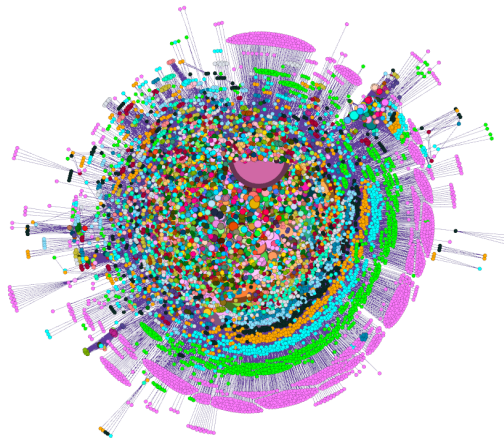


PRÁCTICA 1

ANÁLISIS DE UNA RED

~Red del Universo Marvel~



Grupo 1:

Pablo Álvarez García

Antonio Fernández Martín

Álvaro Corrochano López

Índice

1.-Introducción	2
2.-Tablas de valores significativos	3
3. -Visualizaciones significativas en Gephi	4
Gráficas de valores significativos	4
Distribución de grado	4
Distribución "PageRank"	4
Distribución de agrupamiento(clustering)	5
Distribución de excentricidad	5
Distribución "Harmonic closeness Centrality"	6
Distribución "Closeness Centrality"	6
Distribución de magnitud o tamaño(size)Distribución de centralidad adyacente (Betweenness centrality)	7
Representación gráfica	8
4.- Visualizaciones significativas en Cytoscape	13
Representación gráfica	13
Gráficas de valores significativos (con cytoscape)	14
Distribución de Grado	14
Excentricidad	14
Centralidad	15
Agrupamiento	15
Longitud media del recorrido más corto	16
5.- Visualizaciones significativas en Python	17
6.- Análisis de la red y conclusión	18
7.- Referencias	19

1.-Introducción

El objetivo de esta práctica es analizar un grafo de gran volumen de nodos (en este caso 10469) utilizando las herramientas vistas en clases y laboratorios.

-Gephi

-Cytoscape

-Network X (Python)

En la red se muestran todos los personajes que han aparecido en el universo de Marvel tanto en cómics, películas y series como nodos que están conectados mediante aristas que representan si los dos personajes han interactuado entre sí en algún momento.

The Marvel Social Network Networks of super heroes, constructed by Cesc Rosselló, Ricardo Alberich, and Joe Miro from the University of the Balearic Islands. Collected by [Infochimps](#) and transformed & enhanced by Kai Chang.

Enlace de descarga de la red: [GEPHI file](#).

2.-Tablas de valores significativos

Medidas Básicas	
Número de Nodos	10469
Número de Aristas	178115
Densidad L / Lmax	0,003
Núm. de componentes conexas	8

Grado	
Grado medio $\langle k \rangle = 2L / N$	34,027

Medias de distancia		
Distancia media	2,889	$\langle d \rangle$ =media de todas las distancias
Diámetro	7	dMAX= máxima distancia entre 2 nodos

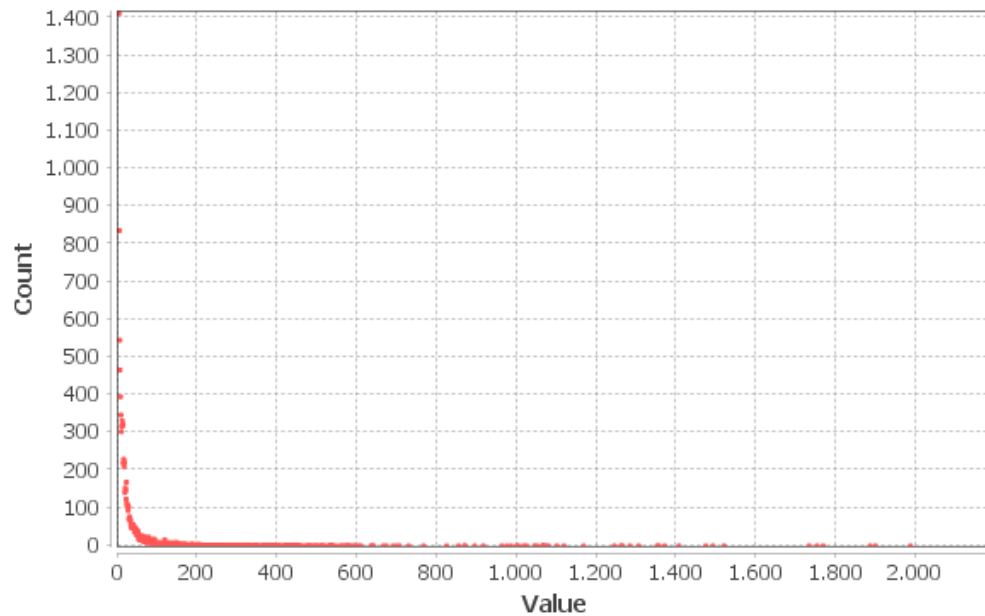
Clustering (Agrupamiento)		
Coeficiente de clustering medio	0,53	Total triangles: 2794418

Componente gigante	
Nodos	10440 nodos (99, 72% visibles)
Aristas	178049 aristas (99,96% visibles)

