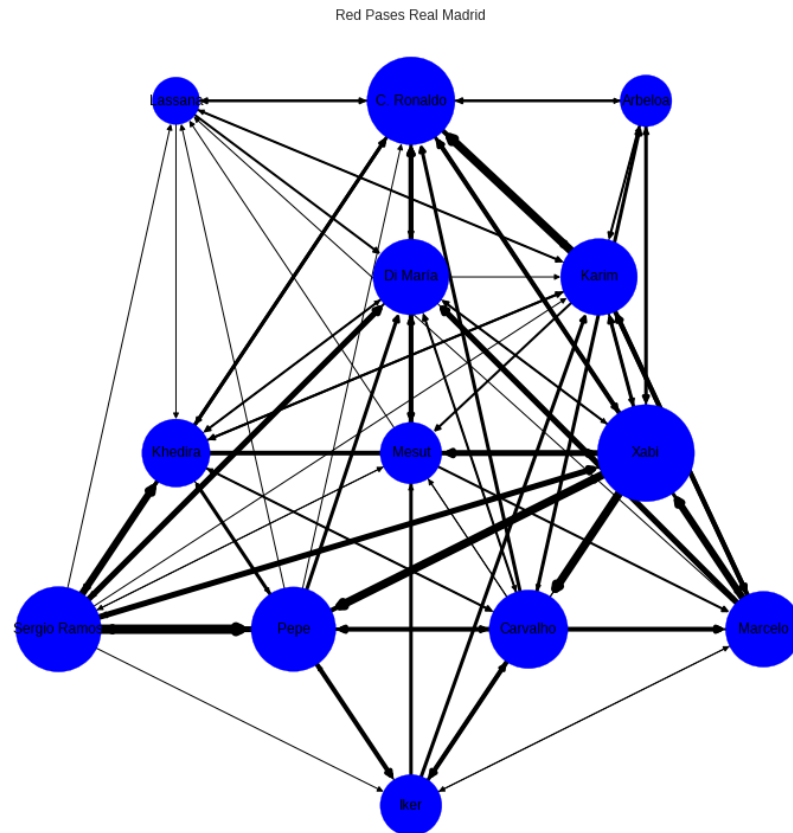
- c. **(2 puntos)** Presenten y analicen una gráfica de la red de pases de cada uno de los equipos donde se haga evidente quiénes fueron los jugadores que más dieron y recibieron pases.

Red de pases F.C. Barcelona



A partir de la red se puede ver que los nodos más grandes son aquellos jugadores que reciben o ejecutan más pases a sus compañeros de equipo. En este caso se destacan Xavi Hernández, Sergio Busquets y A. Iniesta.

Red de pases Real Madrid



A partir de la red se puede ver que los nodos más grandes son aquellos jugadores que reciben o ejecutan más pases a sus compañeros de equipo. En este caso se destacan Xabi Hernández, C. Ronaldo y Sergio Ramos.

- d. **(2 puntos)** Si el Real Madrid quisiera desarticular el juego del FC Barcelona, ¿a quién deberían marcar y por qué?

Dados los resultados del inciso B, el real Madrid debería marcar muy bien a Xavi Hernández e interceptar el juego del Barcelona. En este sentido, si el real Madrid controla los balones que llegan a Xavi Hernández se evita la formación de una jugada ofensiva del Barcelona. Por otro lado, si luego evita que Xavi Hernández pueda realizar pases el Barcelona posiblemente se le dificulte llegar al arco del Real Madrid.

- e. (2 puntos) ¿Qué se puede concluir acerca del estilo de juego del FC Barcelona y qué recomendaciones se le daría a un equipo que posiblemente se fuera a enfrentar a él, de manera que no le pase lo mismo que al Real Madrid? Utilicen las medidas de centralidad que consideren adecuadas para justificar su respuesta.

El estilo de juego del Barcelona se puede resumir a modo de comentarista como "Mucho toque toque". En particular, al analizar la centralidad de grado se observa que la mayoría de sus jugadores participan activamente en el juego mediante el movimiento continuo del balón entre jugadores. Observamos que la mayoría del juego se concentra en la parte media del campo y luego hay pases claves a los delanteros del equipo. Este "Toque toque" permite que el contrincante se canse y que el balón fluya con mayor rapidez de arco a arco. Como consecuencia, los equipos que enfrenten en futuras ocasiones al Barcelona deben preparar una defensiva clara que intercepte los pases generados por los mediocampistas.

Punto 3 (35 puntos): Análisis de las redes de prostitución en Brasil

Los patrones de contacto sexual pueden influir en la propagación de enfermedades de transmisión sexual (ETS). Ustedes han sido contratados por el gobierno de Brasil para analizar el comportamiento de las redes sociales creadas en torno a la prostitución en varios pueblos del país. Para esto, ustedes cuentan con la base de datos de prostitución tomada de un foro público donde, después de un encuentro sexual, hombres heterosexuales (llamados compradores de sexo) califican y hablan acerca de su experiencia con prostitutas (Rocha, Liljeros, & Holme, 2011).

Desde la perspectiva de redes sociales, este foro puede analizarse como una red bipartita (ver Punto 2) donde se conectan los compradores de sexo con las prostitutas a las cuales han contratado. De acuerdo a esto:

- Los nodos representan a las prostitutas y a los compradores de sexo.
- Los arcos se generan cuando una prostituta y un comprador de sexo han tenido sexo.

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 prostitutas están conectadas si tienen al menos un comprador en común.

Entonces la proyección sería:

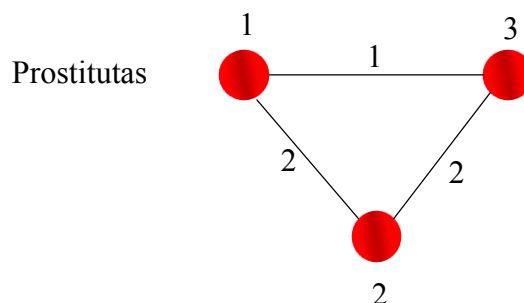


Ilustración 2 Proyección de la red bipartita para prostitutas

Por ejemplo: las prostitutas 1 y 3 tienen 1 comprador en común y las prostitutas 2 y 3 tienen 2 compradores en común.

