

Práctica 2 Grafo Marvel

Ignacio Arias Barra

February 25, 2017

1. Carga el grafo `hero-social-network.gephi` en Gephi.

Se trata de un grafo no dirigido, en el que cada nodo corresponde a un superhéroe y cada arista a una colaboración entre ambos. Hay un total de 10469 nodos y 178115 aristas.

2. Modifica el grafo para que solo sean visibles aquellos superhéroes que hayan colaborado en, al menos, 1000 ocasiones con otros superhéroes.

Aplicamos un filtro de rango que tenga en cuenta el grado de cada nodo. Establecemos un valor mínimo para el rango del grado de 1000. De esta manera, obtenemos el siguiente gráfico:

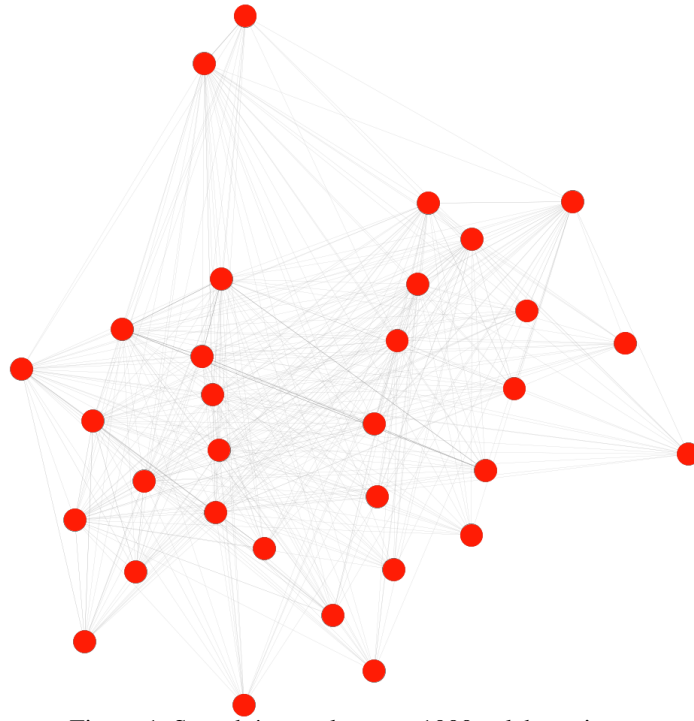


Figura 1. Superhéroes, al menos 1000 colaboraciones

Se trata de un grafo de 32 nodos, en el que se ha aplicado un color naranja para cada nodo, así como un color negro para cada arista, de tal manera que se mejore la visualización del mismo.

Se ha aplicado un tamaño igual para cada nodo, de tal forma que hayamos obtenido un grafo como respuesta en el que sólo nos tenemos que centrar en el número de nodos obtenidos.

Con el objetivo de centrarnos en los nodos, se ha aplicado un reescalado de los pesos de las aristas en la parte de visualización de Gephi, haciendo que todas valgan lo mismo. De esta forma, la visualización del tamaño cada arista es la misma.

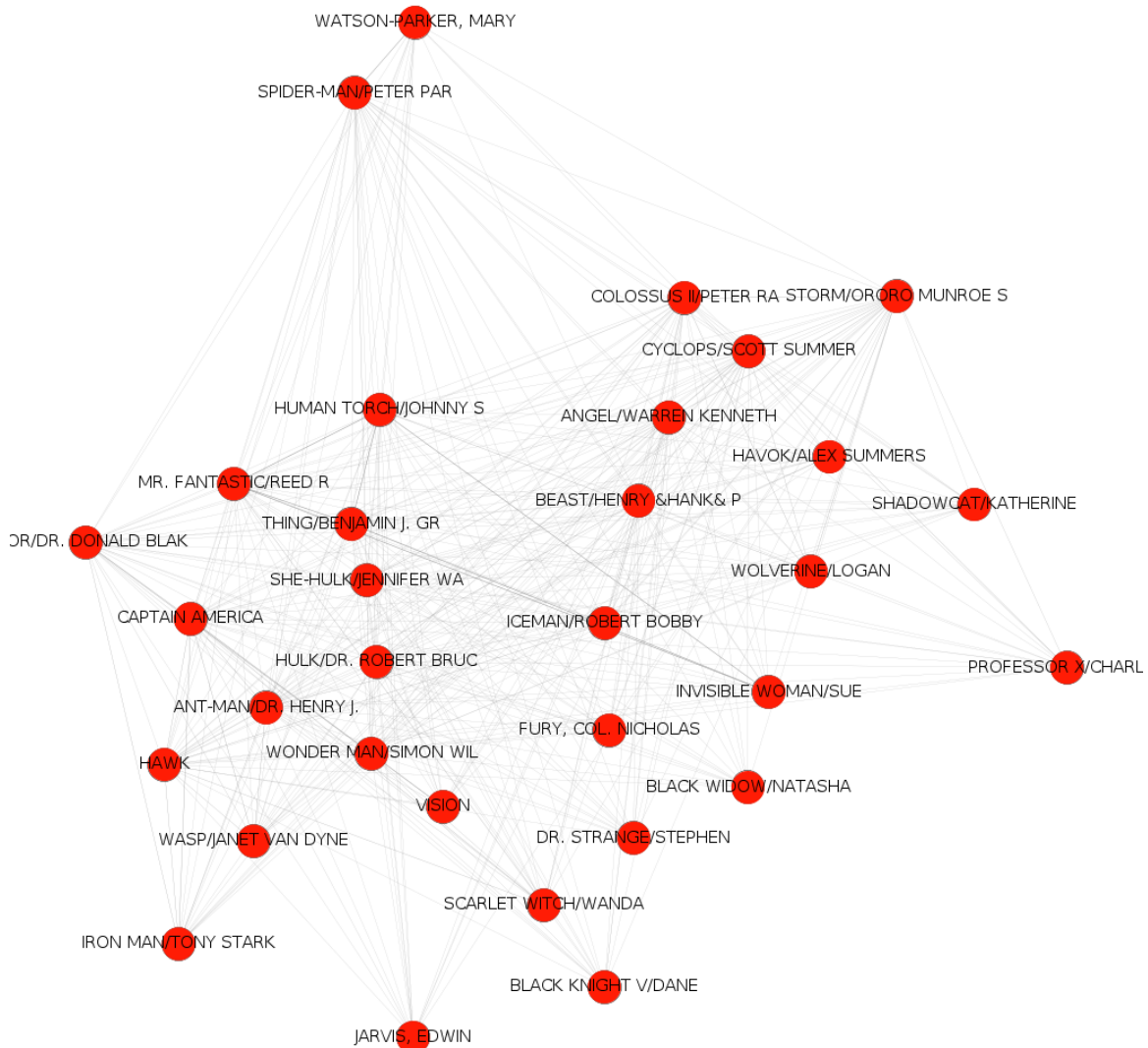


Figura 2. Superhéroes, al menos 1000 colaboraciones con nombres.