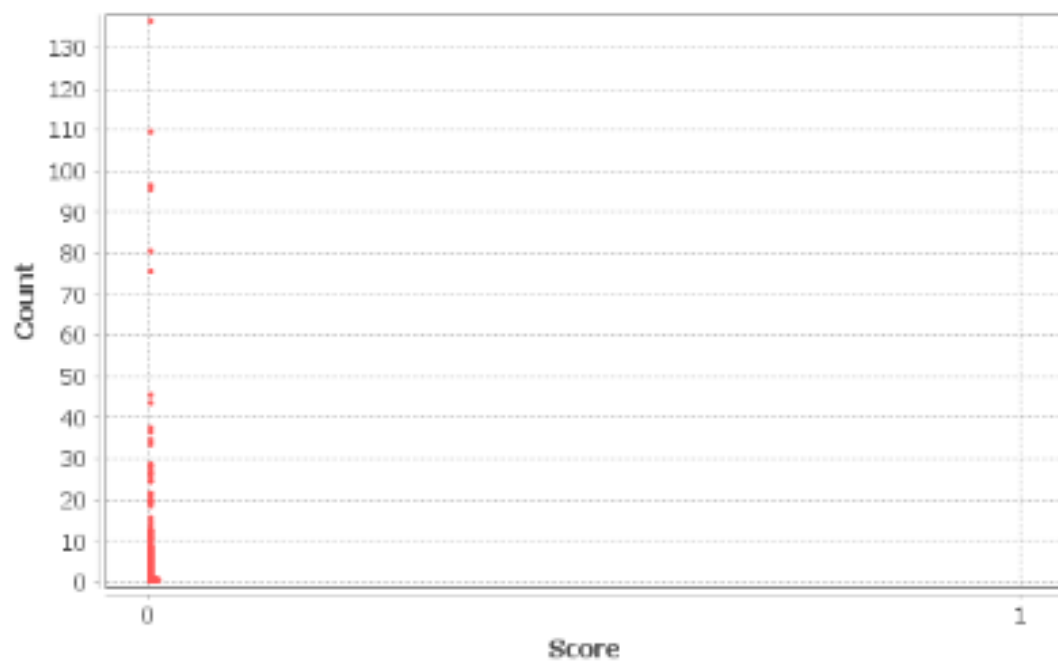
3. -Visualizaciones significativas en Gephi