Por otro lado, se podría entender la red desde una proyección que refleje el punto de vista de los compradores de sexo, donde cada conexión entre compradores representa la existencia de prostitutas en común:

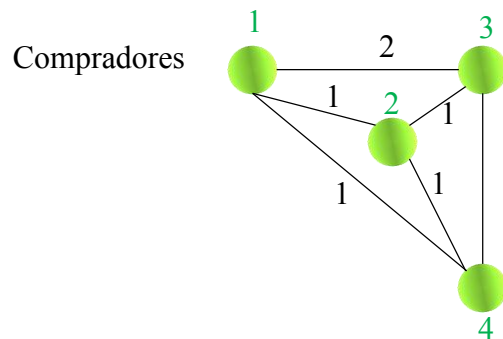


Ilustración 3 Proyección de la red bipartita para compradores

Por ejemplo: los compradores 1 y 3 tienen 2 prostitutas en común y los compradores 2 y 4 tienen una prostituta en común.

Como se pudo ver, en las proyecciones anteriores, cada arco entre prostitutas/compradores tiene un peso asociado a la cantidad de prostitutas/compradores que tenían en común en la red bipartita.

Para hacer la primera proyección desde la matriz de incidencia a una matriz de adyacencia debemos (Newman, 2009):

- I. Calcular la matriz $P = B^T B$
- II. Igualar los valores de la diagonal de la matriz P a cero.

Para hacer la segunda proyección desde la matriz de incidencia a una matriz de adyacencia debemos (Newman, 2009):

- I. Calcular la matriz $P = B B^T$
- II. Igualar los valores de la diagonal de la matriz P a cero.

En el archivo adjunto “Dataset_S1.csv” encontrarán la información correspondiente al foro y la descripción de cada una de las variables que lo componen (KONECT, 2017).

De acuerdo a lo anterior, realicen lo siguiente:

1. **(10 puntos)** Su primera tarea es realizar las proyecciones en Python de la red bipartita donde se evidencien:

Para construir las proyecciones en Python, en primer lugar se crea la red bipartita y se calcula su matriz de adyacencia, dicha matriz está conformada únicamente por 0 y 1, por lo cual es binaria y 1 significa que el hombre x ha tenido sexo con la mujer y, 0 de lo contrario:

	female0	female2	female3	female5	female7	female9	female11	female13	female15	female17	...	female16714	female16716	female16718	female16720
male1	1	1	1	0	0	0	0	0	0	0	...	0	0	0	0
male4	1	0	0	0	0	0	0	0	0	0	...	0	0	0	0
male6	0	0	0	1	0	0	0	0	0	0	...	0	0	0	0
male8	0	0	0	0	1	0	0	0	0	0	...	0	0	0	0
male10	0	0	0	0	0	1	0	0	0	0	...	0	0	0	0
male12	1	0	0	0	0	0	1	1	0	1	...	0	0	0	0
male14	0	0	0	0	1	0	0	0	0	0	...	0	0	0	0
male16	0	0	0	0	0	0	0	0	1	0	...	0	0	0	0
male19	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
male20	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0

Posteriormente, se halla la matriz transpuesta de la matriz de adyacencia:

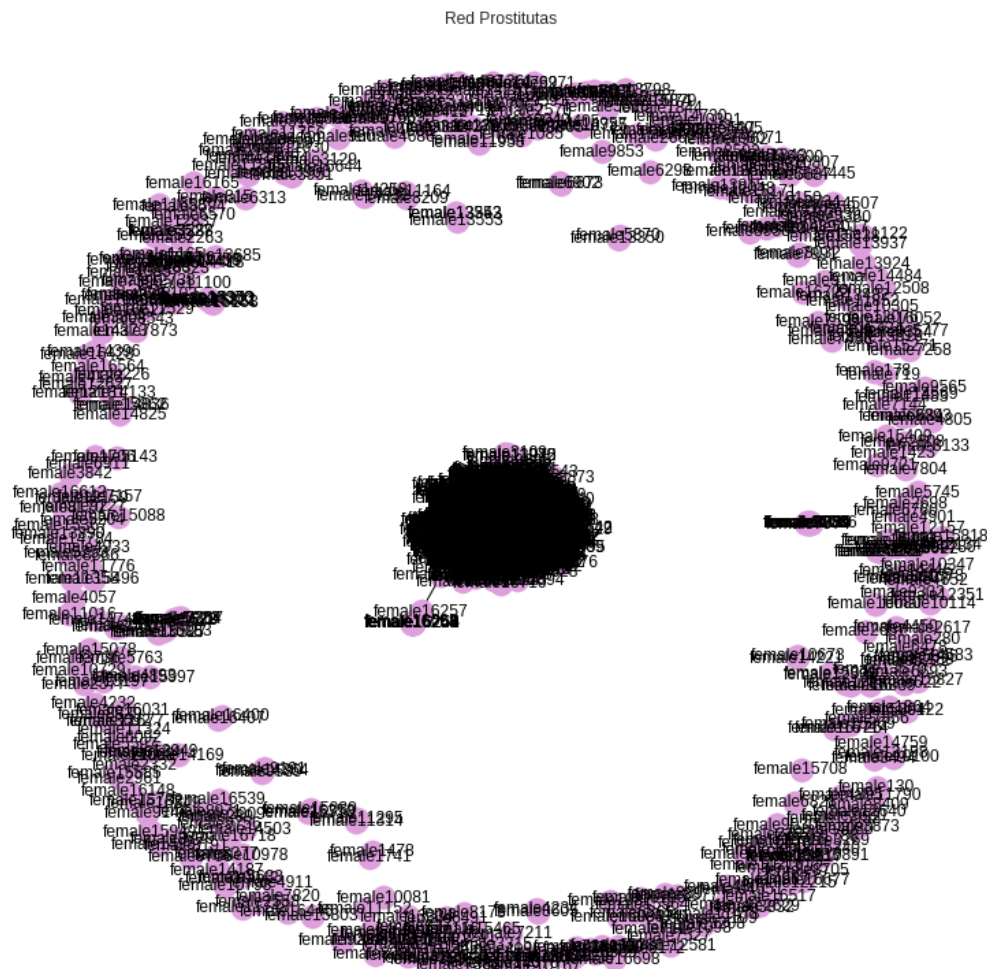
	male1	male4	male6	male8	male10	male12	male14	male16	male19	male20	...	male16706	male16707	male16708	male16709	male16715	male16717
female0	1	1	0	0	0	1	0	0	0	0	...	0	0	0	0	0	0
female2	1	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
female3	1	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
female5	0	0	1	0	0	0	0	0	0	0	...	0	0	0	0	0	0
female7	0	0	0	1	0	0	1	0	0	0	...	0	0	0	0	0	0
female9	0	0	0	0	1	0	0	0	0	0	...	0	0	0	0	0	0
female11	0	0	0	0	0	1	0	0	0	0	...	0	0	0	0	0	0
female13	0	0	0	0	0	1	0	0	0	0	...	0	0	0	0	0	0
female15	0	0	0	0	0	0	0	1	0	0	...	0	0	0	0	0	0
female17	0	0	0	0	0	1	0	0	0	0	...	0	0	0	0	0	0

- a. **(5 puntos)** La red de compradores sexuales que tuvieron sexo con las mismas prostitutas. Es decir, una red donde los nodos representan a los compradores de sexo, las aristas se forman cuando los compradores tienen prostitutas en común y los pesos de las aristas corresponden a la cantidad de prostitutas que los compradores tienen común.

