

Análisis y Visualización Básica de una Red con Gephi

Francisco Pérez Hernández, 20076629K
fperezhernandez92@gmail.com

Tabla de contenido

Introducción.....	2
1 Visualizaciones, medidas y gráficos.....	2
1.1 Visualización de la red completa y de la componente gigante.....	2
1.2 Medidas estudiadas.....	3
1.3 Gráficos de las distribuciones de grado, distancia, etc.....	4
2 Análisis de la red (2.1)	4
3 Análisis de la centralidad de los actores (2.2).....	8
4 Estudio de las comunidades extraídas de la red (2.3)	10
5 Referencias.....	12

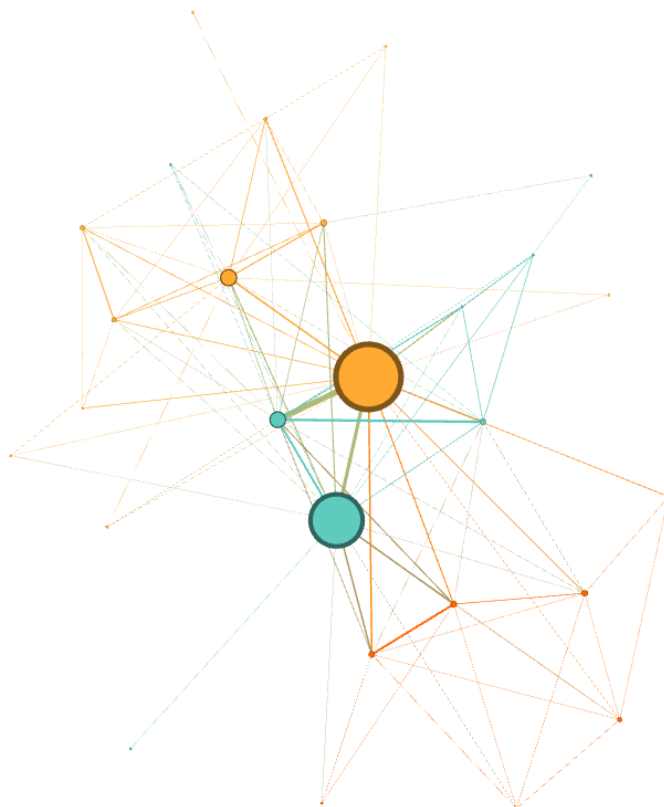
Introducción

En este trabajo vamos a realizar el estudio de una red con la herramienta Gephi. La red que se ha escogido para este estudio es la de la película Piratas del Caribe, ya que he considerado que es mucho mejor a la hora de introducirnos a esta herramienta sabiendo sobre que se está hablando, ya que no soy el creador de ningún grupo en Facebook. La red se ha obtenido de [1].

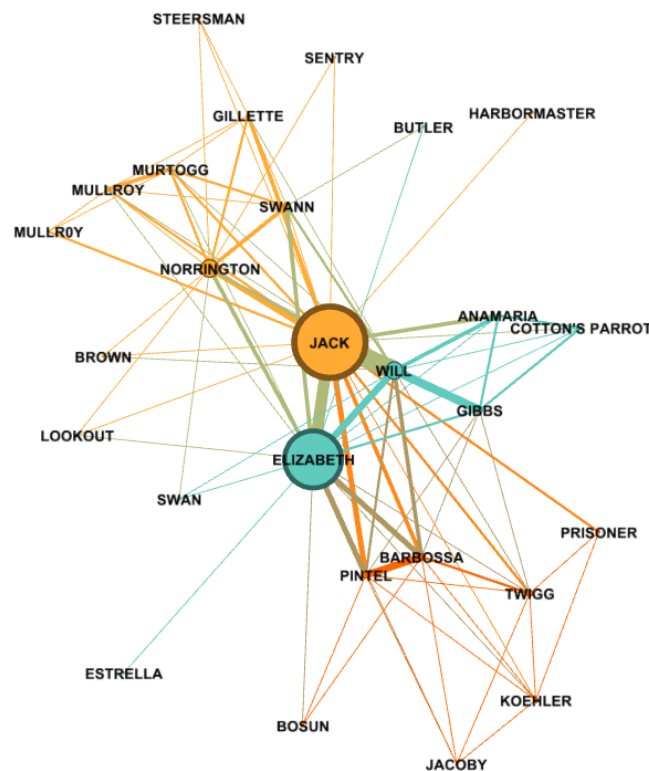
1 Visualizaciones, medidas y gráficos

1.1 Visualización de la red completa y de la componente gigante

La red, una vez que se descarga, tiene este aspecto:



Una vez que la modelamos a nuestro gusto y que añadimos los nombres de cada actor, queda la siguiente red:



En esta red, una vez que vamos a realizar la componente gigante, nos damos cuenta de que el valor de la medida “Componentes conexos” es igual a 1, por lo que la red no va a cambiar si hacemos el cálculo de la componente gigante.

1.2 Medidas estudiadas

Veamos las medidas que se indican en el Excel, a modo de resumen de lo que se verá en los apartados siguientes:

Medida	Valor
Número de nodos N	27
Número de enlaces L	93
Número máximo de enlaces L_{max}	25
Densidad del grafo L/L_{max}	0,265
Grado medio $\langle k \rangle$	7
Diámetro d_{max}	3
Distancia media d	1,818
Coeficiente medio de clustering $\langle C \rangle$	0,802
Número de componentes conexas	1
Número de nodos componente gigante (y %)	No se ha podido
Número de aristas componente gigante (y %)	No se ha podido

Además, veamos la tabla de grado, intermediación, cercanía y vector propio:

Centralidad de Grado	Intermediación	Cercanía	Vector propio
JACK: 21	JACK: 0.3403	JACK: 0.8387	JACK: 1
ELIZABETH: 19	ELIZABETH: 0.2695	ELIZABETH: 0.7878	ELIZABETH: 0.9497
WILL: 15	NORRINGTON: 0.0728	WILL: 0.7027	WILL: 0.8756
NORRINGTON: 13	WILL: 0.0707	NORRINGTON: 0.666	NORRINGTON: 0.7034
GIBBS: 10	SWANN: 0.0211	GIBBS: 0.6190	GIBBS: 0.6564

1.3 Gráficos de las distribuciones de grado, distancia, etc.

A modo de una mejor visualización del documento, las gráficas de este apartado las he adjuntado en las secciones siguientes.