Gráficas de valores significativos

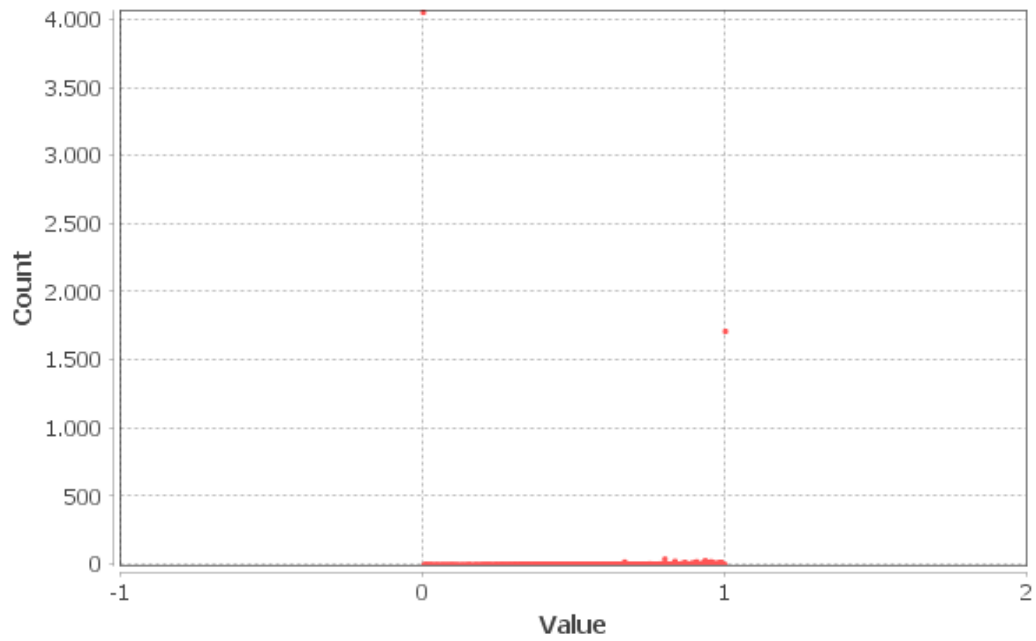
Distribución de grado



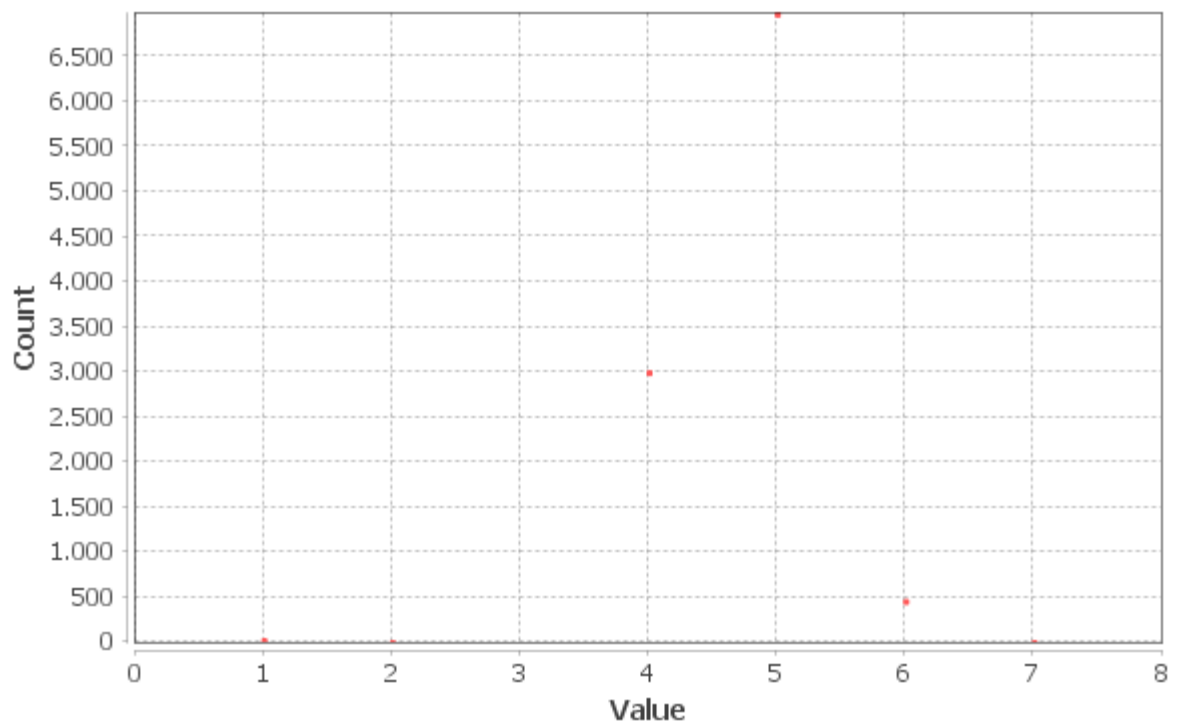
Distribución "PageRank"



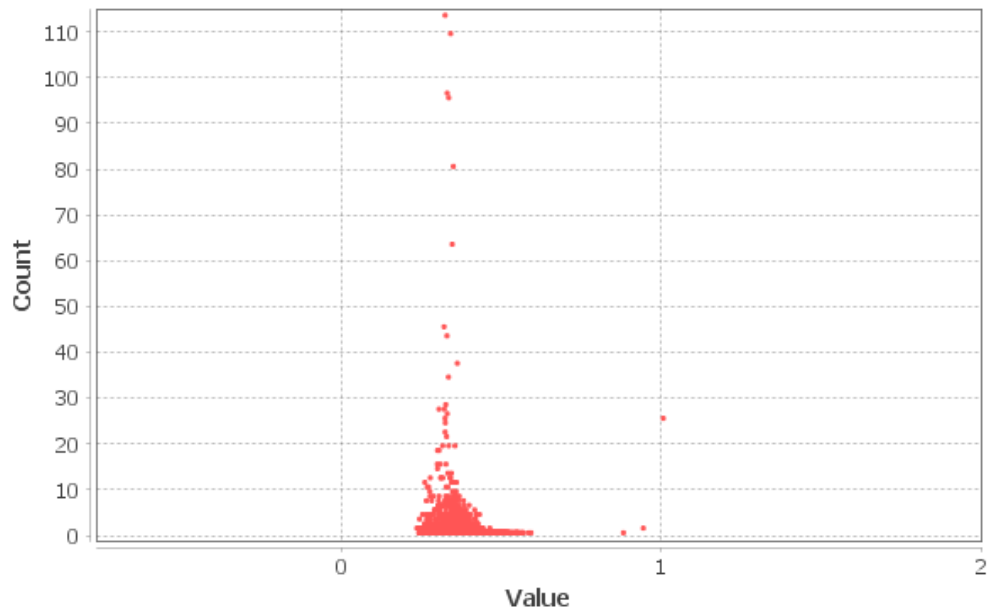
Distribución de agrupamiento(clustering)