Para crear la red de compradores sexuales que tuvieron sexo con las mismas prostitutas, se multiplica la matriz de adyacencia transpuesta por la matriz de adyacencia y los valores de la diagonal de la matriz resultante se igualan a cero. Así, la matriz de adyacencia de esta proyección es:

	female0	female2	female3	female5	female7	female9	female11	female13	female15	female17	...	female16714	female16716	female16718	female16720
female0	0	7	1	0	0	1	6	1	0	2	...	0	0	0	0
female2	7	0	1	0	0	3	2	0	1	1	...	0	0	0	0
female3	1	1	0	0	0	0	0	0	0	0	...	0	0	0	0
female5	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
female7	0	0	0	0	0	1	0	0	0	0	...	0	0	0	0
female9	1	3	0	0	1	0	3	0	0	0	...	0	0	0	0
female11	6	2	0	0	0	3	0	1	0	2	...	0	0	0	0
female13	1	0	0	0	0	0	1	0	0	1	...	0	0	0	0
female15	0	1	0	0	0	0	0	0	0	0	...	0	0	0	0
female17	2	1	0	0	0	0	2	1	0	0	...	0	0	0	0
female18	3	3	0	0	1	14	3	0	1	0	...	0	0	0	0
female21	1	2	1	0	0	6	0	0	2	0	...	0	0	0	0
female22	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
female24	2	1	0	0	0	0	0	0	0	0	...	0	0	0	0
female26	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0

A partir de la matriz de adyacencia se crea la proyección de la red bipartita. Esta red, tiene 6624 nodos, 183465 arcos y su grado promedio es de 55.3940. Gráficamente la red se puede ver así:



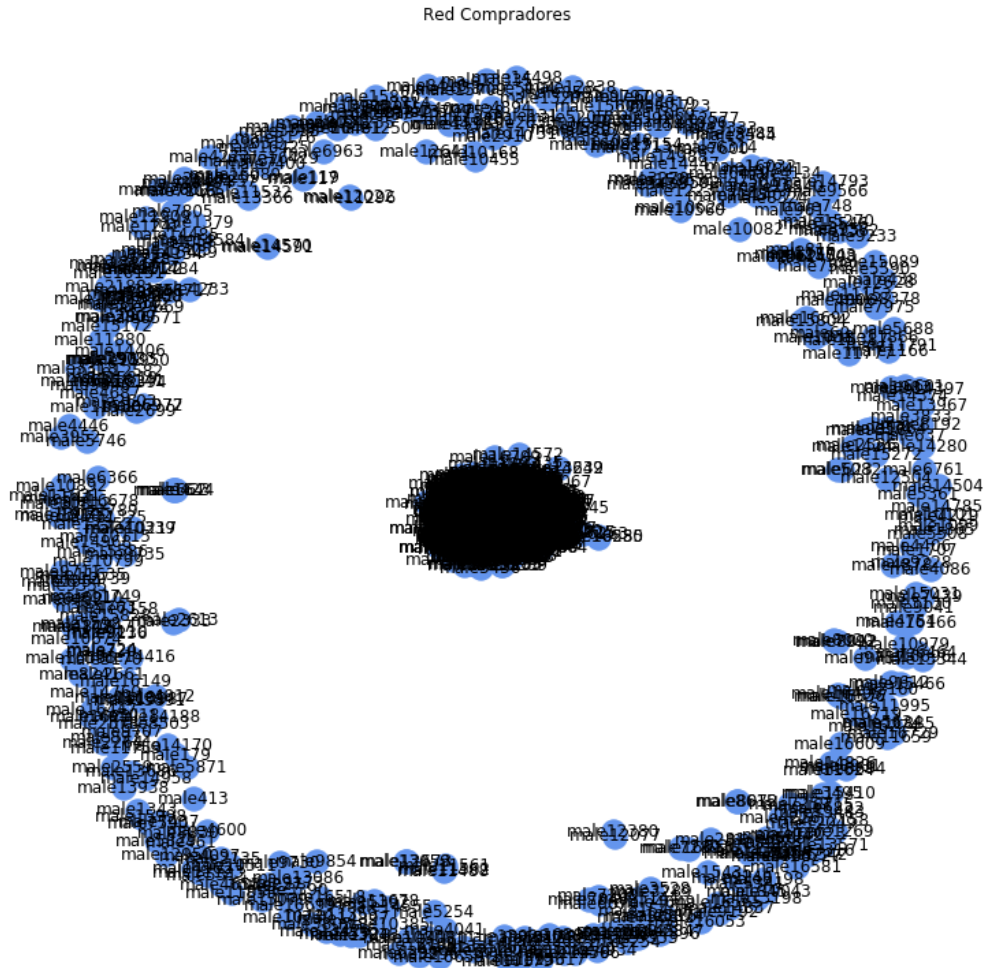
- b. (5 puntos) La red de prostitutas que tuvieron sexo con los mismos compradores sexuales. Es decir, una red donde los nodos representan a las prostitutas, las aristas

se forman cuando las prostitutas tienen compradores en común y los pesos de las aristas corresponden a la cantidad de compradores que las prostitutas tienen común.