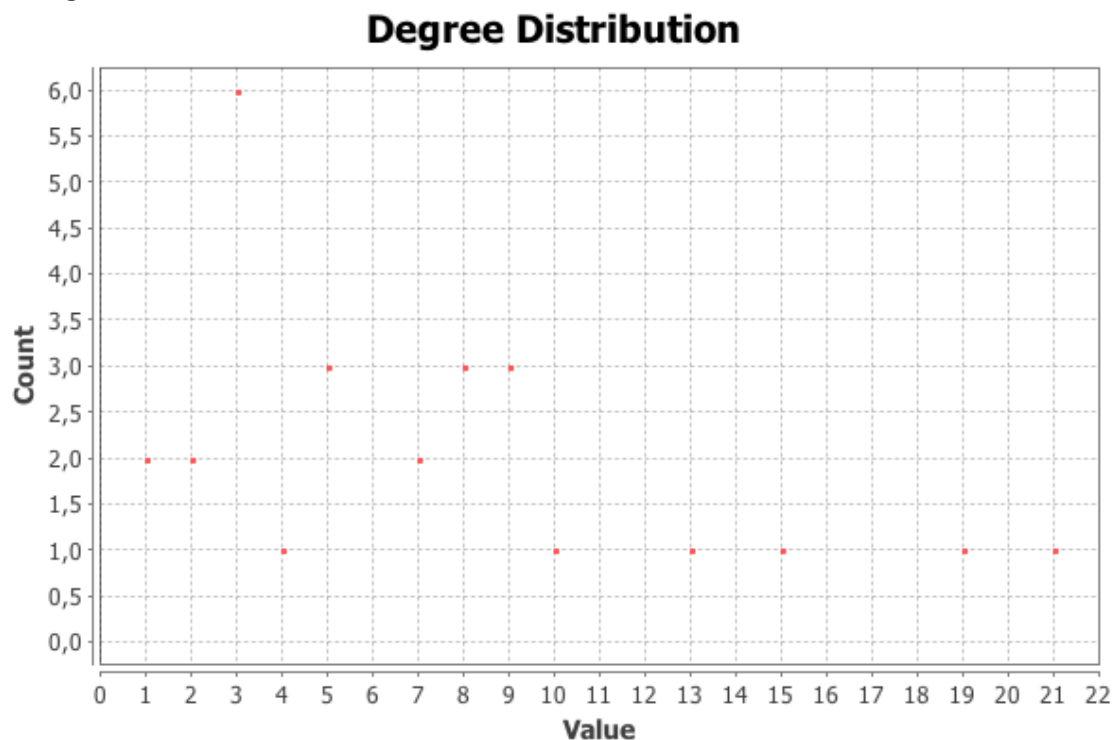
2 Análisis de la red (2.1)

Como hemos visto, tenemos que esta red es una red no dirigida. Por lo tanto, lo primero que realizamos sobre nuestra red es aplicar Force Atlas como algoritmo de layout. La configuración que modificamos será la fuerza de repulsión, usando el valor de 100000 y la gravedad a 10.0.

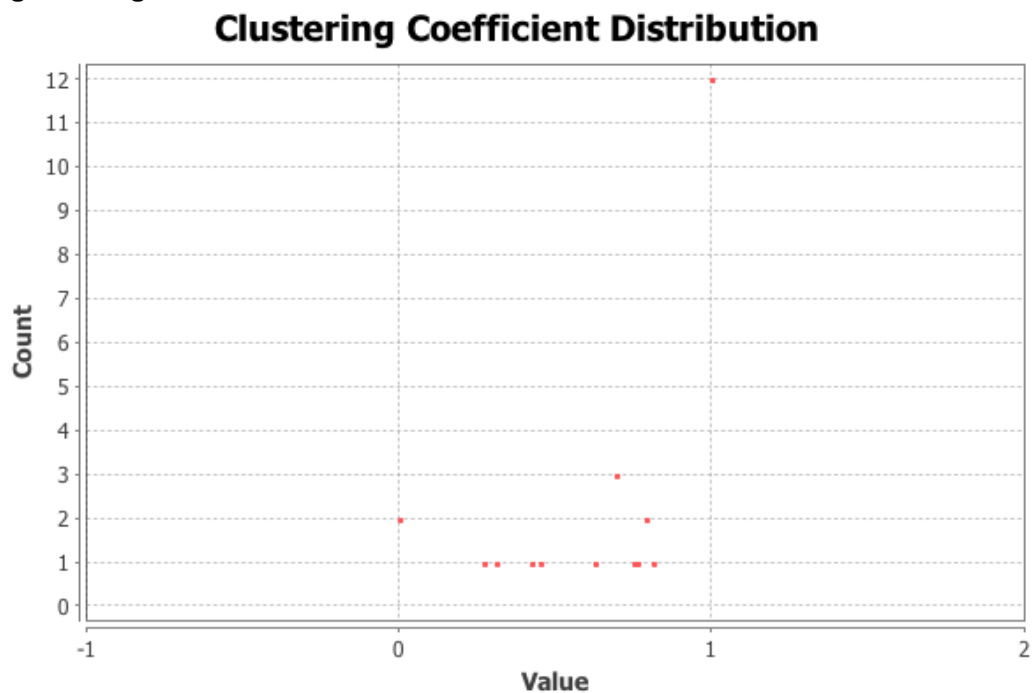
Medidas globales

Como hemos visto en el apartado 1.2, las medidas que realizamos en los primeros pasos del análisis son el número de nodos N igual a 27, el número de enlaces L igual a 93 y el número máximo de enlaces L_{max} igual a 24,645.