Este grafo representa lo mismo que el anterior, pero habiendo activado las etiquetas en la visualización.

3. Modifica el grafo para que el tamaño de los vértices sea proporcional al número de colaboraciones realizadas.

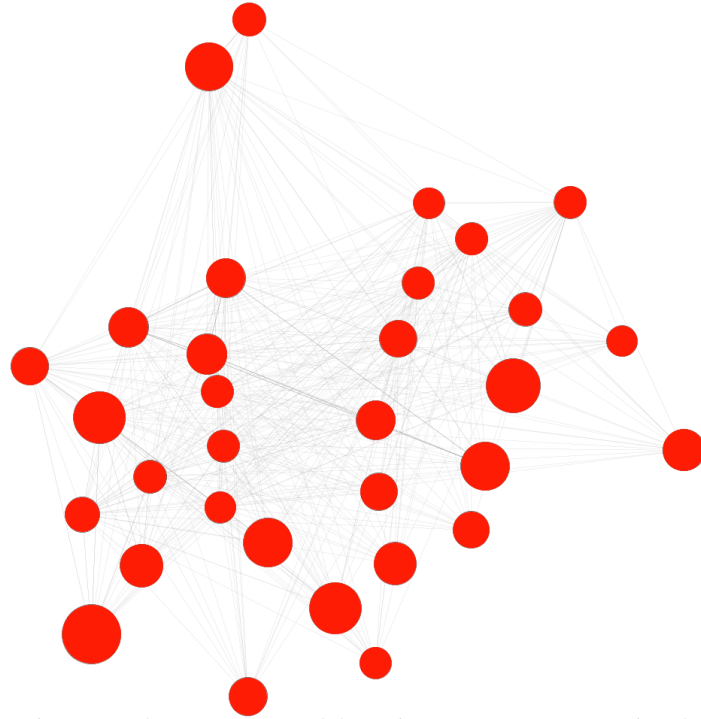


Figura 3. Al menos 1000 colaboraciones, tamaño proporcional.

En la figura 3 podemos observar como hay unos nodos que son más grandes que otros, debido al número total de colaboraciones realizadas.

Se puede observar un grupo de 4 o 5 nodos que tienen un tamaño mayor que el resto.

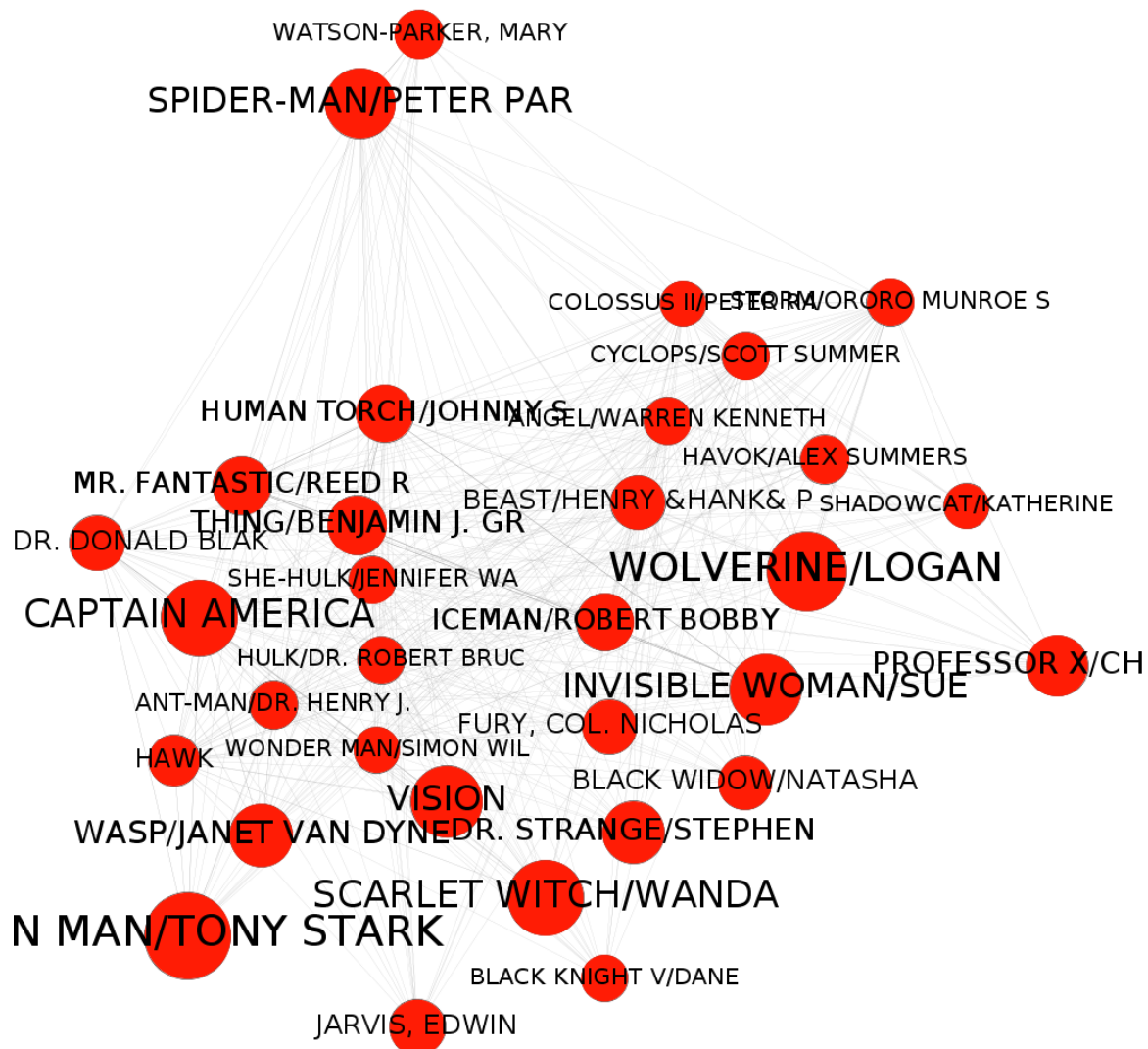


Figura 4. Al menos 1000 colaboraciones, tamaño proporcional (nombres).