Para crear la red de prostitutas que tuvieron sexo con los mismos compradores sexuales., se multiplica la matriz de adyacencia por la matriz de adyacencia transpuesta y los valores de la diagonal de la matriz resultante se igualan a cero. Así, la matriz de adyacencia de esta proyección es:

	male1	male4	male6	male8	male10	male12	male14	male16	male19	male20	...	male16706	male16707	male16708	male16709	male16715	male16717
male1	0	1	0	0	0	1	0	1	1	0	...	0	0	0	0	0	0
male4	1	0	0	0	0	1	0	0	0	0	...	0	0	0	0	0	0
male6	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
male8	0	0	0	0	0	0	1	0	0	0	...	0	0	0	0	0	0
male10	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
male12	1	1	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
male14	0	0	0	1	0	0	0	0	0	0	...	0	0	0	0	0	0
male16	1	0	0	0	0	0	0	0	1	0	...	0	0	0	0	0	0
male19	1	0	0	0	0	0	0	1	0	1	...	0	0	0	0	0	0
male20	0	0	0	0	0	0	0	0	1	0	...	0	0	0	0	0	0
male23	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
male25	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
male27	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
male28	0	0	0	0	0	1	0	0	0	0	...	0	0	0	0	0	0
male29	0	0	0	0	0	0	0	0	1	1	...	0	0	0	0	0	0

A partir de la matriz de adyacencia se crea la proyección de la red bipartita. Esta red, tiene 10106 nodos, 668183 arcos y su grado promedio es de 132.2349. Gráficamente la red se puede ver así:



2. **(15 puntos)** Su segunda tarea es analizar la topología de cada red del literal anterior. Para esto usted debe:

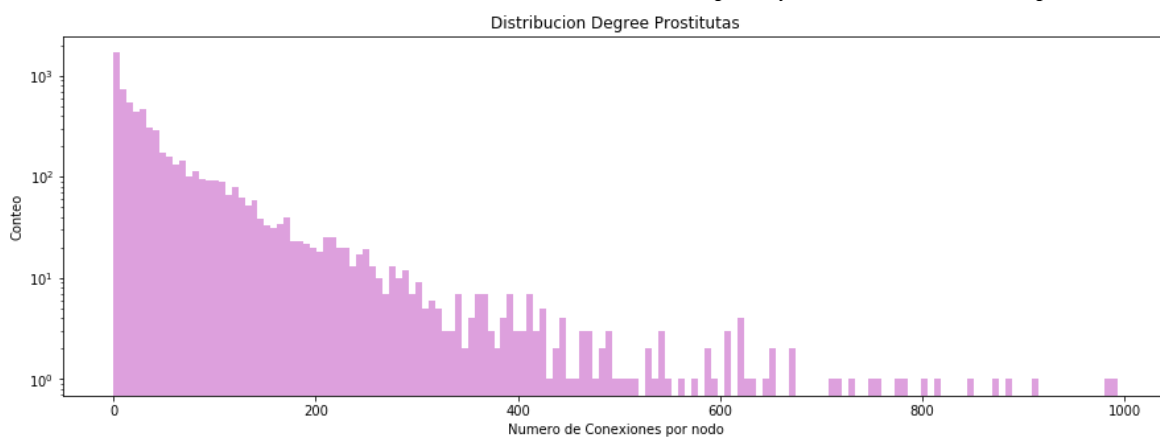
- (4 puntos)** Calcular y analizar el grado promedio con y sin pesos y explicar la diferencia entre ambas medidas.

Red de prostitutas

Análisis del grado promedio con pesos

```
female4585 : 993
female4569 : 984
female4200 : 913
female7714 : 887
female10964 : 874
female18 : 844
female98 : 812
female11717 : 803
female6813 : 782
female8344 : 777
```

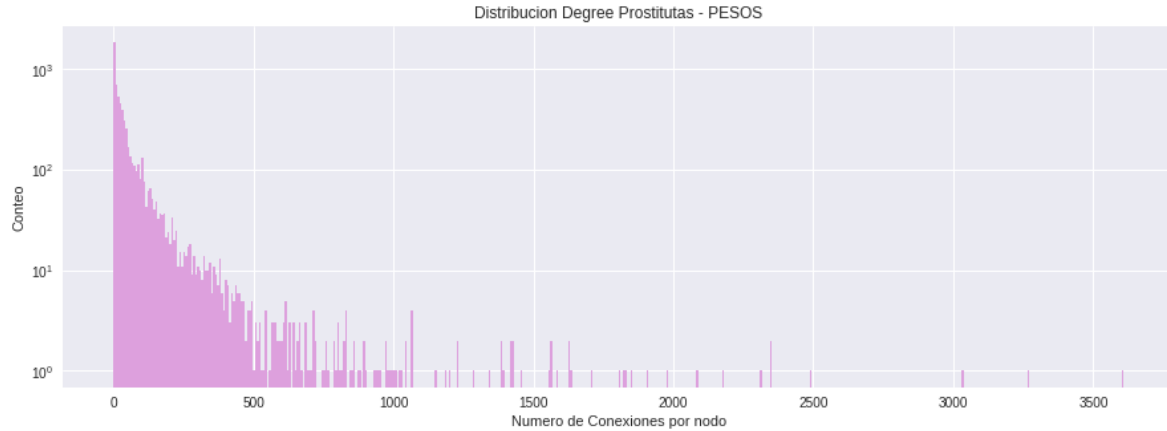
Output Python, literal a Redes prostitución



Análisis del grado promedio sin pesos

```
female4569 : 3607
female4585 : 3267
female4200 : 3031
female7714 : 2492
female6813 : 2349
female98 : 2348
female11717 : 2314
female10964 : 2180
female18 : 2084
female7730 : 1981
```

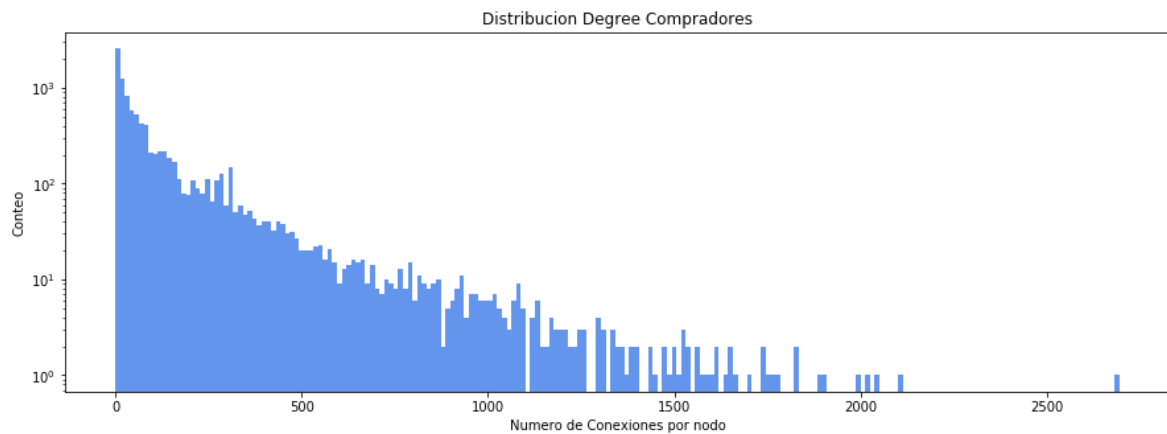
Output Python, literal a Redes prostitución



