

Práctica 1: Análisis de una red con Gephi

GRUPO 03

EDERSON ALDAIR FUNES CASTILLO

JOSE LUIS GOMEZ ALONSO

ZHIIHAO ZHENG

UNIVERSIDAD DE COMPLUTENSE DE MADRID

Índice

1. Introducción	2
2. Estudios de la red.....	4
a) Medidas globales de la red	4
Distribución de grados de la red completa	4
Distribución del coeficiente de clustering.....	5
b) Medidas de centralidad	6
PageRank	6
c) Conectividad	6
Modularidad	7
d) Visualización.....	8
e) Comentarios.....	11
3. Bibliografía.....	12

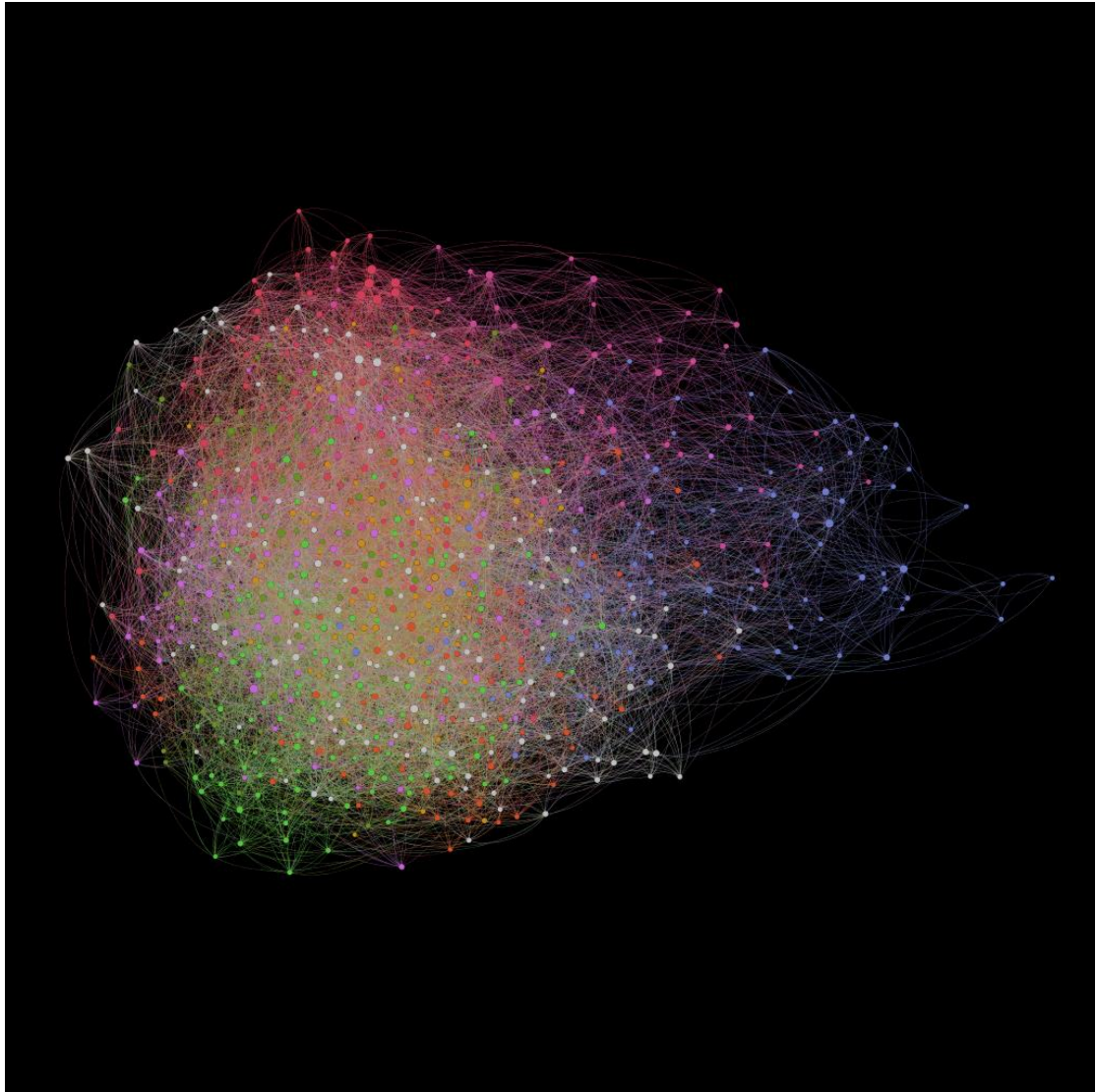
1.Introducción

La práctica consiste en construir una red a partir un archivo csv que contiene los datos de las películas y los actores que participaron.