En este nuevo grafo, se representan de nuevo los superhéroes con nodos proporcionales a las colaboraciones. El grupo de 4-5 nodos que parece predominar en tamaño, es el formado por Iron Man, Scarlet Witch, Wolverine, Captain America y Spider-Man. Esto quiere decir que son los que más colaboraciones con otros superhéroes han realizado.

4. ¿Quién es el superhéroe que ha hecho más colaboraciones? ¿Y el que menos?

IRON MAN / TONY STAR

Figura 5. Más colaboraciones.

Si elevamos el filtro de rango, respecto al grado, hasta el máximo, nos da como resultado que IRON MAN / TONY STAR es el superhéroe que más colaboraciones ha realizado, con un total de 2189.

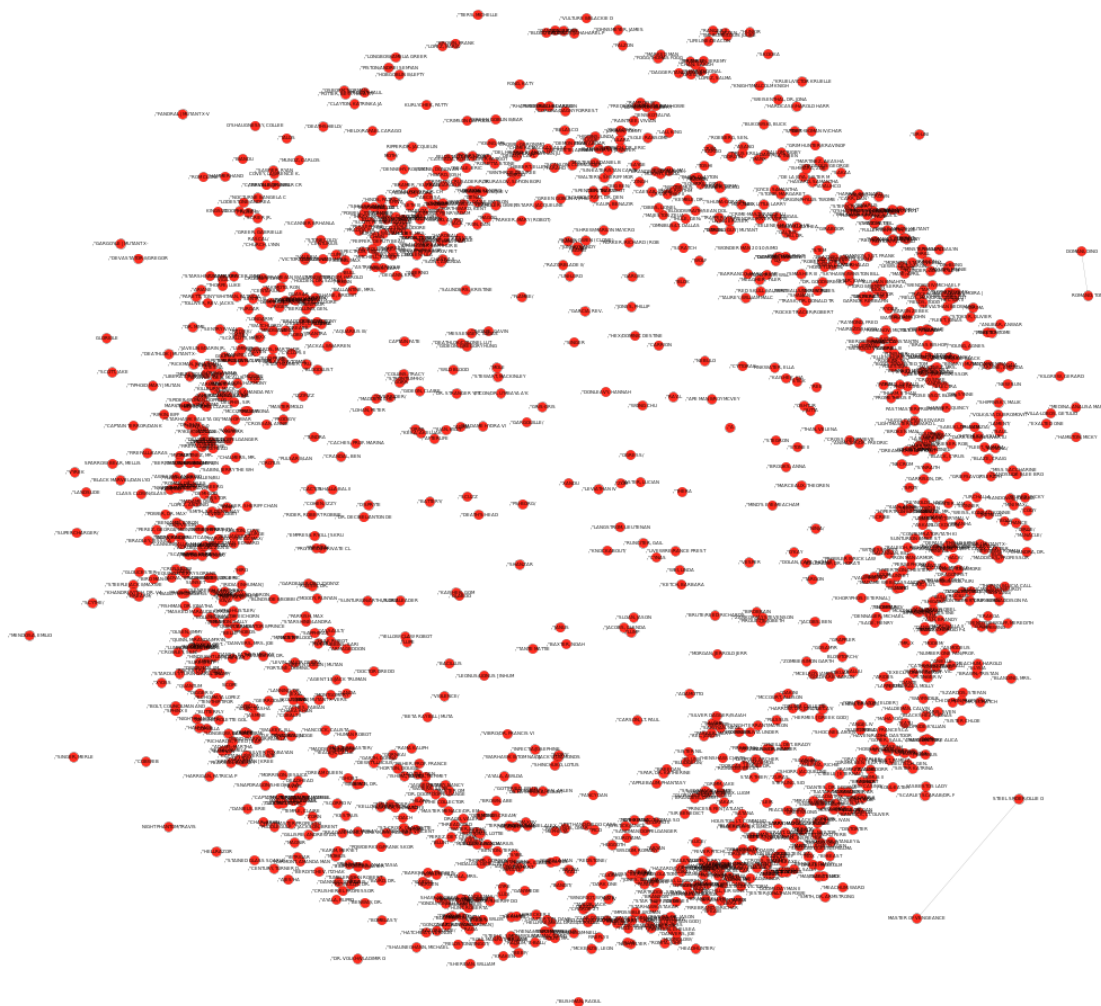


Figura 6. Superhéroes con menos colaboraciones.

Por otro lado, si reducimos el filtro de rango, respecto al grado de los nodos, hasta el mínimo, obtenemos el anterior grafo. En el podemos observar 1414 nodos que no realizan ninguna aportación.

Si nos fijamos, en la parte derecha superior e inferior, podemos ver 4 nodos que están conectados entre sí por pares. Esto quiere decir que hay 2 parejas de superhéroes que sólo han realizado una colaboración.

Estos pares de superhéroes son:

- Par 1: Master of Vengeance – Steel Spider / Ollie o
- Par 2: Domani, Dino – Romano, Tony

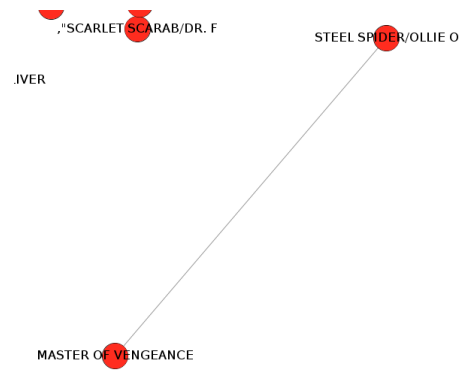


Figura 8. Pareja 1 con menos colaboraciones.

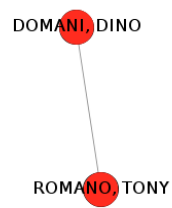


Figura 9. Pareja 2 con menos colaboraciones.

SHADOWCAT/KATHERINE

Figura 7. Superhéroe con menos colaboraciones.

Si tenemos en cuenta sólo el grafo filtrado, obtenemos que el que realiza menos colaboraciones es Shadowcat. Para ello, se ha deslizado el filtro de rango respecto al grado de un nodo a partir de 1000 hacia arriba hasta que ha aparecido un nodo. En concreto, el valor obtenido de rango ha sido 1003, convirtiéndole en el superhéroe con menos colaboraciones respecto a los que más colaboraciones realizan.

5. ¿Qué nodo o nodos podemos considerar importantes? Indica por qué son importantes y cómo lo has calculado.