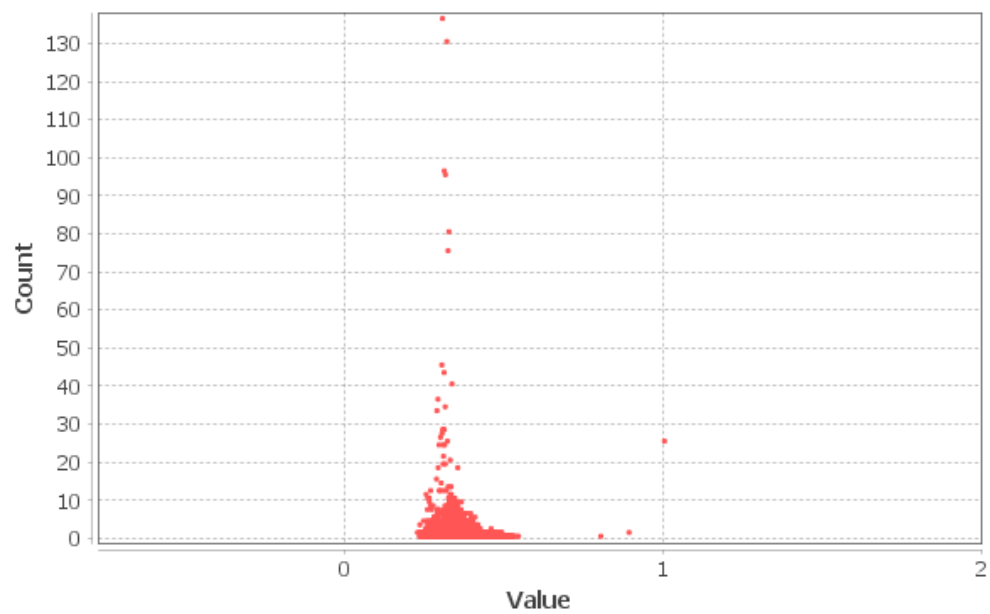
Distribución de excentricidad



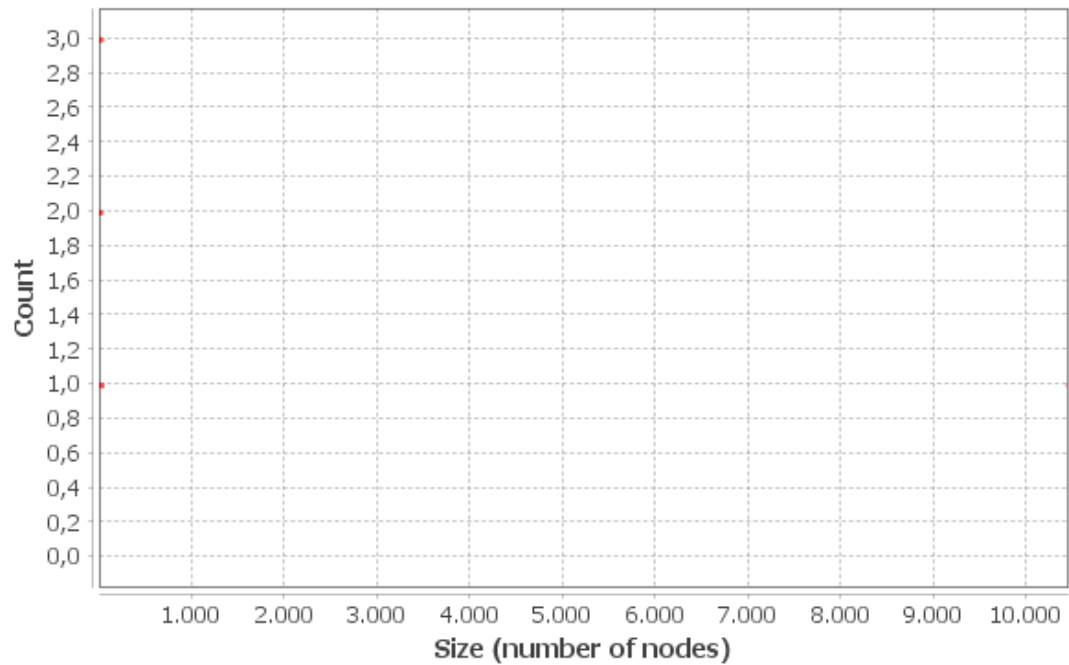
Distribucion “Harmonic closeness Centrality”



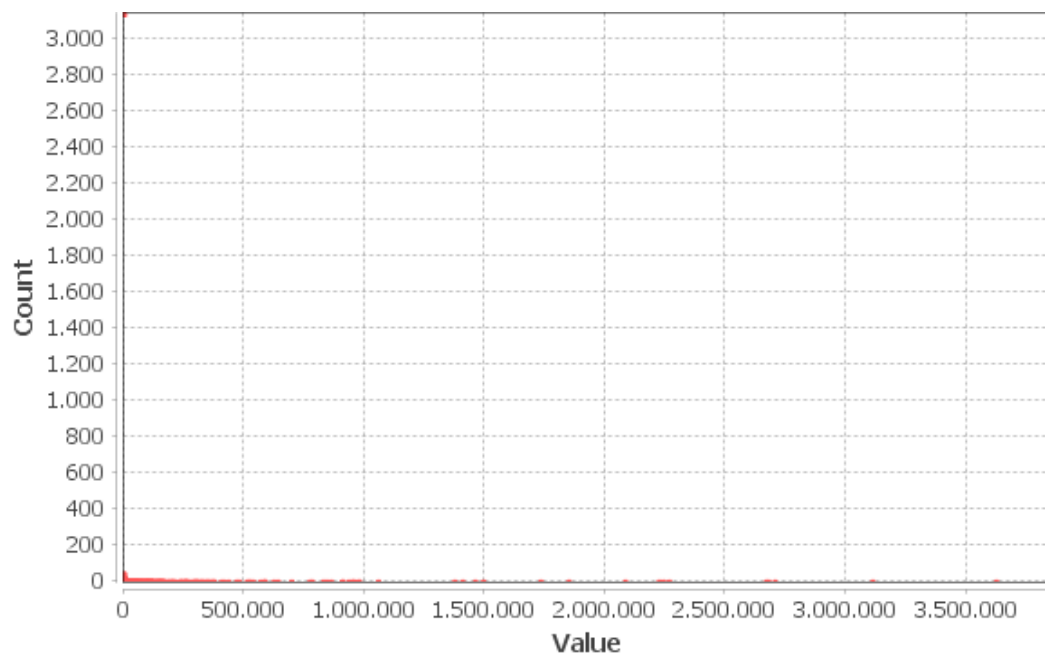
Distribución “Closeness Centrality”