Para poder extraer los datos útiles para construir la red, hemos programado una pequeña aplicación java, para procesar los datos, y pasarlos al formato que pueda leer Gephi.

Los pasos de la aplicación son los siguientes:

- 1) Abrir los datos de entrada.
- 2) Leer una línea de datos y pasar las películas que ha participado a un array de String.
- 3) Cogemos la primera película del array.
- 4) Si esta película no ha aparecido antes, lo guardamos en un árbol junto con su id (un contador) y lo guardamos también al archivo de salida node.csv
- 5) Cogemos la segunda película del array y si no ha aparecido antes, también guardamos en el árbol.
- 6) Miramos si existe una arista entre dos películas (Buscar en otro árbol que guarda las aristas).
 - Si no existe: insertamos en el árbol de aristas con su peso a 1.
 - Si ya existe: cogemos el peso de la arista y sumamos 1.
- 7) Así unimos todas las películas del array:
 $0 \rightarrow 1, 0 \rightarrow 2, \dots, 0 \rightarrow (N-1), 1 \rightarrow 2, 1 \rightarrow 3, \dots, 1 \rightarrow (N-1), \dots, (N-2) \rightarrow (N-1)$
- 8) Volvemos al paso 2) hasta que termina todos.
- 9) Recorrer el árbol de aristas y guardamos en el archivo de salida edge.csv.



Para las visualizaciones significativas hemos aplicado varios filtros como componente gigante, modularidad y rango de grado y peso de aristas. También layout de ForceAtlas 2 y expansión.

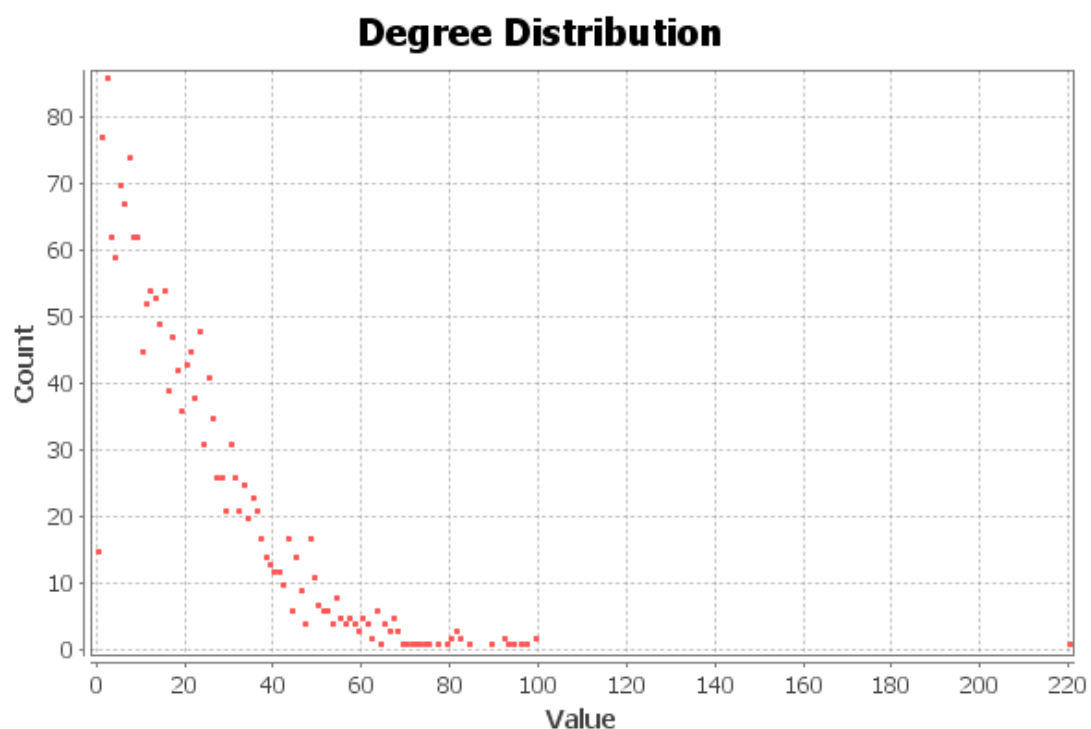
Y para sacar una mejor información desde la red, hemos repartido la red por comunidades utilizando el cálculo del panel de estadística y el tamaño del nodo según el grado con el peso.