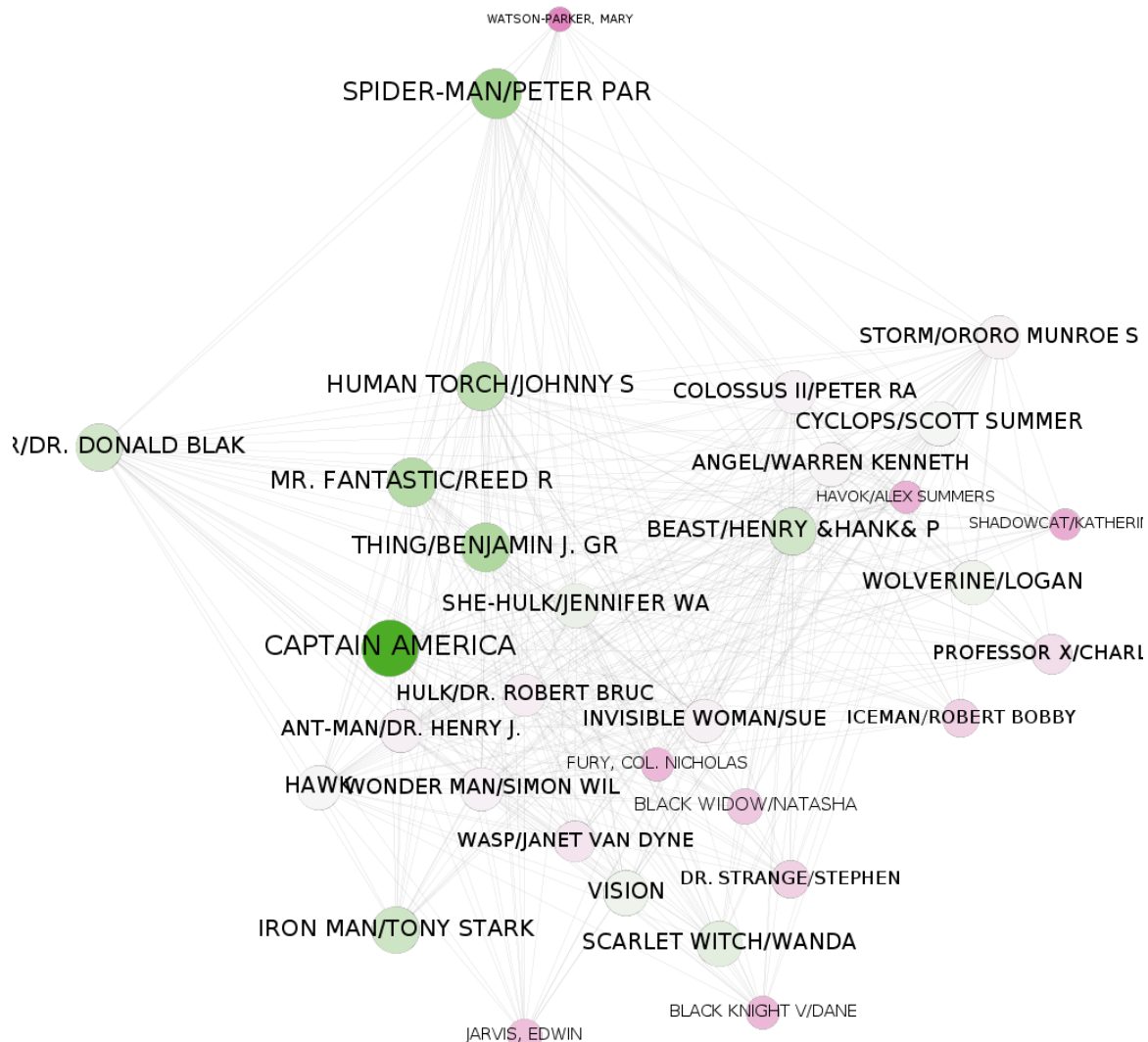


Figura 10. Superhéroes más importantes

Definimos que un nodo es importante cuando los nodos con aristas incidentes en el mismo son importantes por sí mismos. En el caso de un grado no dirigido, este cálculo se centrará en ver si los nodos a los que esté conectado, son importantes en cuanto al número de aristas totales. En este contexto, un superhéroe es más importante si le ayudan/ayuda a otros importantes.

Para el cálculo de dicho algoritmo, usamos la opción Eigenvector Centrality.

Para la visualización del grafo, se ha tenido en cuenta el color y el tamaño en función de la importancia del nodo. A mayor tamaño, mayor importancia. Respecto a los colores, el verde se asignará a los nodos con mayor importancia, gris para los nodos medios y rosa para los nodos menos importantes.

Observando el grafo, podemos ver como Captain America es el superhéroe más importante. Después, podemos observar un grupo de nodos importantes. A saber Thor, IronMan, Beast, Human torch, Spider man, Mr.Fantastic y Thing.

6. Calcula la densidad del grafo e interpreta su valor.

Definimos la densidad de un grafo como un valor entre 0 y 1 que nos indica cuánto está un grafo de ser un

grafo completo, es decir, cuánto está de cerca de tener todas las aristas posibles.

En este caso, se han realizado dos análisis. En el primer análisis, se han tenido en cuenta todos los nodos del grafo. El valor obtenido es de 0.003. Si calculamos cuántas aristas habría en caso de ser un grafo completo, tendríamos 54794746 aristas en total. Como observamos, teniendo en cuenta que en este grafo tenemos 178155 aristas, vemos que está muy lejos del número máximo potencial de aristas.

El resultado obtenido indica que tenemos un grafo en el que, debido a la gran cantidad de héroes, no hay muchos que realicen muchas colaboraciones con otros héroes.

El segundo análisis se ha realizado sobre el subgrafo resultante de aplicar el filtro de al menos 1000 colaboraciones. El valor obtenido es de 0.754. Si calculamos cuántas aristas habría en caso de ser un grafo completo, tendríamos $\text{rformat}(\text{pot.ari.1000})$ arista en total. Como observamos, teniendo en cuenta que en este grafo tenemos 374 aristas, vemos que bastante cerca del número máximo potencial de aristas.

El resultado obtenido en este subgrafo indica que tenemos un grafo en el que los héroes con mayor número de colaboraciones, también realizan muchas colaboraciones entre ellos.

En ambos casos, la densidad ha sido calculada teniendo en cuenta que el grafo es no dirigido.

7. Aplica un método de clustering y representa los nodos de cada cluster con un color diferente. ¿Qué sentido encuentras a los personajes que forman parte de cada cluster?