Distribución de magnitud o tamaño(size)



Distribución de centralidad adyacente (Betweenness centrality)

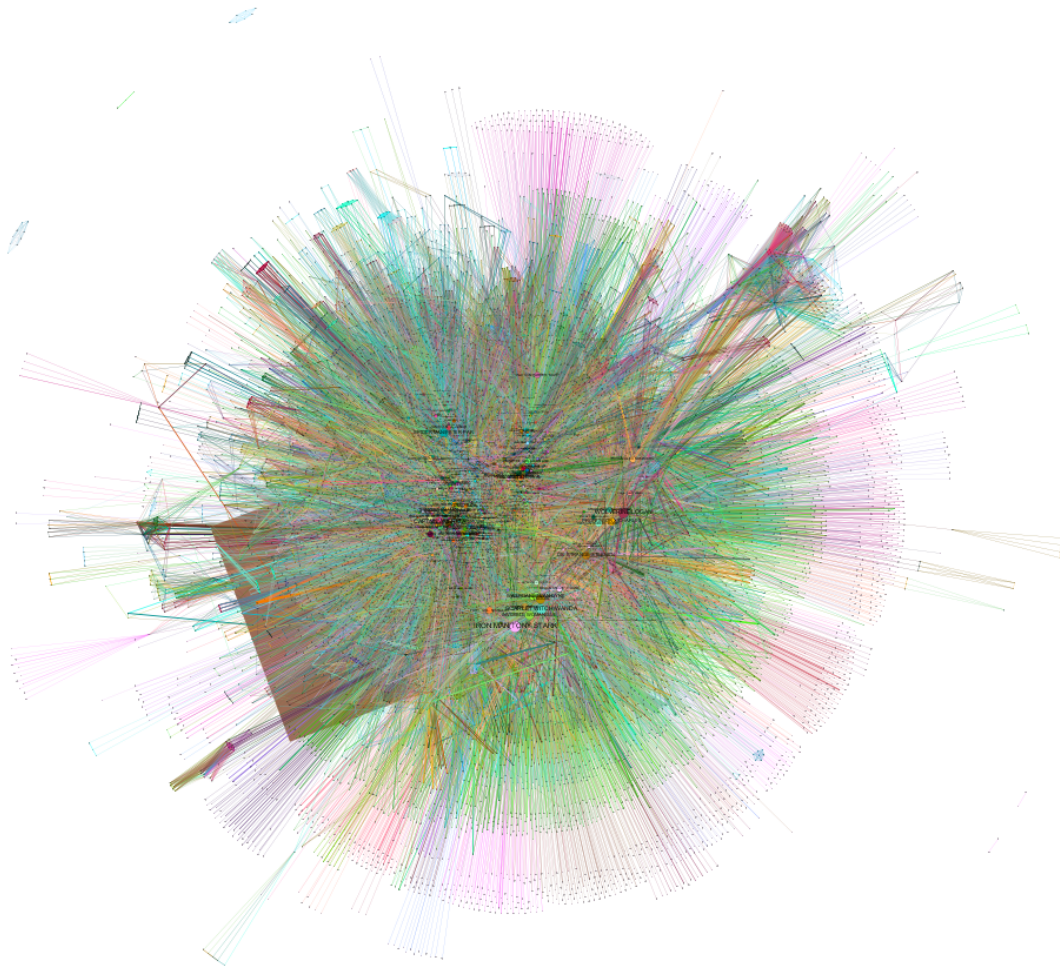


Empezamos con las visualizaciones en Gephi, luego en Cytoscape y por último en Python.

Representación gráfica

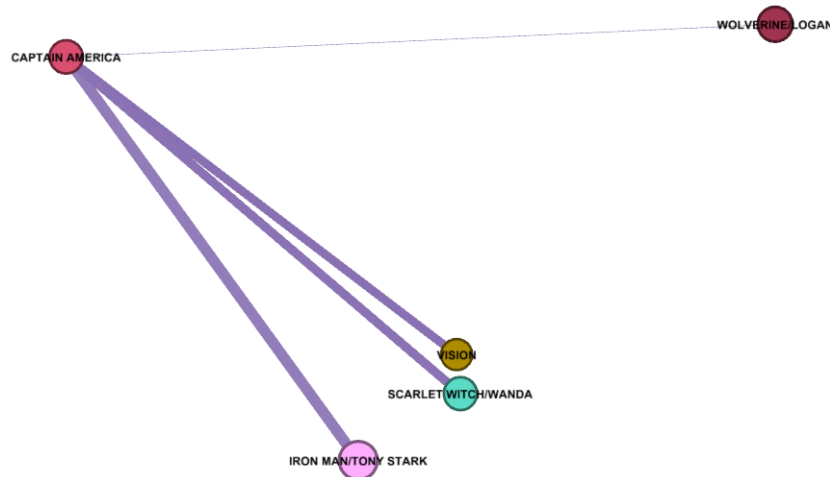
Aquí vemos el grafo entero, y gracias a la herramienta Gephi, hemos asignado colores según su grado.