Seguidamente, el grado medio $\langle k \rangle$ lo calcularemos con Gephi dando un valor de 6,889 y una gráfica:

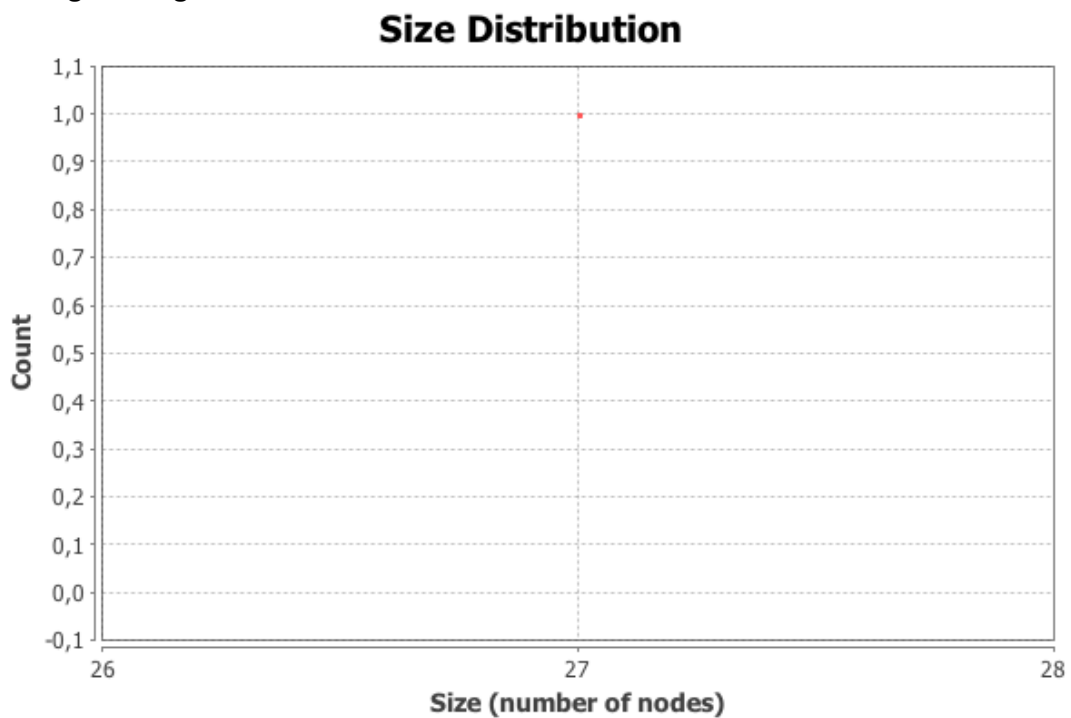


Además, calculamos el valor de la medida de Densidad de grafo, que mide la relación entre número de enlaces L y máximo de enlaces L_{max} , dando como resultado 0,265. El Coeficiente medio de clustering $\langle C \rangle$ nos da el valor de 0,802, además de que hay 149 triángulos. La gráfica de esta medida es:



Conectividad de la red

Veamos primero el número de componentes conexas siendo como resultado 1, junto con la siguiente gráfica:

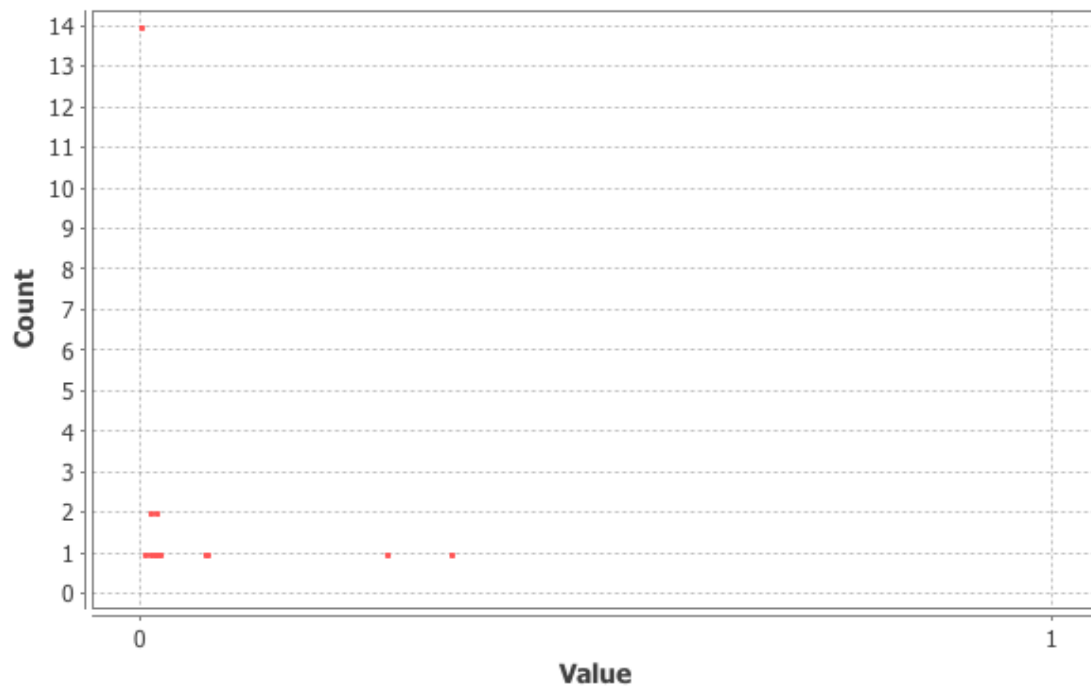


Para la componente gigante, en nuestra red no va a ser necesario ya que como el valor de las componentes conexas ha sido 1 y no vamos a obtener ningún resultado.