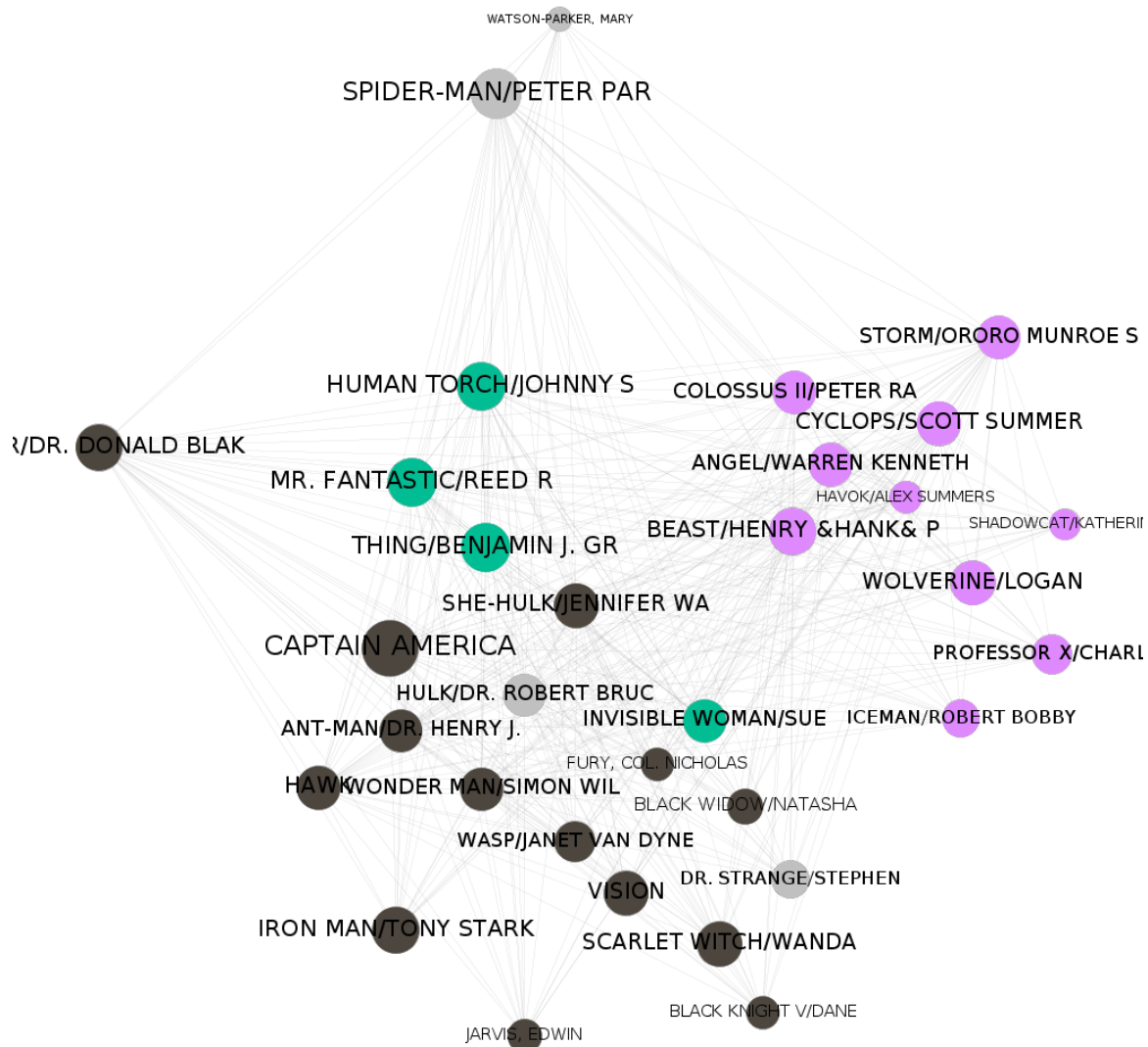


Figura 11. Superhéroes, al menos 1000 colaboraciones con nombres.

En este apartado se tendrá en cuenta sólo el grafo de superhéroes con al menos 1000 colaboraciones.

Para aplicar el método clustering, se ha aplicado el cálculo de la modularidad del grafo. A partir de aquí, se ha aplicado un color para cada cluster resultante.

Se obtiene un grafo con 4 clusters. Si observamos los componentes de cada cluster, podemos ver que lo forman superhéroes que forman parte de los mismos equipos en la ciencia ficción. En realidad, si tenemos en cuenta esta relación, deberíamos juntar los negros con los grises en el mismo cluster.

Los grupos son:

- Verdes: 4 fantásticos
- Morados: Xmen
- Negros y grises: Los vengadores

De acuerdo con esto, es lógico que los clusters formados tengan estos componentes, ya que son los superhéroes que más van a colaborar entre sí.

8. Modifica el layout para que separe gráficamente los clusters. ¿Qué layout deberías utilizar?

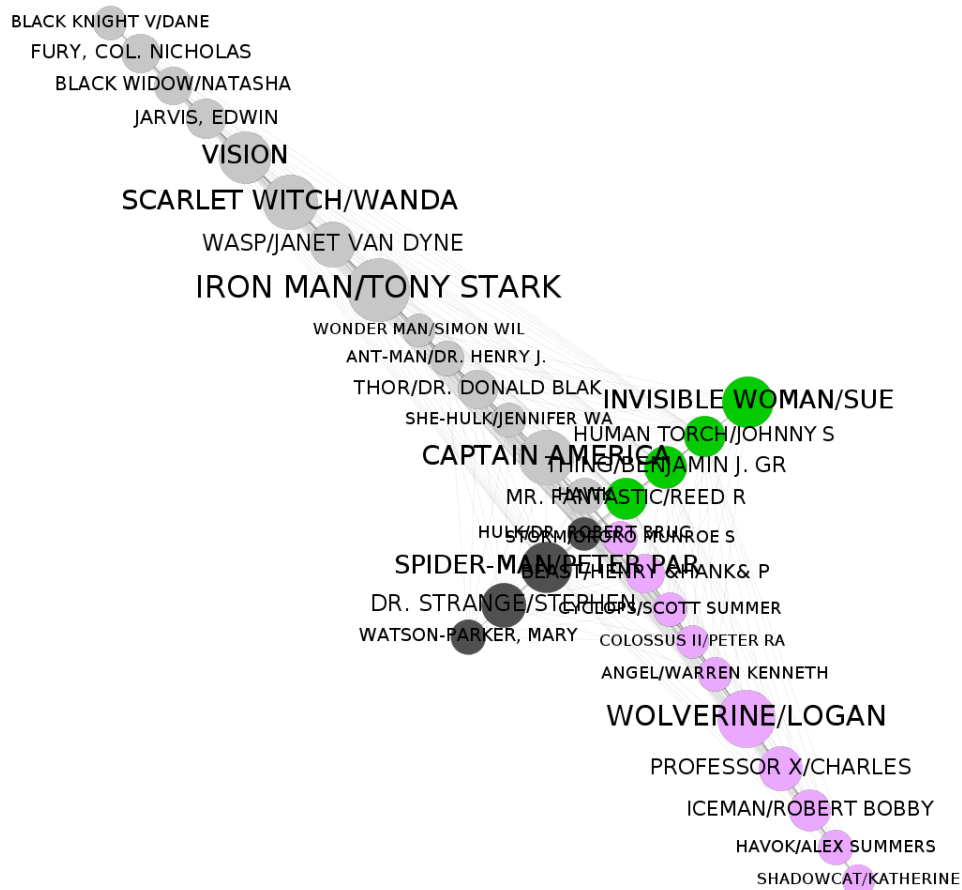


Figura 12. Clustering superhéroes.