2. Estudios de la redes

a) Medidas globales de la red

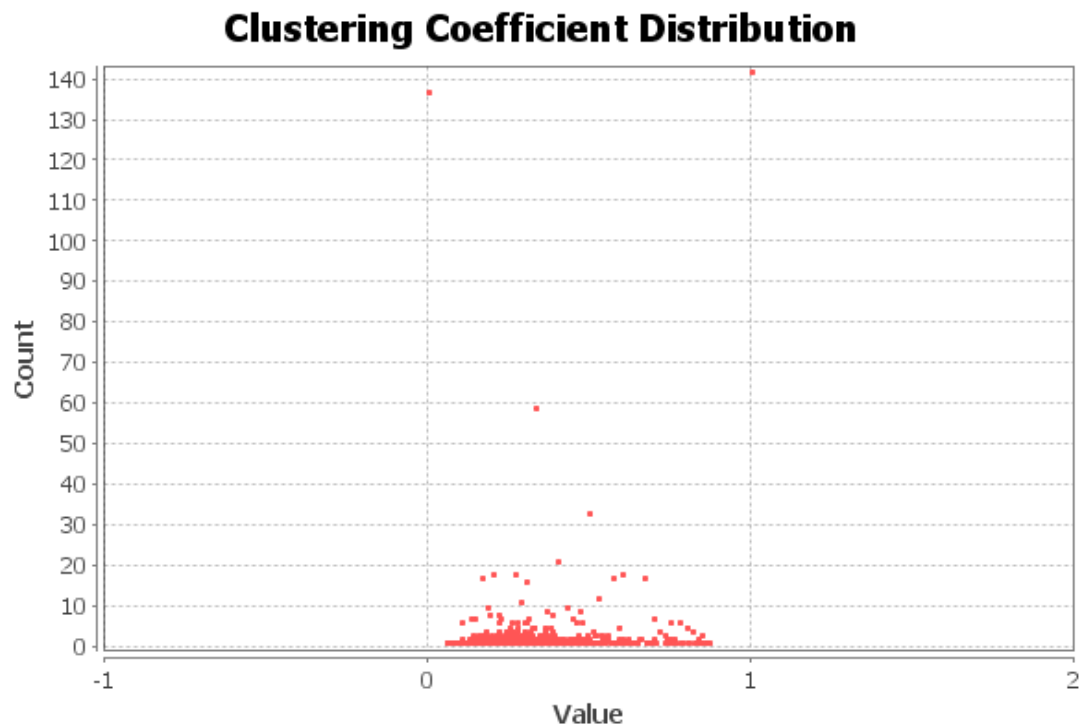
Nodos	1894
Número de enlaces	18793
Densidad de la grafo	0,01
Grado medio	19.845
Grado medio con el peso	21.408
Diámetro de la red	9
Distancia media	3.132
Modularidad	0.383
Componentes conexas	27
Coficiente medio de clustering	0.361

Distribución de grados de la red completa



Podemos observar que hay una gran cantidad de películas con grados relativamente pequeños en las que pocos actores han coincidido en participación con otras, pero a medida que este aumenta, la cantidad de películas lo hace de manera inversamente proporcional quedando las películas restantes por norma general como parte de una misma relación fílmica. Hamlet es la película con mayor participación de actores coincidentes del resto de películas.