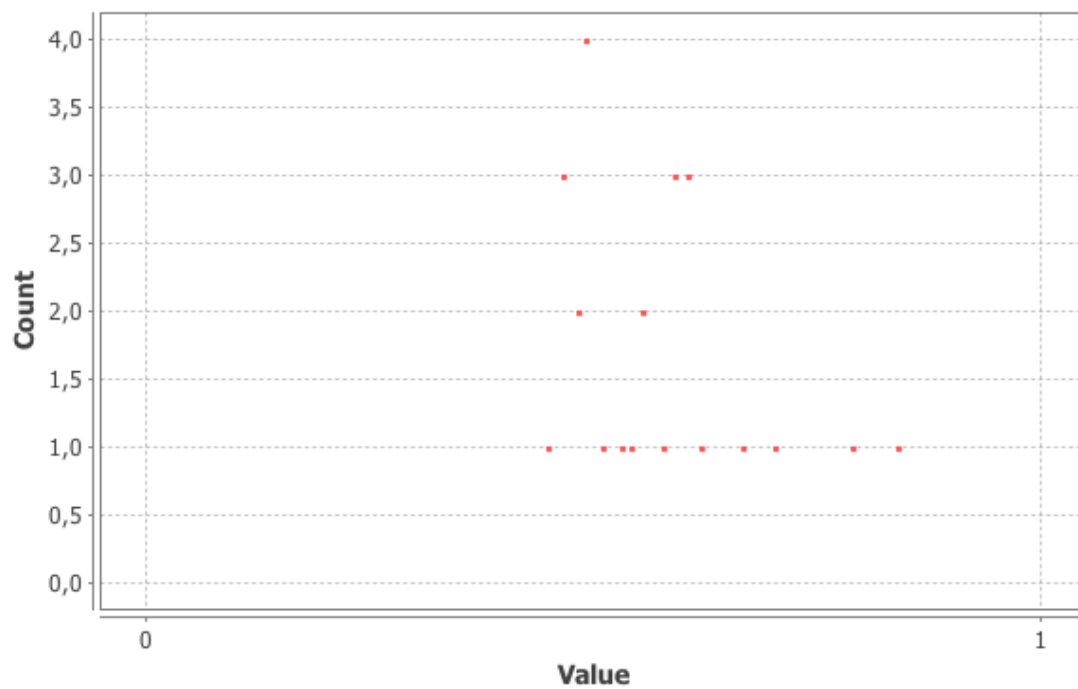
Medidas globales

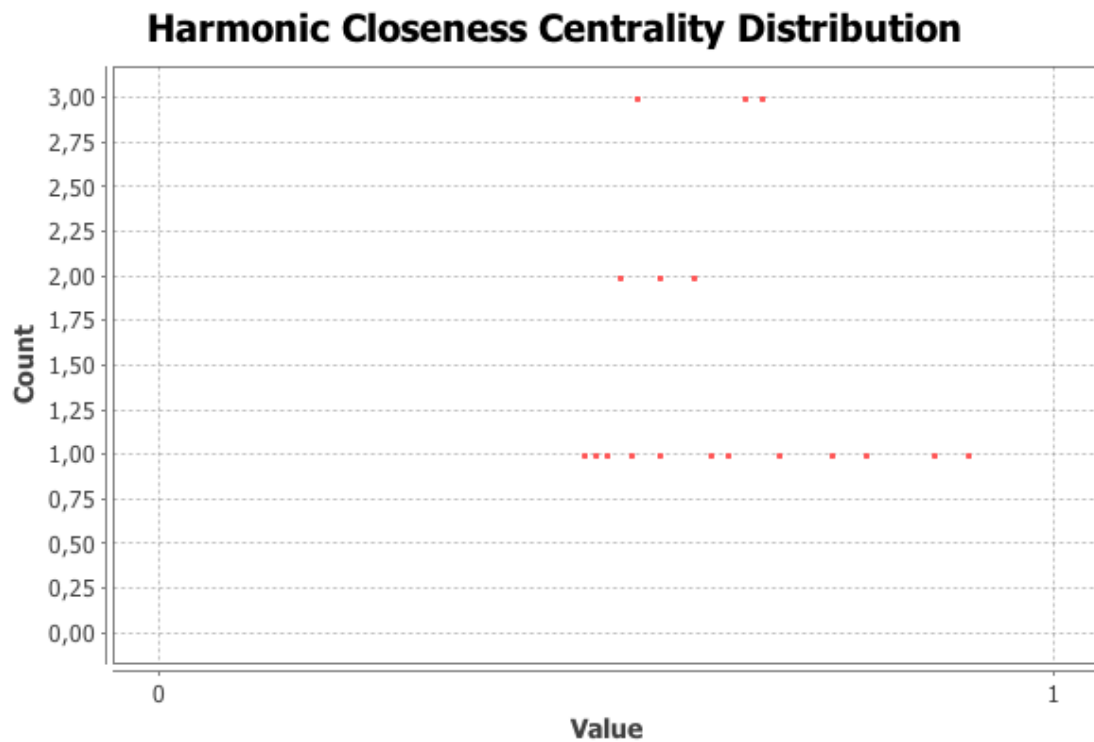
Si calculamos el Diámetro de la red obtenemos que el diámetro d_{max} vale 3, con radio 2 y distancia media d igual a 1,81766. Además, obtenemos estas gráficas:

Intermediación:

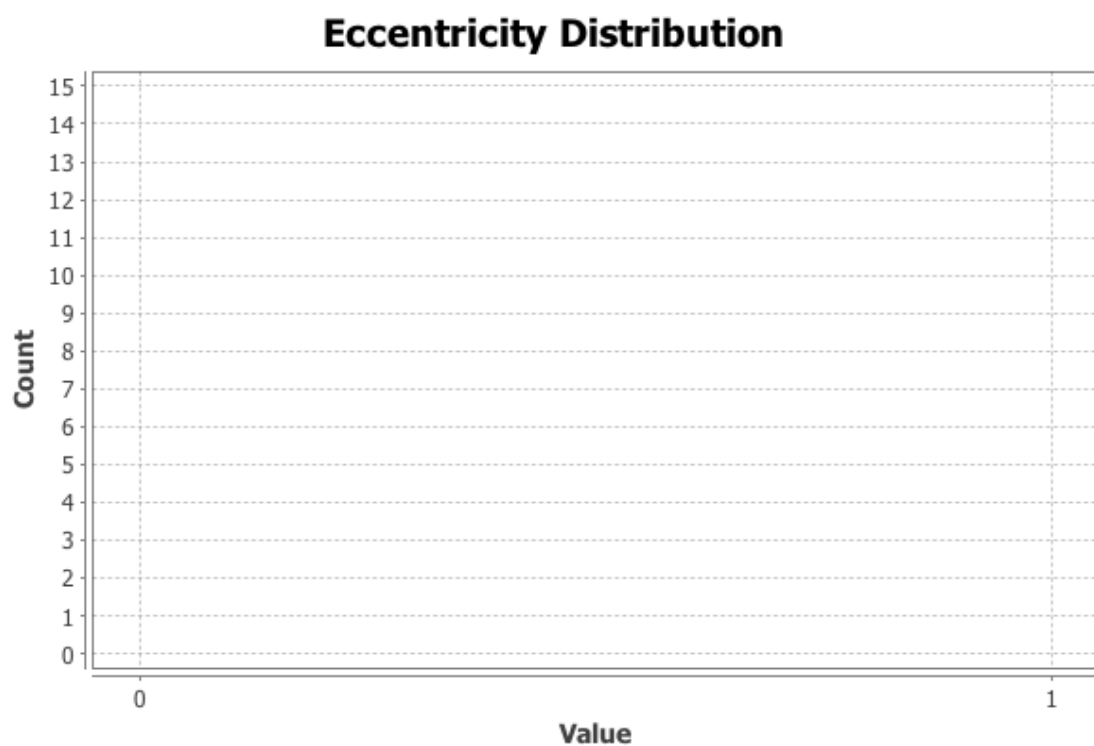
Betweenness Centrality Distribution

Cercanía:

Closeness Centrality Distribution



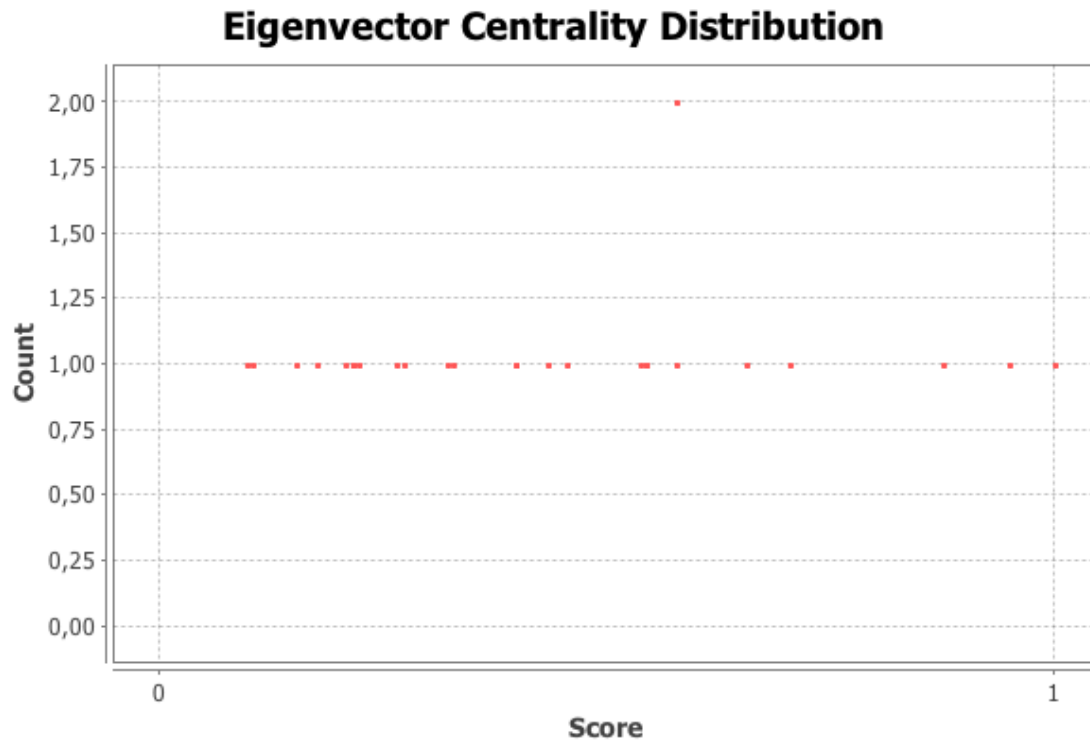
Excentricidad:



Hagamos un resumen de lo que hemos visto en esta sección. Hemos tenido un valor de grado medio de 6,889, por lo que un personaje interactúa con otras 6-7. Además, existen unos pocos nodos fuertemente conectados (hubs). El valor del diámetro máximo ha sido 3, con una media de 1,817. En relación a la conectividad, nuestra red tiene solo una componente conexas. Para el coeficiente de clustering medio hemos tenido 0,802, siendo bastante alto indicando un grado significativo de clustering local.