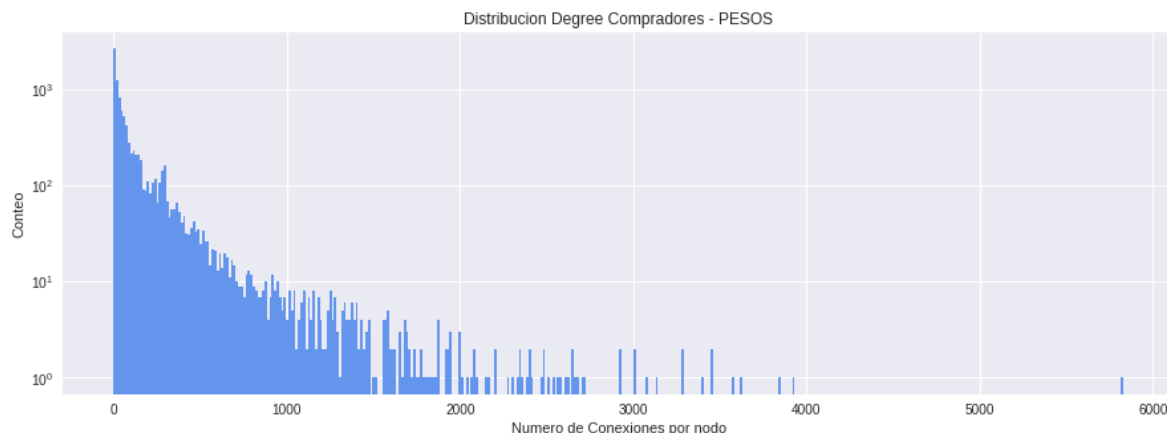
A partir de lo anterior se puede evidenciar la diferencia entre cálculo del grado con pesos y sin pesos ya que el último no incluye el número de compradores que comparten las prostitutas y, por tanto, cambia el valor esta medida de centralidad

Red de compradores

Análisis del grado promedio sin pesos



Análisis del grado promedio con pesos

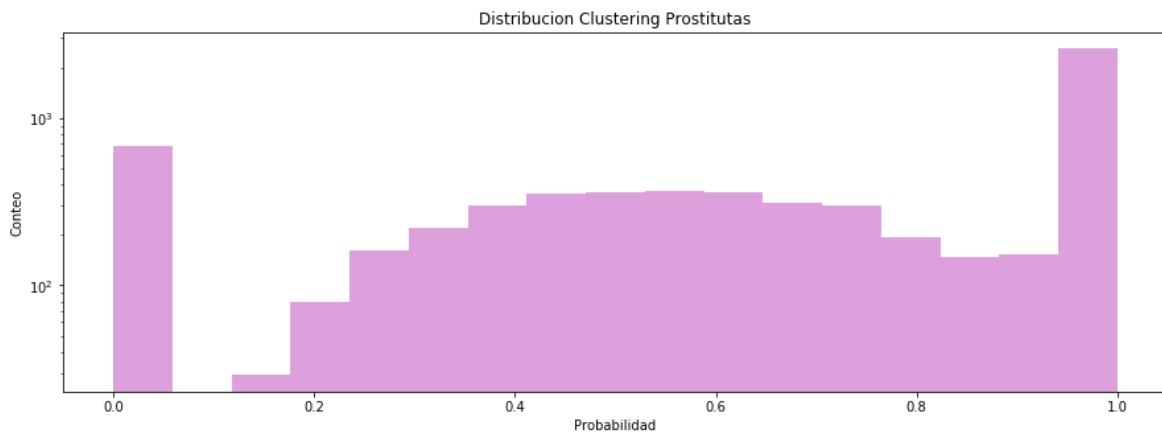


A partir de lo anterior se puede evidenciar la diferencia entre cálculo del grado con pesos y sin pesos ya que el último no incluye el número de prostitutas que comparten los compradores y, por tanto, cambia el valor esta medida de centralidad

- b. **(4 puntos)** Calcular y analizar las medidas de centralidad sin pesos: coeficiente de clustering promedio, betweenness, diámetro y densidad.

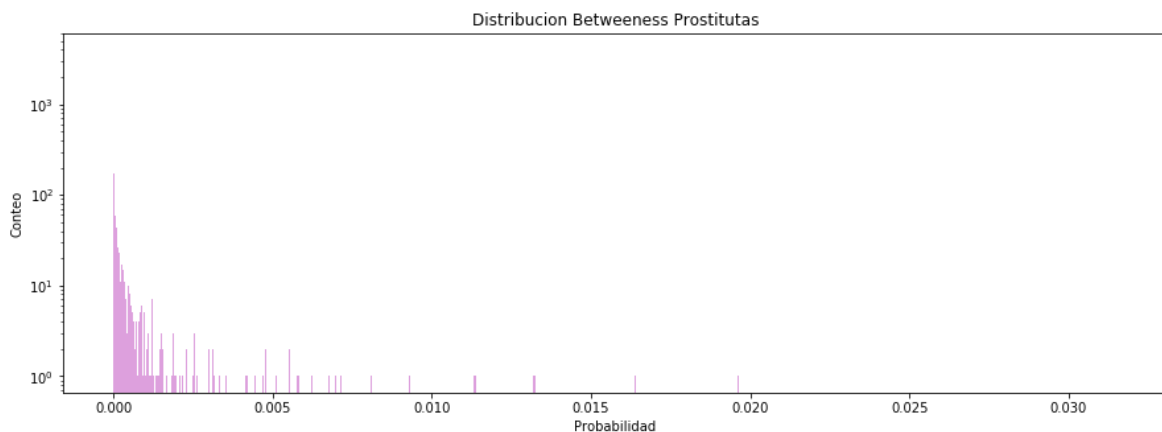
Red de prostitutas

Coeficiente de clustering promedio



El clustering promedio indica el número de triángulos en el grafo, es decir, un grupo de 3 prostitutas que tienen sexo con los mismos clientes. Así pues, como se observa en la gráfica anterior, la mayoría de prostitutas tienen un clustering promedio igual a 1, lo cual indica que muchas de ellas tienen clientes en común. De igual forma, se puede ver que la mayoría de prostitutas tienen un clustering diferente de cero por lo que no están aisladas y comparten variedad de compradores entre ellas.