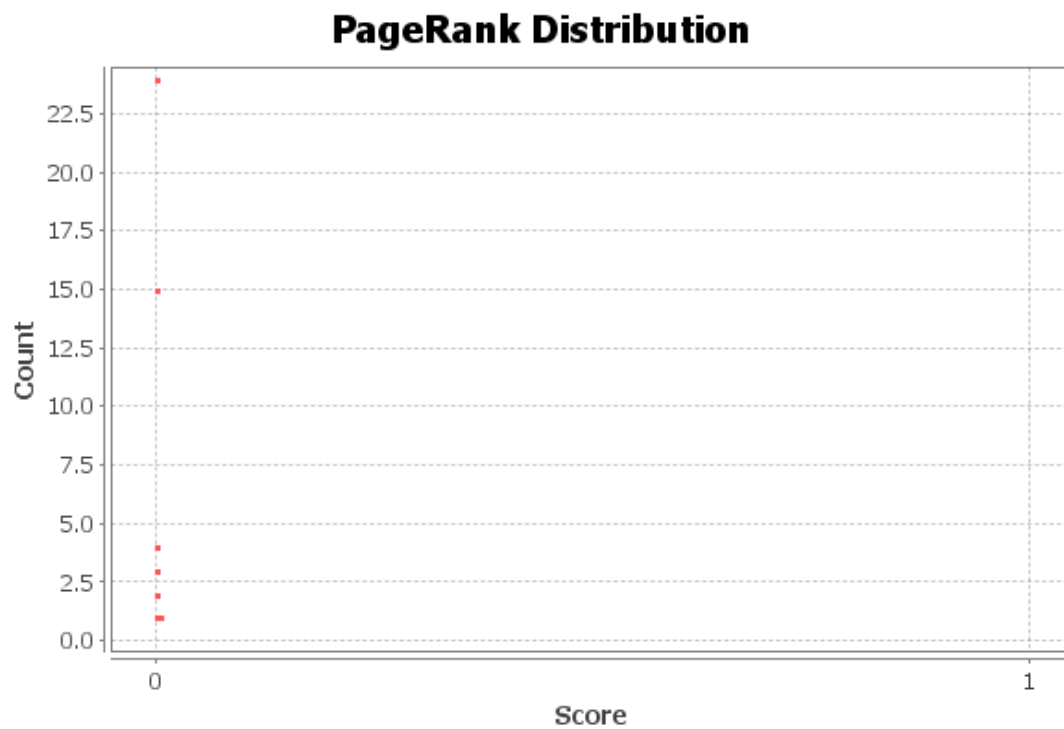
Distribución del coeficiente de clustering



El coeficiente de agrupación es de 0.359. Este alcanza por poco 1/3 de la agrupación del grafo. Esto quiere decir que de media, por cada película conectada a otras 3, solo 2 de estas estarán conectados entre sí (solo un lado del "triangle"). Esto es normal tratándose de un grafo en la que la mayor parte de las películas no tienen una relación directa entre sí.

b) Medidas de centralidad

PageRank



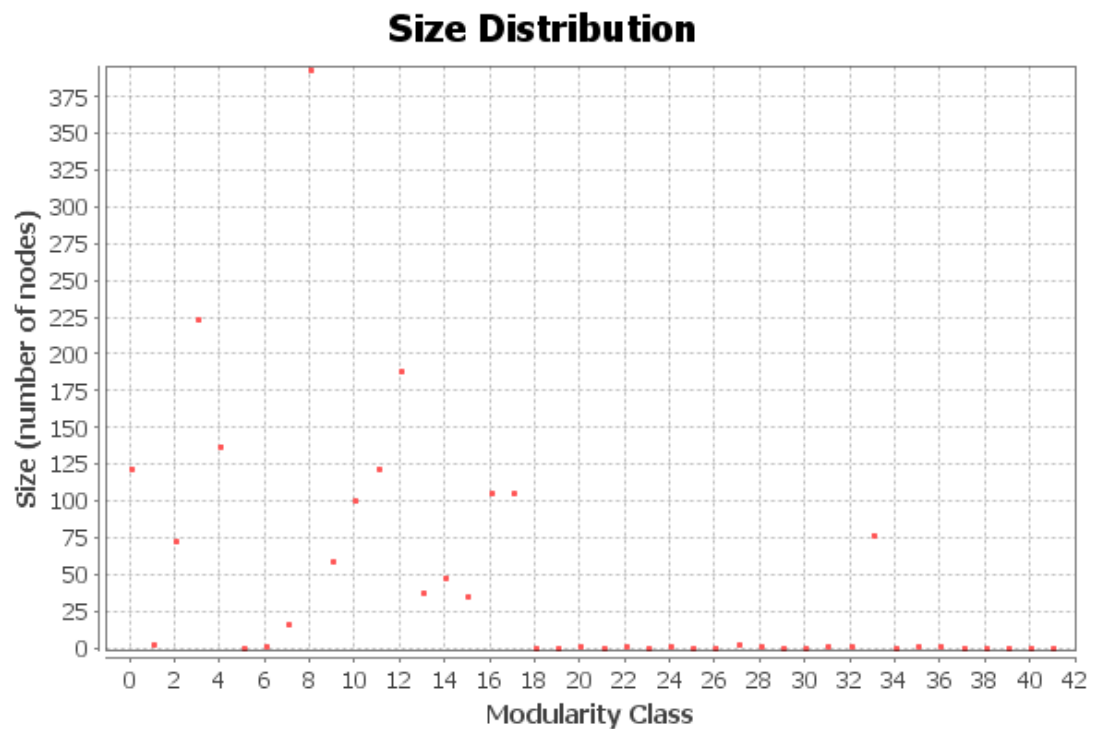
c) Conectividad

Al aplicar el filtro de componente gigante, aparecen 1855 nodos (97.94% visibles) y 18778 aristas (99.92% visibles).