Para modificar el gráfico de tal manera que se separen los clusters gráficamente, necesitamos un layout que permita agrupar por la modularity class previamente calculada.

En este caso, se ha utilizado el layout Radial Axis Layout. Como se puede observar, en cada radio del gráfico obtenido tenemos cada cluster previamente generado.

9. Modifica la visualización para obtener el subgrafo de cada cluster por separado.

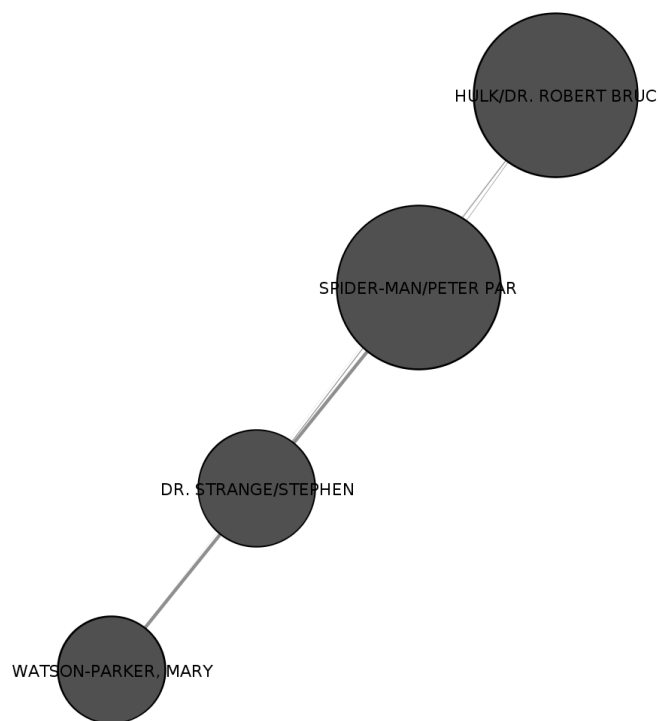


Figura 13. Cluster perteneciente a Los Vengadores.

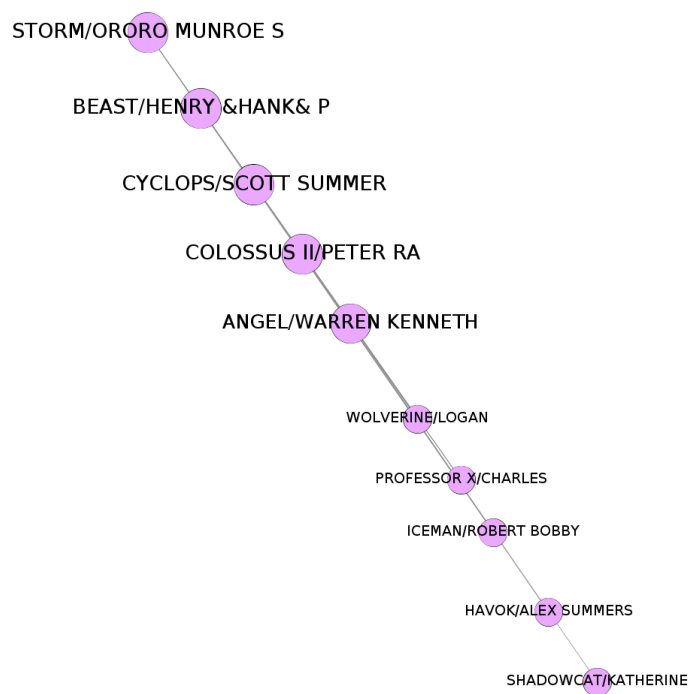
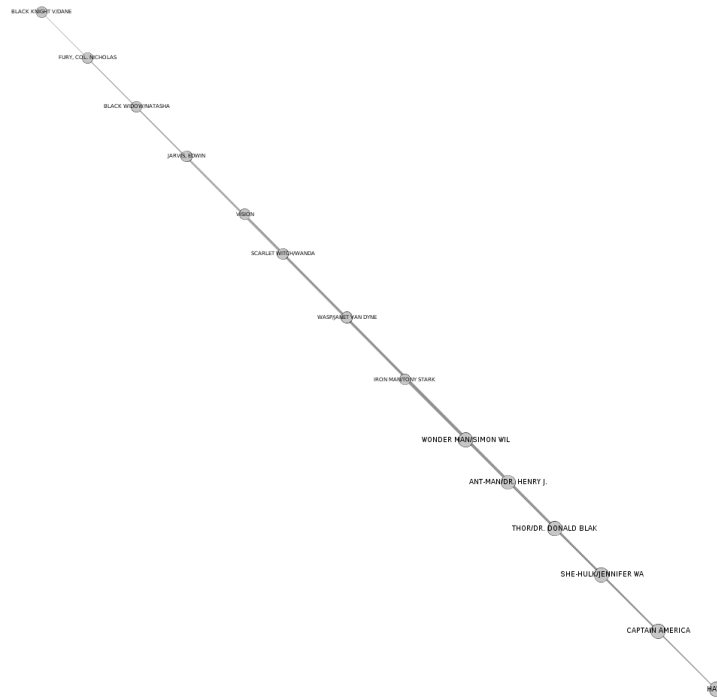


Figura 14. Cluster Xmen



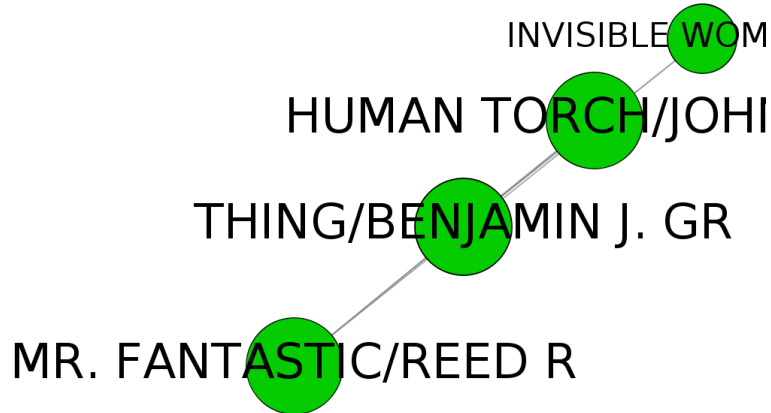


Figura 16. Cluster 4 fantásticos.