Betweenness



Las prostitutas que se encuentran más veces dentro del camino más corto entre todos los nodos de la red se encontraron con la medida de centralidad intermediación o betweenness centrality. Esta, es la frecuencia con la que un nodo aparece en el camino más corto entre todos los pares de nodos de la

red. Es decir, muestra qué nodos actúan como 'puentes' entre nodos en una red (Disney A., 2014). En este contexto, se puede ver que esta medida de centralidad es muy pequeña para la mayoría de las prostitutas, lo cual indica que no hay prostitutas que sean focos, sino que se puede acceder a ellas desde cualquier punto de la red sin acudir a una prostituta central o más conectada.

Diámetro

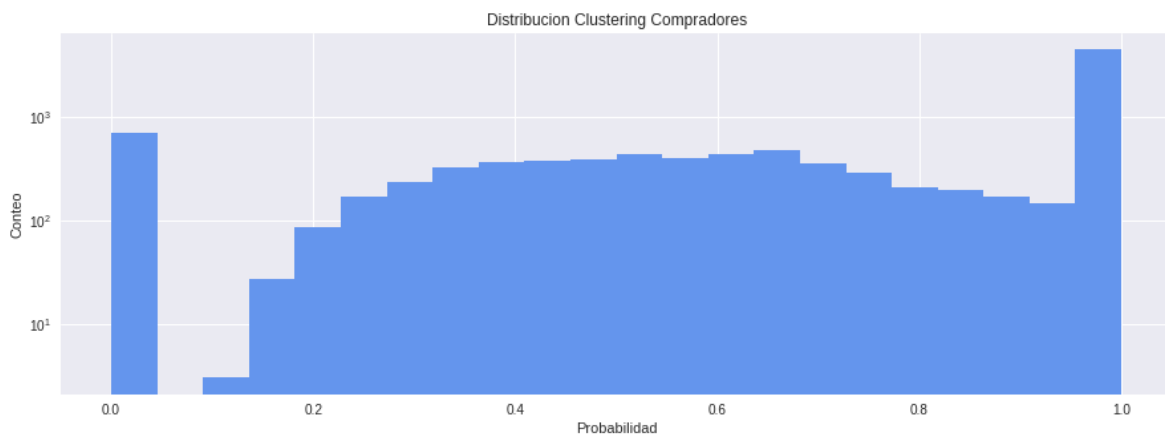
El diametro del subgrafo conexo mas grande es:
8

Teniendo en cuenta que el diámetro es la longitud de la distancia geodésica más larga entre cualquier par de nodos en la red. Es posible decir que la distancia de una prostituta a cualquier otra es de 8 prostitutas entre ellas.

Densidad.

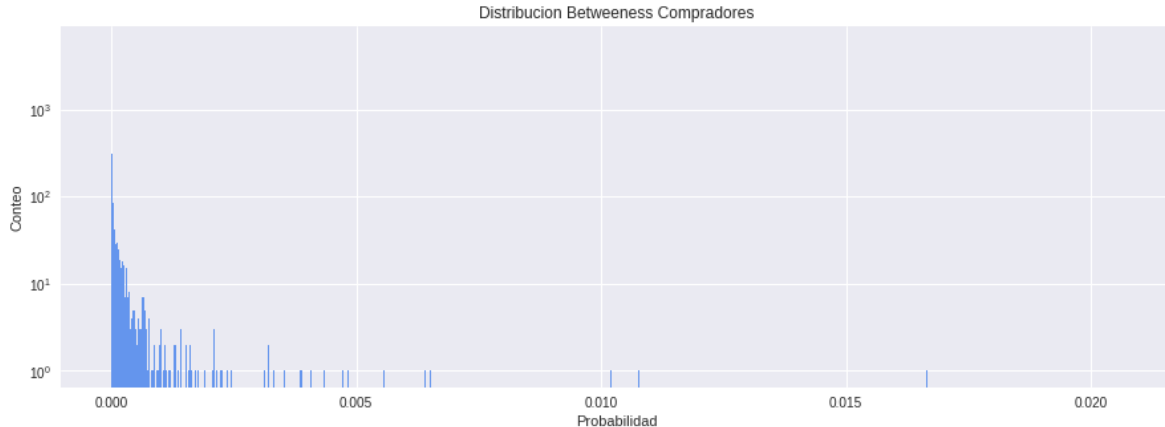
Red de compradores

Coefficiente de clustering promedio



El clustering promedio indica el número de triángulos en el grafo, es decir, un grupo de 3 compradores que tienen sexo con las mismas prostitutas. Así pues, como se observa en la gráfica anterior, la mayoría de compradores tienen un clustering promedio igual a 1, lo cual indica que muchas de ellas tienen clientes en común. De igual forma, se puede ver que la mayoría de clientes tienen un clustering diferente de cero por lo que no están aislados y comparten variedad de prostitutas entre ellos. A su vez, se puede notar que en comparación con la red de prostitutas, el número de compradores que comparten prostitutas es muy superior al número de prostitutas que comparten compradores. Esto se debe al número de nodos de la red y sus conexiones.

Betweenness



Los compradores que se encuentran más veces dentro del camino más corto entre todos los nodos de la red se encontraron con la medida de centralidad intermediación o betweenness centrality. Esta, es la frecuencia con la que un nodo aparece en el camino más corto entre todos los pares de nodos de la red. Es decir, muestra qué nodos actúan como 'puentes' entre nodos en una red (Disney A., 2014). En este contexto, se puede ver que esta medida de centralidad es muy pequeña para la mayoría de los compradores, lo cual indica que no hay compradores que sean focos, sino que las prostitutas pueden ofrecer sus servicios desde cualquier nodo de la red sin acudir a un comprador que sea más central o esté más conectado.