Comunidades dentro de la componentes gigante: 16

Comunidades totales del grafo: 42

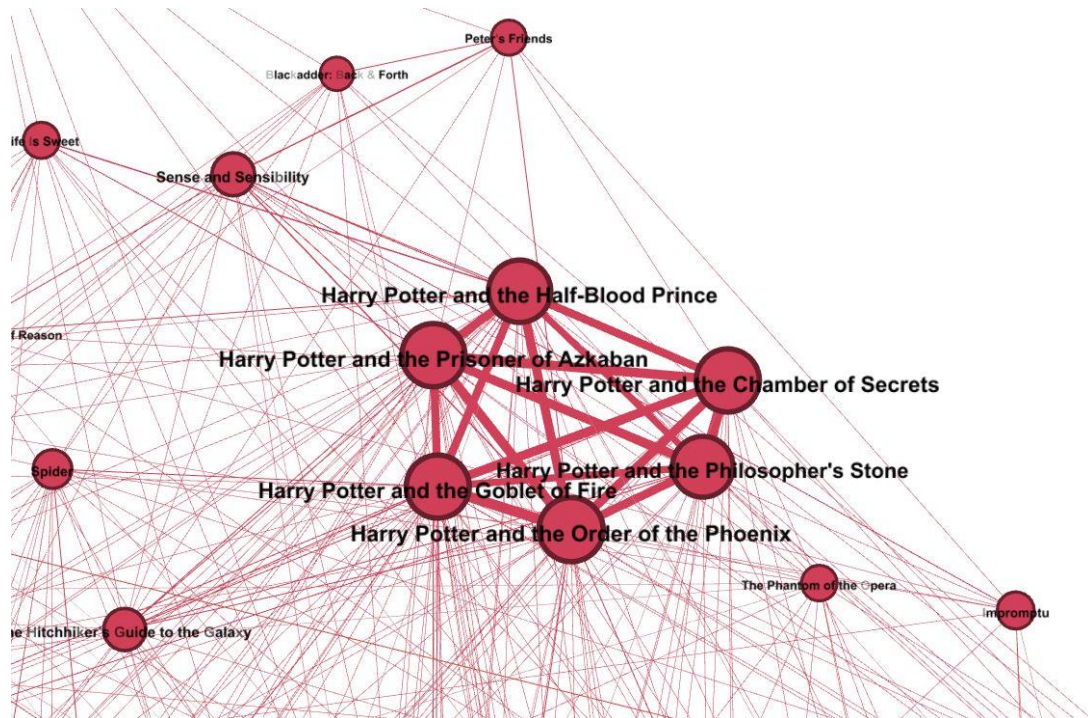
Modularidad



Este grafo tiene un muy bajo nivel de modularidad debido a que son pocas las películas capaces de compartir una gran cantidad de mismos actores en las mismas. Solo son capaces de ganar una alta modularidad, una vez más, las películas pertenecientes a la misma saga cinematográfica por norma.