Para realizar este paso, se ha realizado un filtro INTERSECTION en el que se han incluido los filtros de rango con mínimo el valor de grado 1000 y el filtro Equal de Modularity class. Para este último, se han seleccionado los grupos que formaba cada cluster para poder visualizarlos por separado. Para el tamaño de cada subgrafo ha sido aplicado según la importancia de cada nodo.

10. Aplica las modificaciones y filtros que consideres oportunos para mejorar la visualización del grafo.

A continuación se visualizarán dos posibles casos, el grafo completo y el subgrafo con los superhéroes que han realizado al menos 1000 colaboraciones.

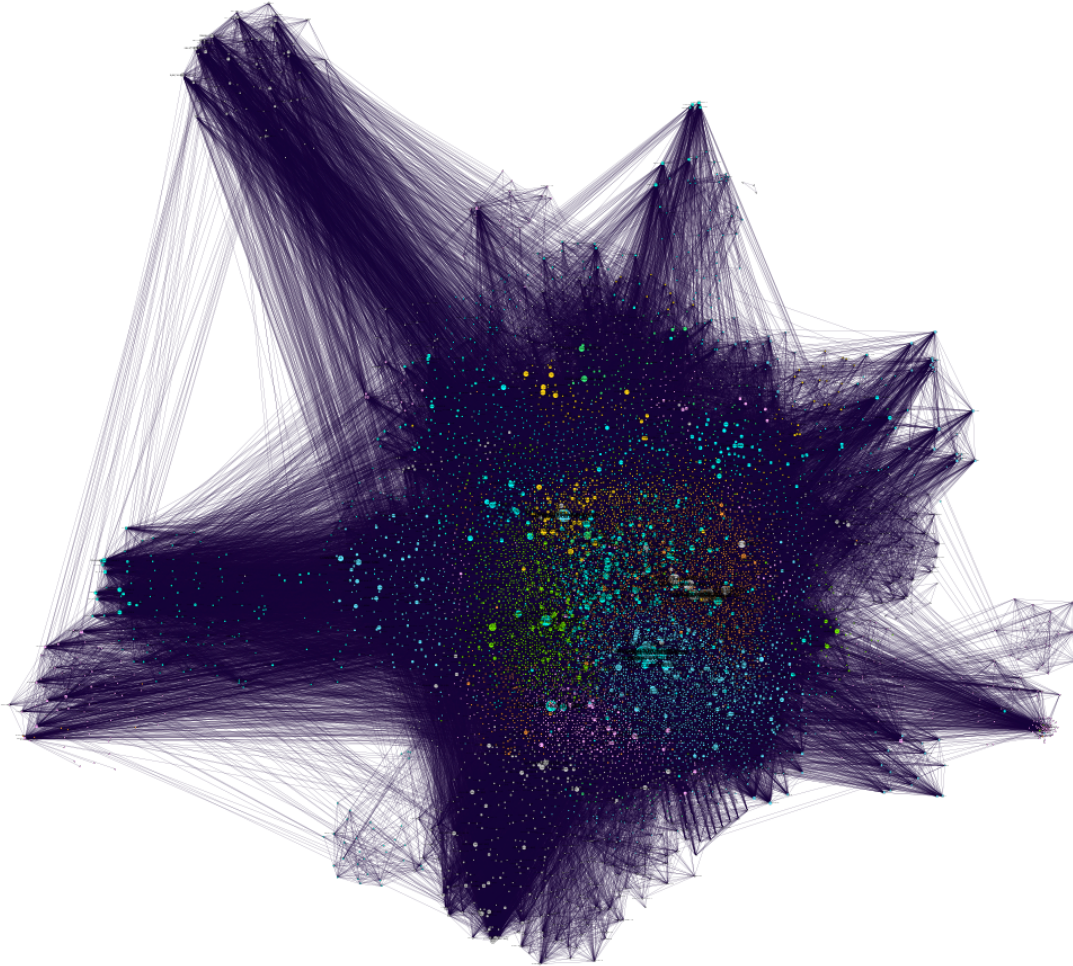


Figura 17. Clustering grafo total.

Las características para la mejora de visualización del grafo total han sido:

- Layout Forceatlas2 (características: prevent overlap, scaling 100)
- Tamaño según importancia, rango de 10 a 100 (Eigenvector Centrality)
- Aristas moradas
- Color de los nodos según ranking en Modularity Class
- Layout Fruchterman Reingold (características: area 20000, gravity 10, speed 10)