3 Análisis de la centralidad de los actores (2.2)

Para esta sección, hacemos uso de lo realizado en la sección anterior y además, he realizado la centralidad de vector propio y he obtenido el valor $2.3442E-4$ y la gráfica:



La tabla de las medidas ha sido:

Centralidad de Grado	Intermediación	Cercanía	Vector propio
JACK: 21	JACK: 0.3403	JACK: 0.8387	JACK: 1
ELIZABETH: 19	ELIZABETH: 0.2695	ELIZABETH: 0.7878	ELIZABETH: 0.9497
WILL: 15	NORRINGTON: 0.0728	WILL: 0.7027	WILL: 0.8756
NORRINGTON: 13	WILL: 0.0707	NORRINGTON: 0.666	NORRINGTON: 0.7034
GIBBS: 10	SWANN: 0.0211	GIBBS: 0.6190	GIBBS: 0.6564

De este modo, podemos ver que tenemos 6 actores en total, para estas 4 medidas:

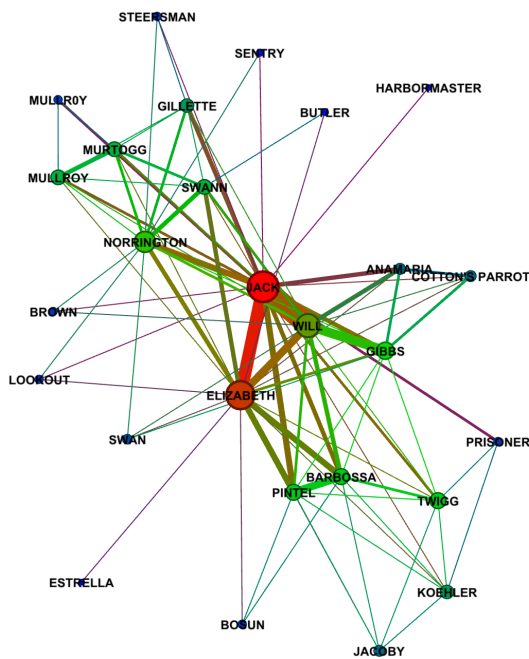
- JACK: es el principal en todas las medidas, ya que es el personaje principal.
- ELIZABETH: siendo otro personaje principal en la película y la segunda con las medidas más importantes.
- WILL: Es normalmente el tercero en las medidas.
- NORRINGTON: Suele ser el cuarto, salvo en la intermediación.
- GIBBS: Otro personaje que sería el quinto más importante.
- SWANN: Solo aparece en la medida intermediación, como el quinto más importante.

Veamos la imagen de cada personaje en [3] para recordar quien era cada uno en la película.

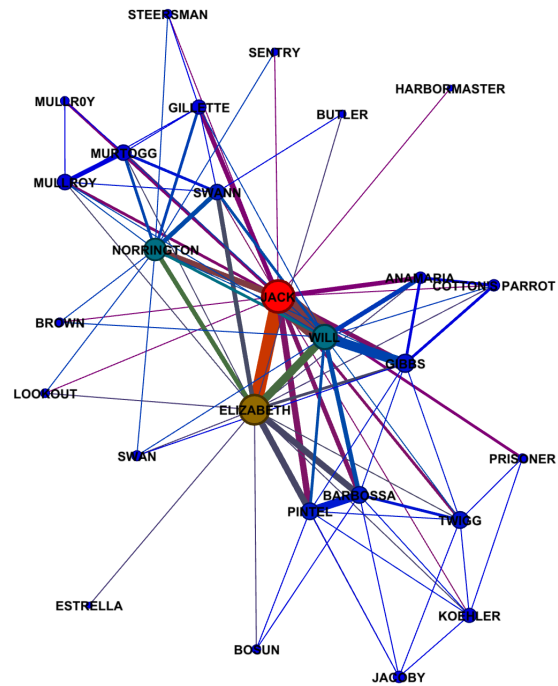
Si mostramos la red, en función de estas medidas y en función del color, siendo:

- Azul: menor valor
- Verde: valor medio
- Rojo: mayor valor

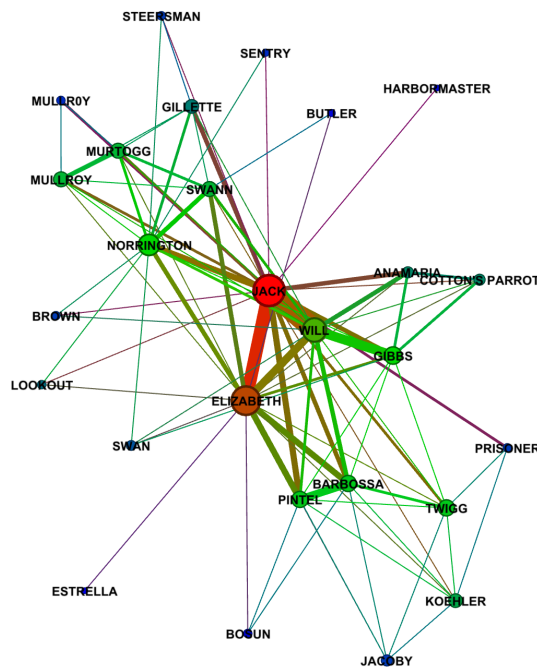
Centralidad de grado



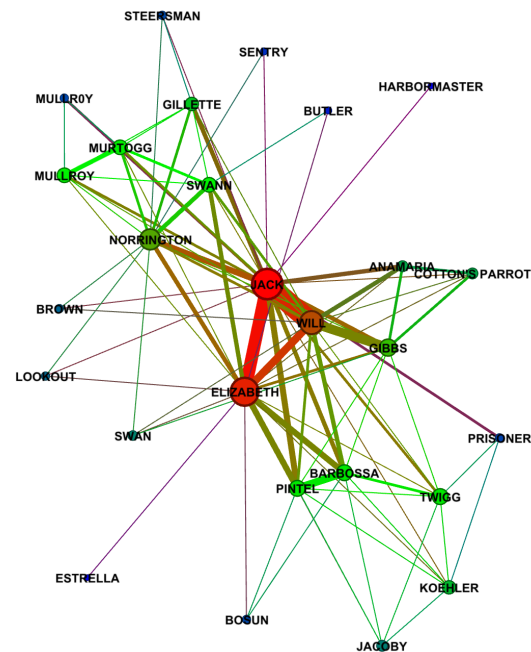
Intermediación



Cercanía



Centralidad de vector propio



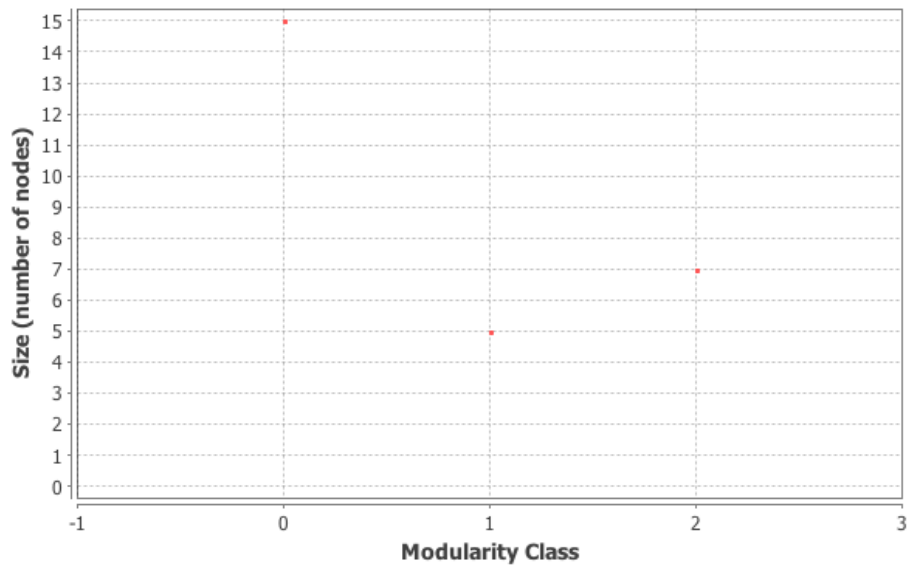
Donde cada medida es:

- **Centralidad de grado** es el número de enlaces que lo conectan con otros.
- **Intermediación**, podemos ver la posición más favorable para conectar a todos los nodos.
- **Cercanía** para medir la centralidad y darle importancia a estar en el centro de las cosas.
- **Centralidad de vector propio**, que se basa en que la centralidad de un nodo concreto depende de cómo de centrales son sus vecinos.

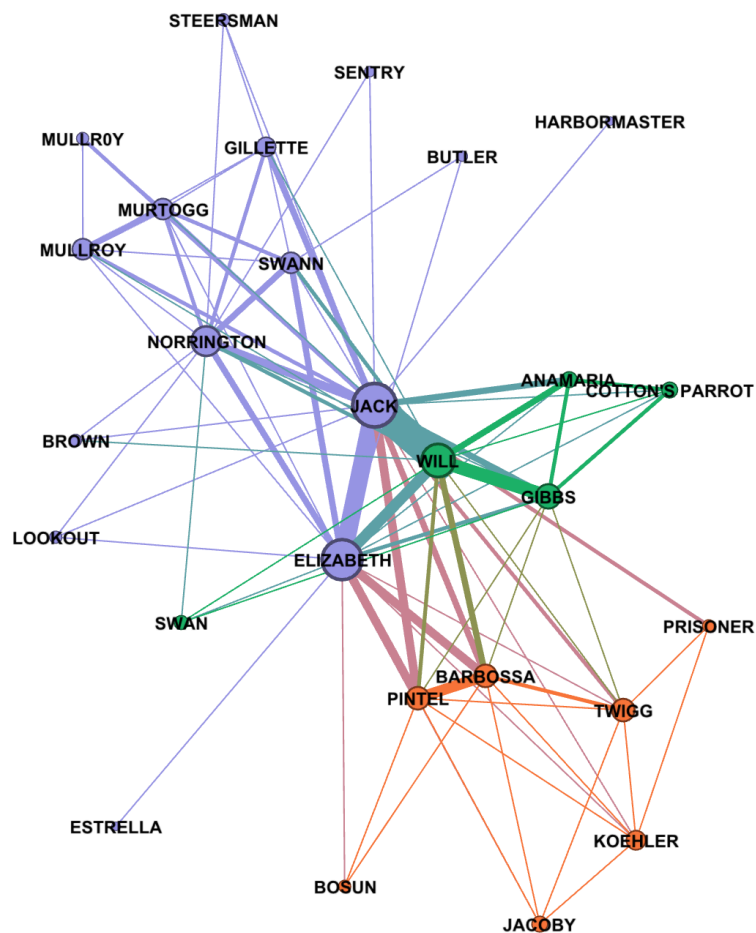
4 Estudio de las comunidades extraídas de la red (2.3)

Con el método de Lovaina, vamos a poder detectar comunidades sobre la red de Piratas del Caribe. Esta medida nos da un valor de 0,174, si usamos el valor de resolución igual a 1.0. Además, nos proporciona 3 comunidades:

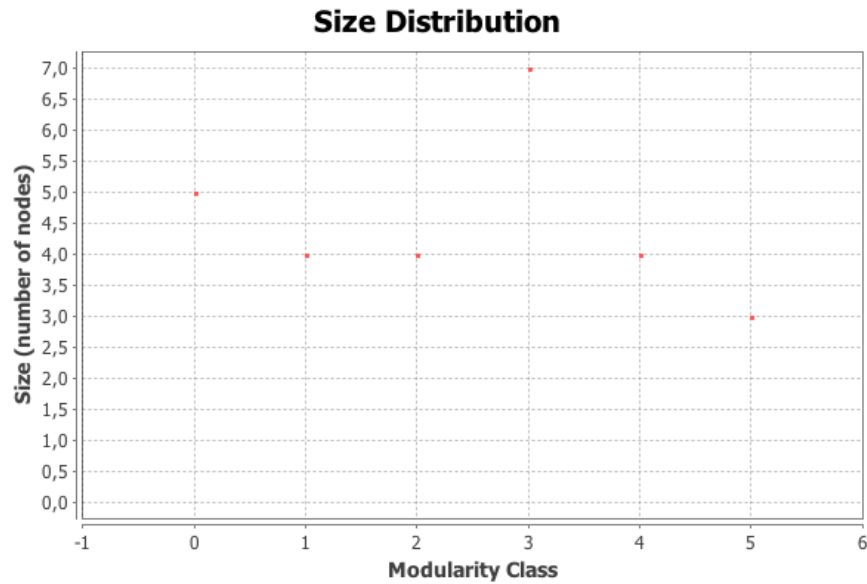
Size Distribution



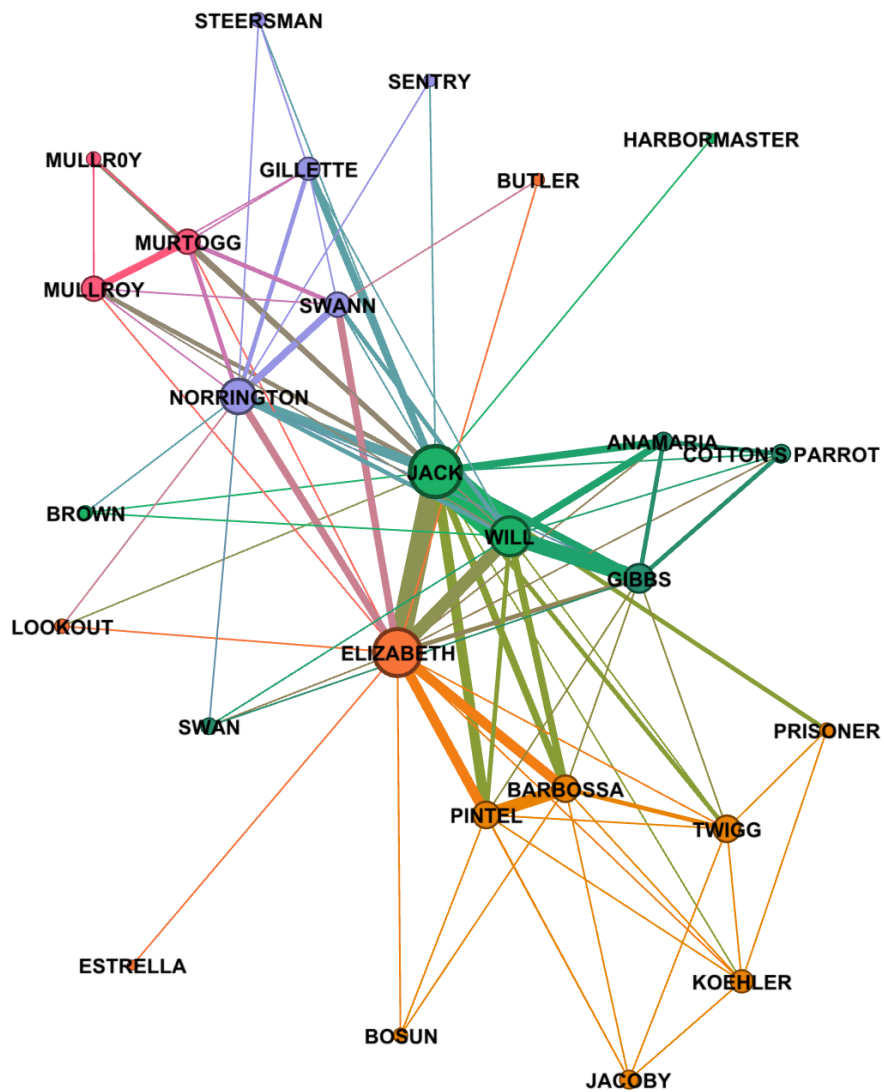
Si visualizamos por comunidades obtenemos:



Si usamos un valor de resolución de 0.5, obtenemos un valor de 0,143 y 6 comunidades:



Y la red quedaría:



Podríamos decir que, en esta película, funcionaría mejor el uso de 3 comunidades, pero para asegurarnos deberíamos ver la película o informarnos más sobre nuestra red para saber que estamos en lo cierto.

5 Referencias

[1] Red Piratas del Caribe: <http://moviegalaxies.com/movies/638-Pirates-of-the-Caribbean:-The-Curse-of-the-Black-Pearl>

[2] Otras redes encontradas: https://figshare.com/articles/Topology_Analysis_of_Social_Networks_Extracted_from_Literature/1373869

[3] Personajes principales según nuestras medidas:



Jack Sparrow



Will Turner



Elizabeth



James
Norrington



Joshamee
Gibbs



Weatherby
Swann