d) Visualización

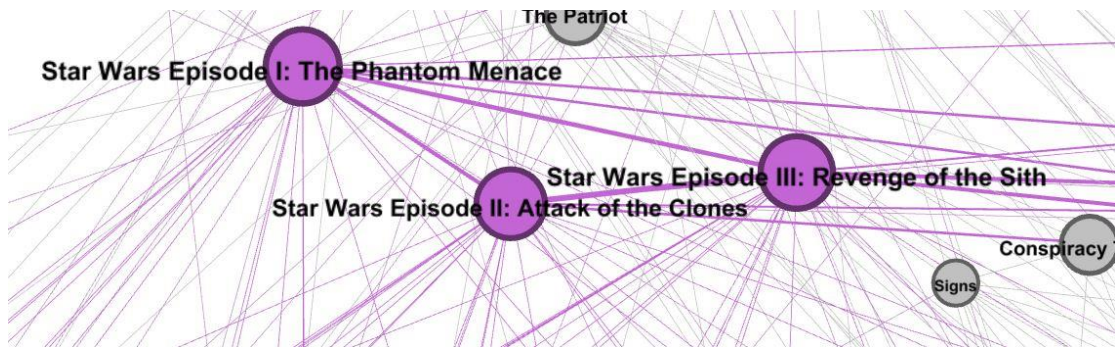
Películas de la saga: Harry Potter



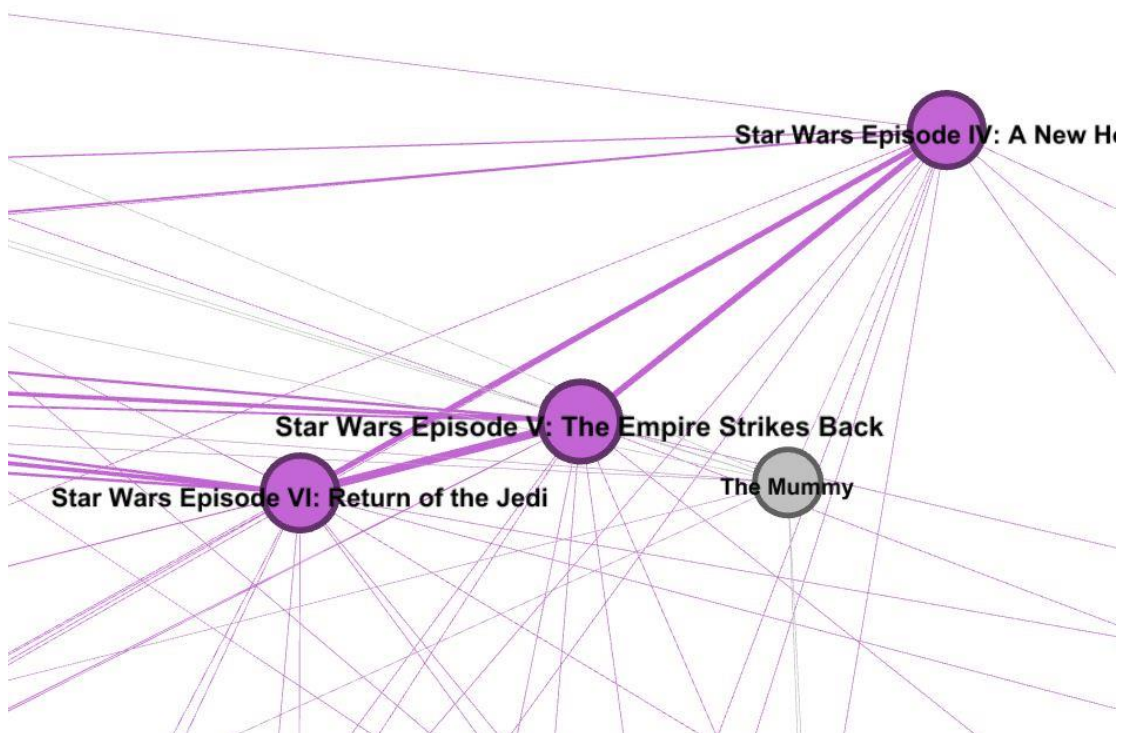
Trilogía: Señor de anillos



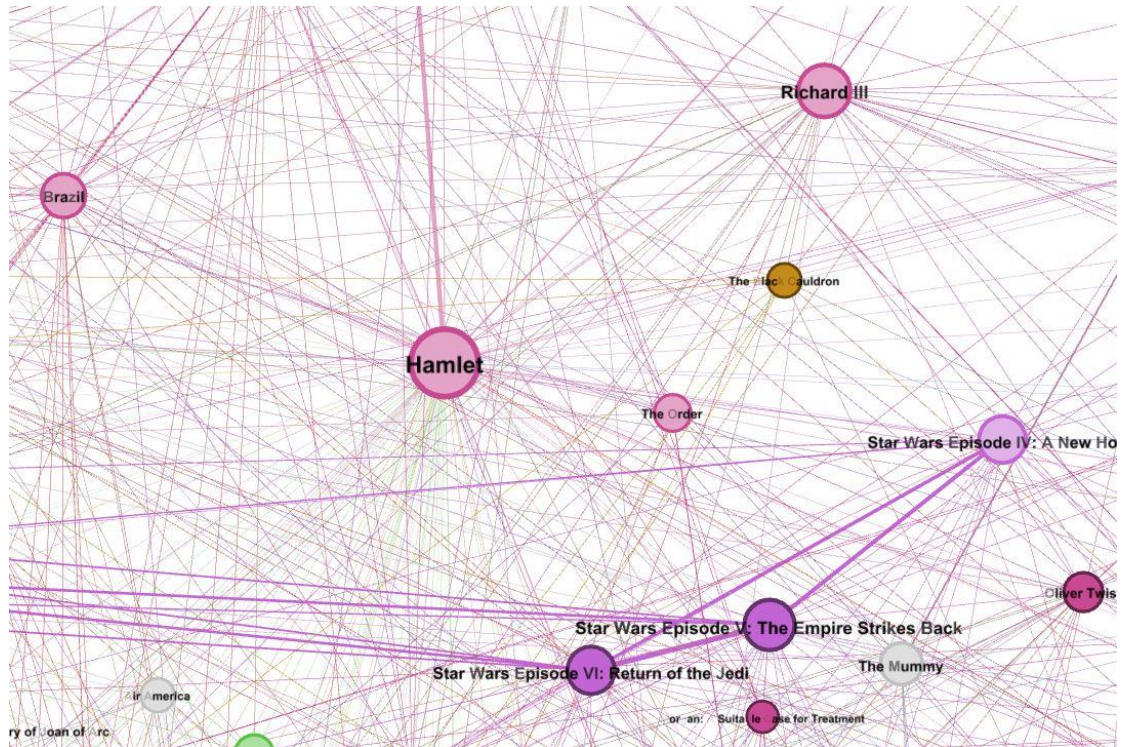
Trilogía: Star Wars (Primera parte)