Para distribuir el grafo, hemos usado el algoritmo de 'Fruchterman Reingold'.



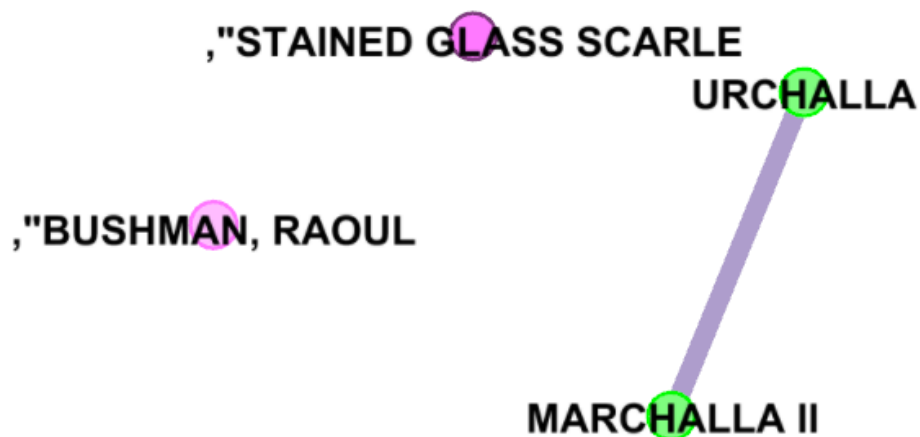
NODOS DE MAYOR GRADO

Estos son los nodos de mayor grado y nos indican los personajes que más relaciones tienen, siendo Iron Man el superheroe con mas relaciones (2189), seguido por Lobezno / Wolverine (1984), Capitán América (1896), Scarlet Witch (1883) y Visión (1765)