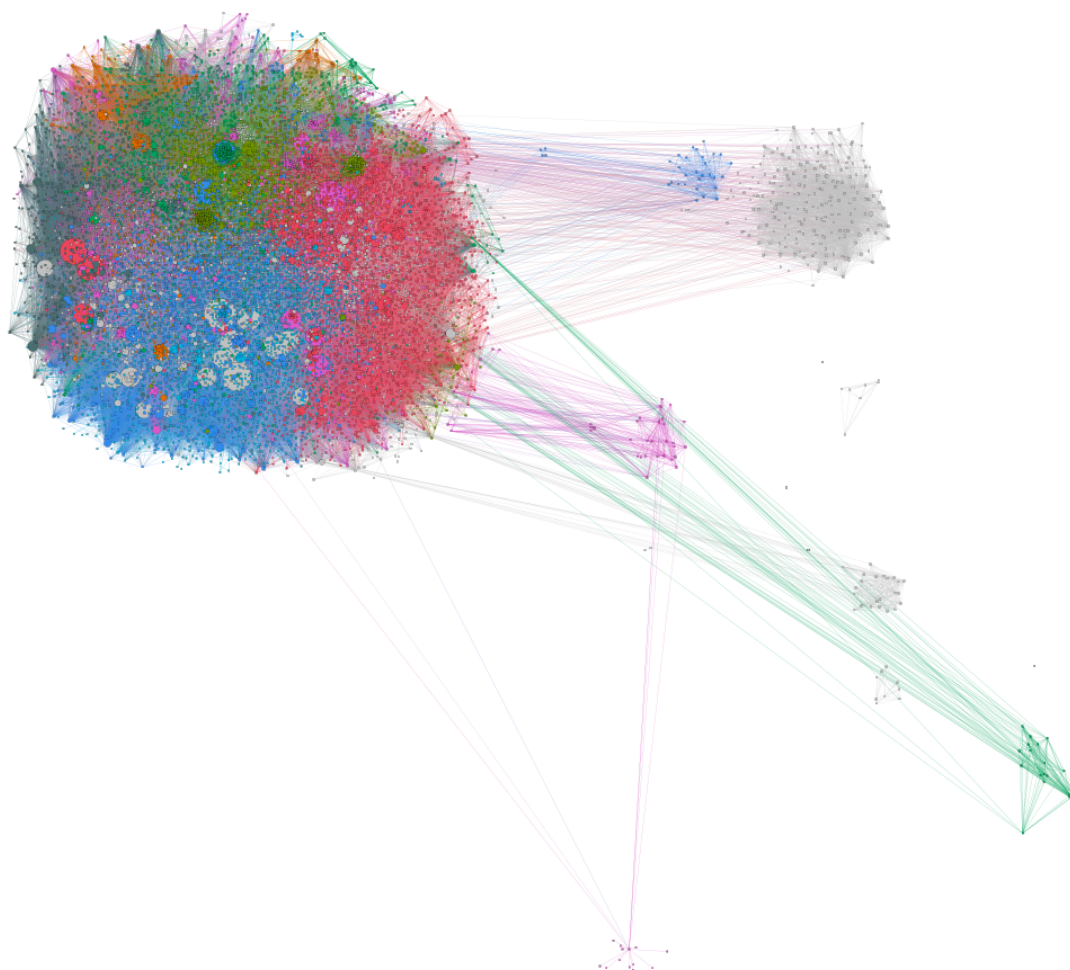
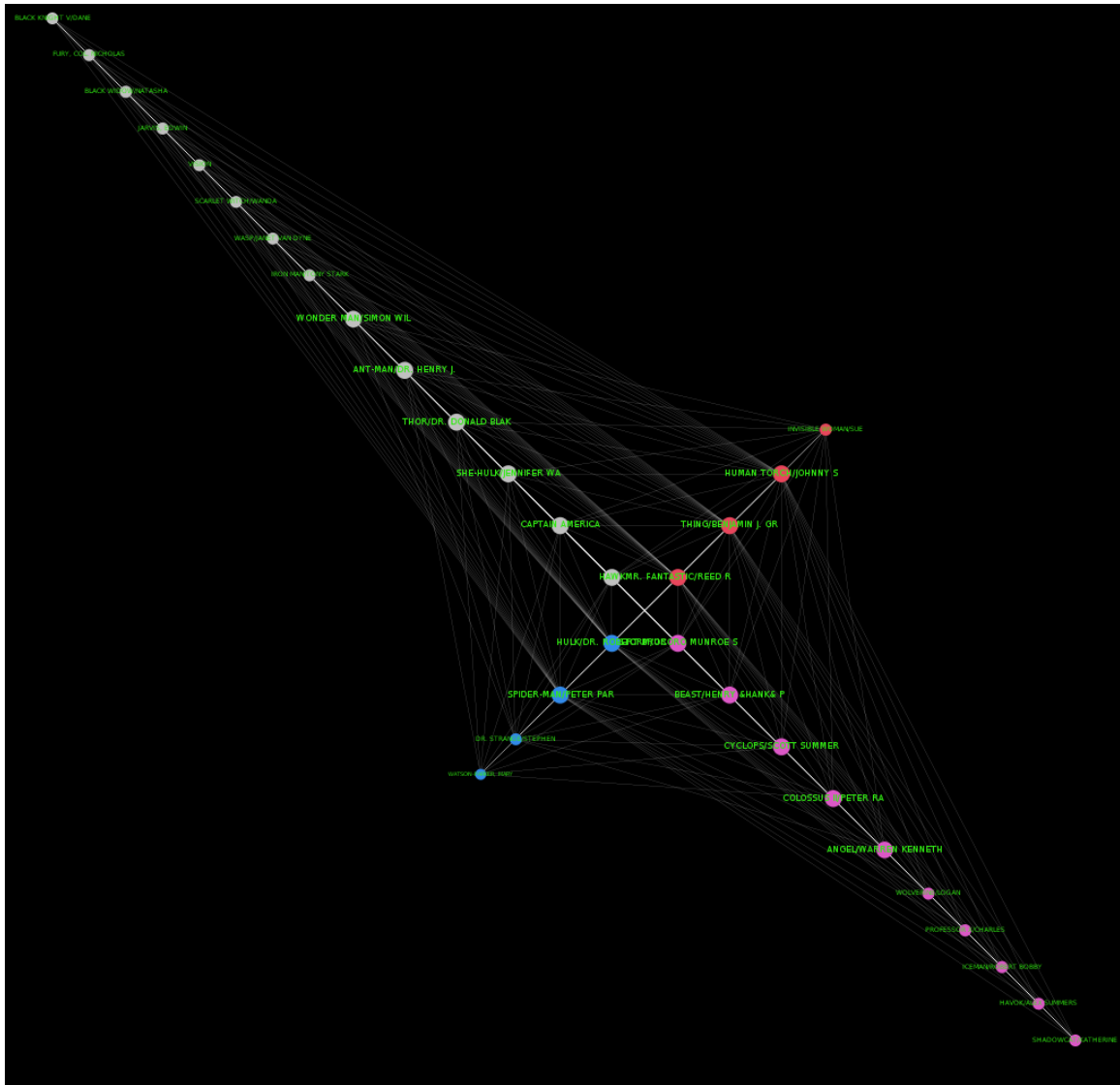


Figura 18. Clustering grafo total.

Para esta segunda opción de mejora del grafo completo, se ha aplicado un layout específico para grafos grandes, el Openord. Esta basado en el layout Fruchterman Reingold pero con un número fijo de iteraciones. Variando la característica Edge-cutting de 0 a 1, podemos hacer que este algoritmo se parezca más (0) o menos (1) a este algoritmo.

En concreto, se han utilizado las siguientes características:

- Liquid -> 90
- Expansion -> 90
- Cooldown -> 90
- Crunch -> 90
- Simmer -> 90
- Edge-cutting -> 1
- Iterations -> 3000



Para este segundo caso, se ha aplicado un layout Radial Axis, combinado con el layout de Expansion y Noverlap. Se ha querido resaltar las etiquetas con el verde fosforito sobre un fondo negro.

En cuanto al color, se ha aplicado la clasificación por Modularity class.