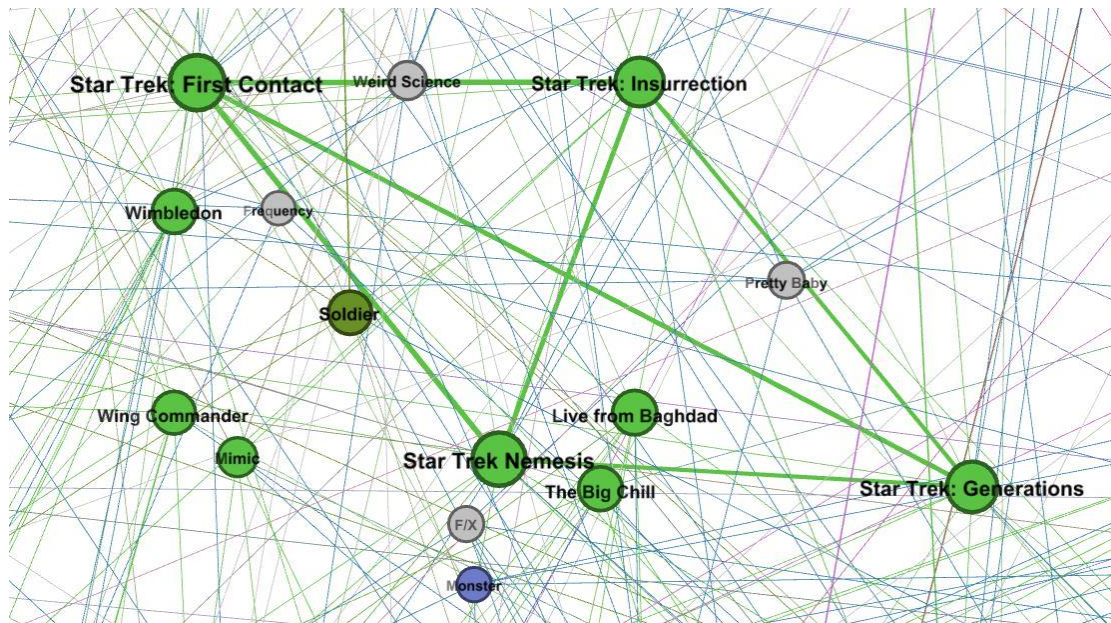
Trilogía: Star Wars (Segunda parte)