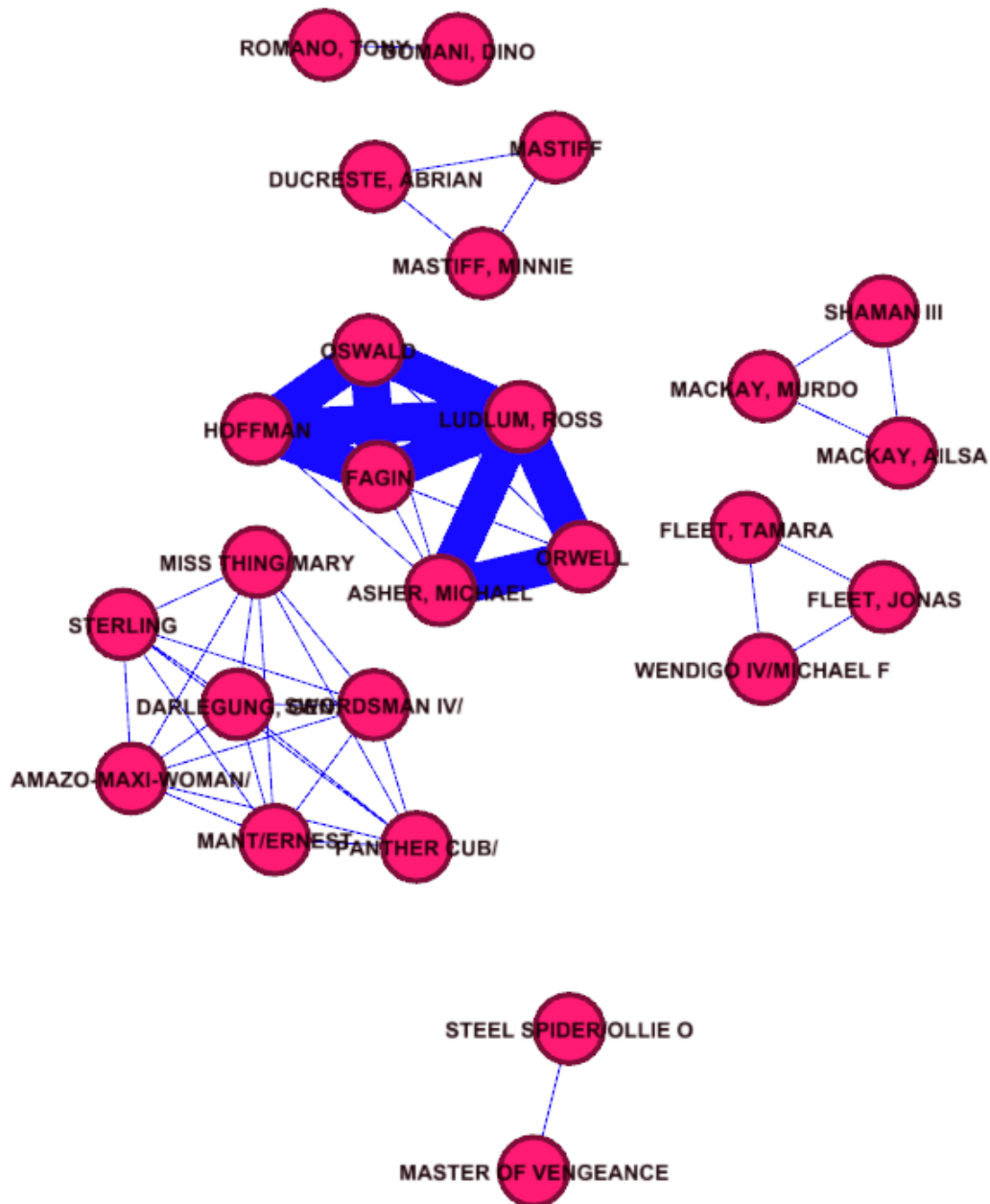
PERIFERIA

El gráfico siguiente muestra la periferia del grafo, que es el conjunto de nodos cuya excentricidad es igual al diámetro, que es 9.



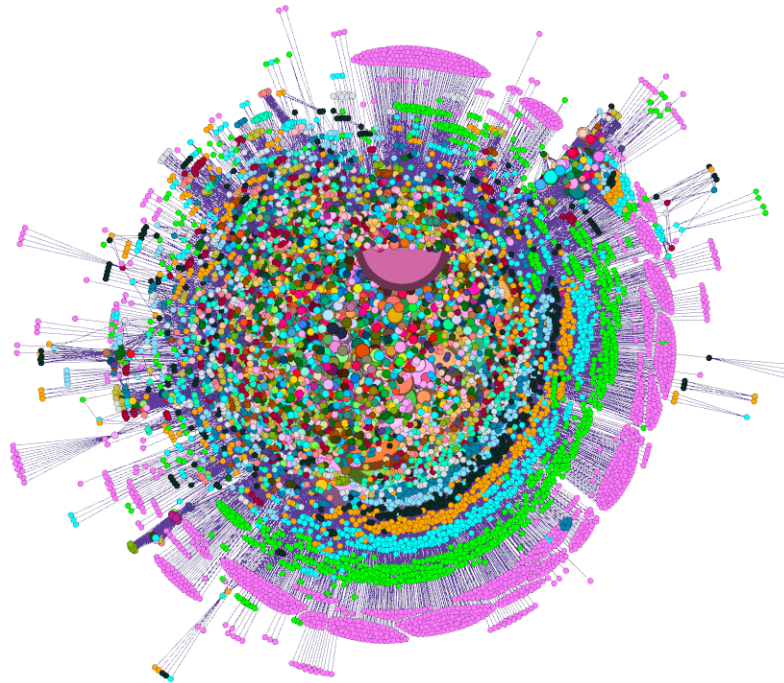
EXCENTRICIDAD

La excentricidad es la distancia máxima entre dos nodos siguiendo el camino mínimo. Por ejemplo, observamos que hay 26 nodos con excentricidad 1, como podemos ver en este gráfico.

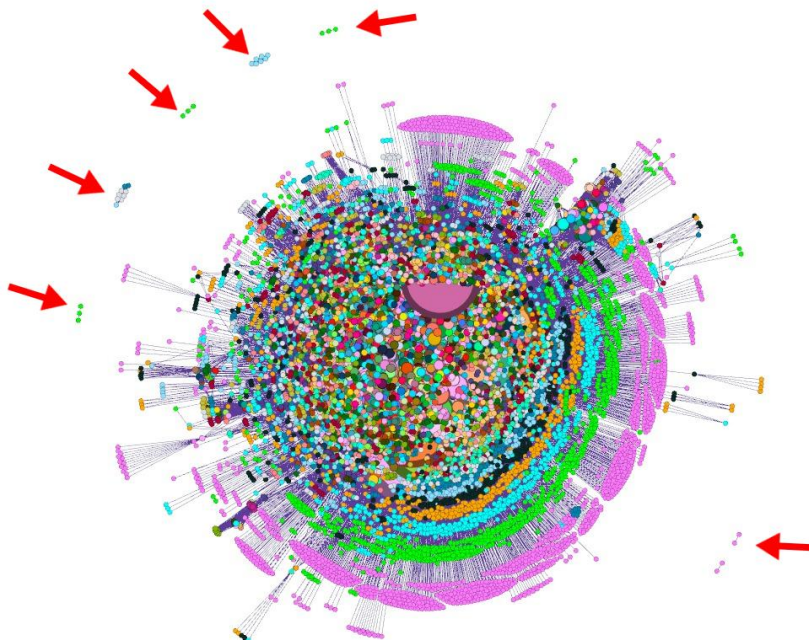


COMPONENTE GIGANTE

La componente gigante de nuestro grafo tiene 10440 nodos (99, 72% visibles) y 178049 aristas (99,96% visibles).