Diámetro

El diametro del subgrafo conexo mas grande es:
8

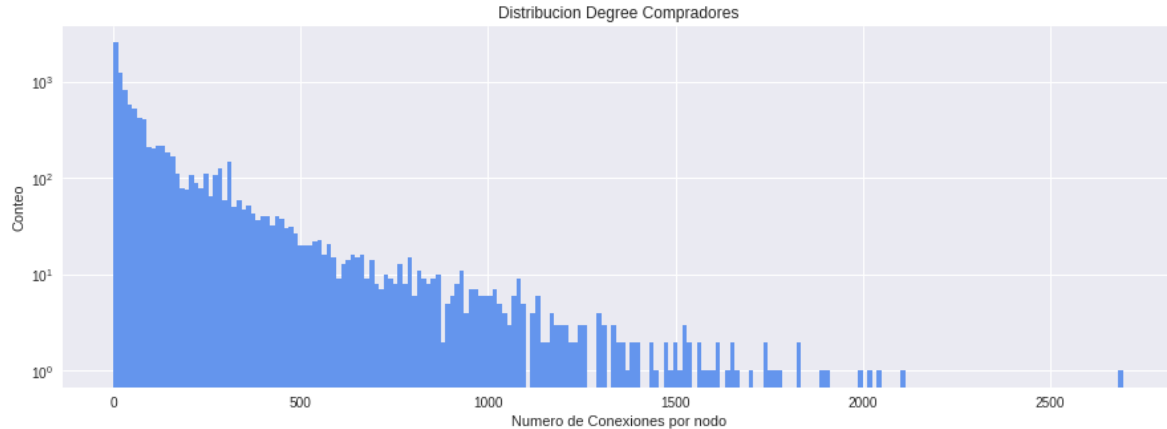
Teniendo en cuenta que el diámetro es la longitud de la distancia geodésica más larga entre cualquier par de nodos en la red. Es posible decir que la distancia de un comprador a cualquier otro es de 8 compradores de la misma manera que en la red de prostitutas.

Densidad.

- c. **(4 puntos)** Estimar la distribución del grado (sin pesos) de cada red (a partir de un fit polinomial o una prueba chi cuadrado o kolmogorov-smirnov).

Red de compradores

Para estimar la distribución del grado sin pesos se retoma el histograma del grado promedio de la red de compradores:



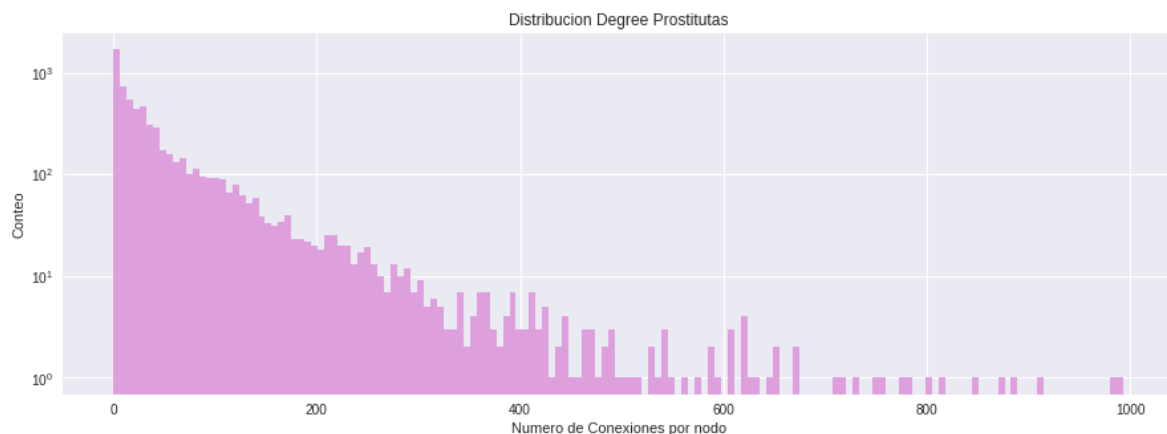
A partir de este, se realiza un ajuste para diversas distribuciones. Luego, por máxima verosimilitud se encuentra cuál de ellas se ajusta mejor:

```
Estamos probando las siguientes distribuciones...
---> laplace
---> norm
---> expon
---> dweibull
---> lognorm
---> gamma
---> powerlaw
---> t
La Distribucion que mejor se ajusta es --> powerlaw -- MLE value: 50933.76155413734
```

Así, se encontró que la distribución que mejor se ajusta a la distribución del grado de la red de compradores es una powerlaw.

Red de prostitutas

Para estimar la distribución del grado sin pesos se retoma el histograma del grado promedio de la red de prostitutas:



A partir de este, se realiza un ajuste para diversas distribuciones. Luego, por máxima verosimilitud se encuentra cuál de ellas se ajusta mejor:

```
Estamos probando las siguientes distribuciones...
---> laplace
---> norm
---> expon
---> dweibull
---> lognorm
---> gamma
---> powerlaw
---> t
La Distribucion que mejor se ajusta es --> powerlaw -- MLE value: 18720.997310755847
```

Así, se encontró que la distribución que mejor se ajusta a la distribución del grado de la red de prostitutas es una powerlaw.

- d. **(3 puntos)** Asociar cada red con alguno de los modelos de redes aleatorias vistos en clase: Erdos-Rényi, Wats-Strogatz y Barabasi-Albert. Para esto usted debe soportar su decisión con los resultados de los literales a, b y c.

En primer lugar, dado que la distribución que mejor se ajusta a la distribución del grado de las dos redes es powerlaw, se presume que es una red Barabasi-Albert ya que estas redes se caracterizan por tener distribuciones de grado que siguen leyes de potencias:

$$P(k) = k^{-\gamma}$$

Donde $\gamma > 0$

Por tanto, como la red de Brabasi-Albert es una red de preferencia se puede decir que aquellas prostitutas que son más deseadas tienden a ser más deseadas y los compradores que tienen más sexo tienden a tener todavía más sexo. Esto con base en las medidas de centralidad halladas en los literales a y b, ya que las distribuciones de grado se comportan de manera similar y las otras medidas de centralidad muestran que son redes libres de escala, por lo cual cumple las características de una red Barabasi-Albert.

3. **(5 puntos)** ¿Podría inferir la existencia de una red de prostitución consolidada con la información suministrada? Justifique su respuesta.