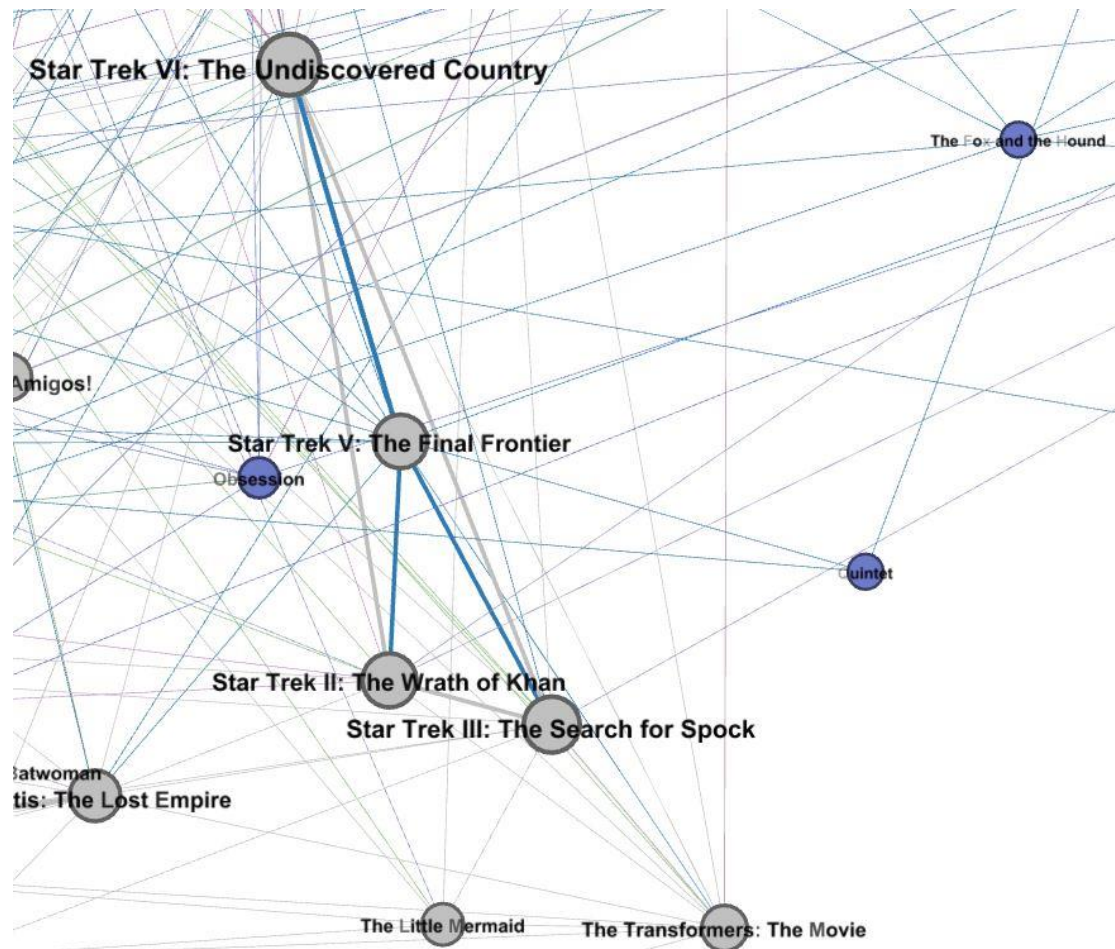
Nodo con el mayor grado: Hamlet



Películas de la saga: Star Trek



Star Trek: segunda parte



e) Comentarios

Con la red que hemos analizado en el apartado anterior, podemos saber que la red está dividida en distintas comunidades, según las participaciones de los actores entre las películas. Cada nodo representa una película, y cada arista representa algún actor que participa en las dos películas, cuando había más actores que participan en dos películas, más peso tiene esa arista (grosor).

La red de las películas de saga como Harry Potter, están unidas con aristas muy gruesas, esto quiere decir que la mayoría de actores que participaron dentro de la saga son los mismos. Sin embargo, en la saga de Star Wars, existe una red separada, esto muestra que los actores que participaron en la primera parte de la película no son los mismos que los de la segunda. También pasó esto en las películas de Star Trek.

Esto no siempre se cumple, ya que sagas cinematográficas como la franquicia Alien (que al tratarse de un género de terror) no comparten la mayor parte del plantel de actores y por lo tanto, aun tratándose de la misma saga fílmica, estas están unidas con finas aristas.

Otra red es la película Hamlet, esta película tiene el mayor grado de toda la red, esto muestra que es la película que tiene más actores que han participado en otras películas. Esto puede deberse a la gran cantidad de reparto en ella, incluyendo a actores muy famosos y activos del mundo del cine de la talla de [Derek Jacobi](#) como [el rey Claudio](#) , [Julie Christie](#) como [la reina Gertrudis](#) , [Kate Winslet](#) como [Ofelia](#) , [Michael Maloney](#) como [Laertes](#) , [Richard Briers](#) como [Polonio](#) , y [Nicholas Farrell](#) como [Horacio](#) . Además de [Robin Williams](#) , [Gérard Depardieu](#) , [Jack Lemmon](#) , [Billy Crystal](#) , [Rufus Sewell](#) , [Charlton Heston](#) , [Richard Attenborough](#) , [Judi Dench](#) , [John Gielgud](#) y [Ken Dodd](#) .

3. Bibliografía

- Slides gephi: http://barabasilab.neu.edu/courses/phys5116/content/Gephi_2016_EKT.pdf
- Apuntes de campus