Nota: Una red de prostitución consolidada sería una red donde las prostitutas comparten los mismos compradores o donde los compradores tienden a contratar a las mismas prostitutas. Se podría ver como un subgrafo conexo de la red o como pequeñas redes aisladas.

```
El componente conexo mas grande tiene el 92.96497584541062 % de los nodos.
```

Output Python Red de prostitución consolidada

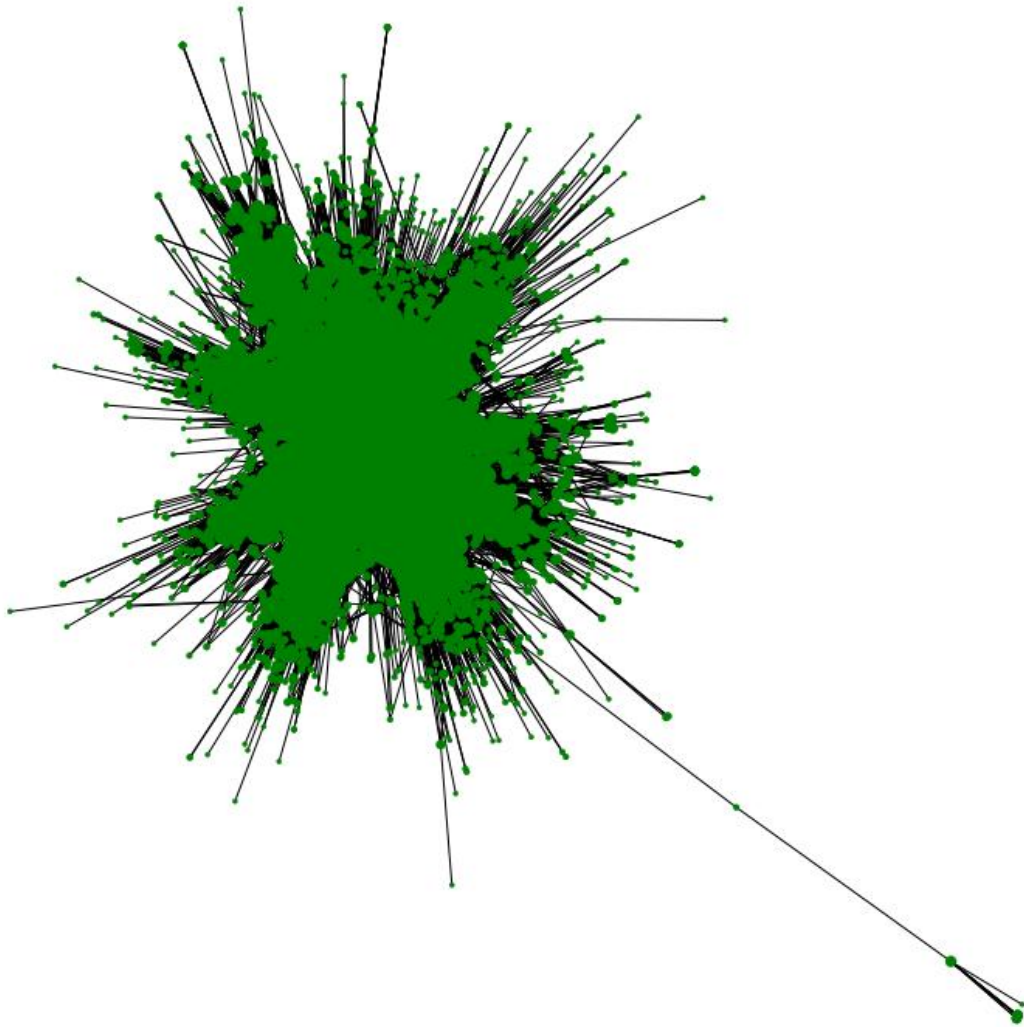
Sí se puede inferir la existencia de una red de prostitución consolidada, ya que el componente conexo más grande tiene el 92.96% de los nodos de la proyección de la red de prostitutas. Lo anterior implica que, el componente conexo más grande maneja a casi todas las prostitutas de los pueblos analizados en el estudio. Lo cual, podría implicar que los clientes acceden a ellas a través de un intermediario ya que su conexión es muy grande.

4. **(5 puntos)** Según el resultado del literal anterior, recomiende una estrategia de destrucción de la red de prostitución en Brasil y otra estrategia para evitar la propagación de ETS a lo largo de la red.

Para destruir la red de prostitución en Brasil se recomienda atacar los nodos más importantes dentro de cada red, tanto prostitutas como compradores. Esto provocará que las conexiones entre ellos desaparezcan y por tanto, las conexiones dentro de la red bipartita también para que ya no se generen oferta y demanda en el mercado de la prostitución. Por otro lado, para evitar la propagación de ETS a lo largo de la red es fundamental influir en las decisiones de las prostitutas y clientes con mayor grado para que se utilicen diferentes métodos con el fin de evitar ETS, aunque también con la destrucción de la red mencionada anteriormente aumentaría la efectividad de una política orientada a disminuir los riesgos de la propagación de enfermedades de transmisión sexual.

5. **Bono (10 puntos):** Grafique en Python el componente conexo más grande de cada proyección, donde se evidencie la topología de la red y los nodos más influyentes.

Componente Conexo mas grande



Bibliografía

- 11tegen11. (2018). *Chelsea vs Barcelona: positions and passing network*. Retrieved from <https://tribuna.com/en/fcbarcelona/news/2615179/>
- Bossaert, G., & Meidert, N. (2013). *Data files and variables for the Harry Potter support networks of Goele Bossaert and Nadine Meidert*. Retrieved from <https://www.stats.ox.ac.uk/~snijders/siena/HarryPotterData.html>
- FIFA. (n.d.). *History*. Retrieved from <https://es.fifa.com/about-fifa/who-we-are/history/index.html>
- Goal. (s.f). *A 7 años del Barcelona 5-0 Real Madrid: la cumbre del Guardiolismo*. Retrieved from <http://www.goal.com/es-co/noticias/a-7-anos-del-barcelona-5-0-real-madrid-la-cumbre-del/1bexvf2nrugxa1njazr7g0esju>

- Gupta, R. (2018). *FIFA 2018*. Retrieved from <https://www.kaggle.com/rohit782192/fifa-2018>
- KONECT. (2017). *Sexual escorts network dataset*. Retrieved from <http://konect.uni-koblenz.de/networks/escorts>
- Newman, M. (2009). *Networks: An introduction*. University of Michigan and Santa Fe Institute.
- Pottermore. (n.d.). Retrieved from <https://www.pottermore.com/>
- Rocha, L., Liljeros, F., & Holme, P. (2011, Marzo 17). *Simulated Epidemics in an Empirical Spatiotemporal Network of 50,185 Sexual Contacts*. Retrieved from <http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1001109#s2>