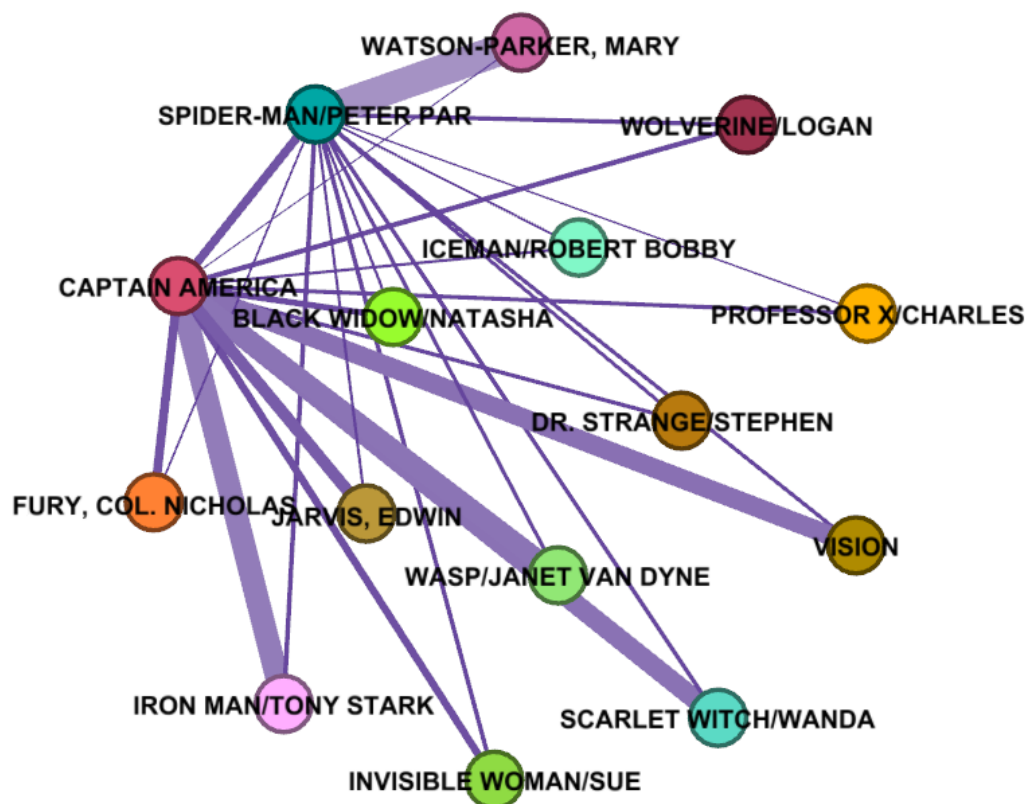
Aquí se muestran las componentes más pequeñas, pertenecientes a diversas series de cómics Marvel, que no establecen ninguna conexión con otro personaje externo.



PAGE RANK

Aquí mostramos el conjunto de nodos con mayor Page Rank.

PageRank
0.008739
0.007836
0.007237
0.006649
0.006225
0.005966
0.005928
0.005434
0.005199
0.005021
0.004743
0.004597
0.00456
0.004277
0.004252

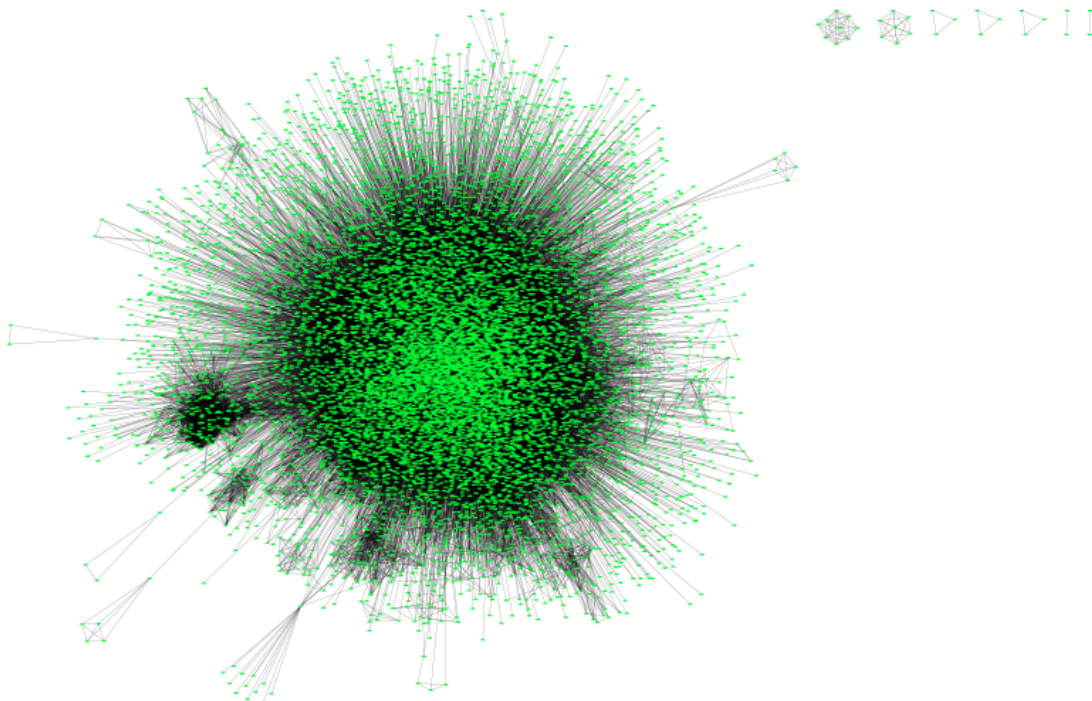


4.- Visualizaciones significativas en Cytoscape

Datos obtenidos tras analizar la red en Cytoscape:

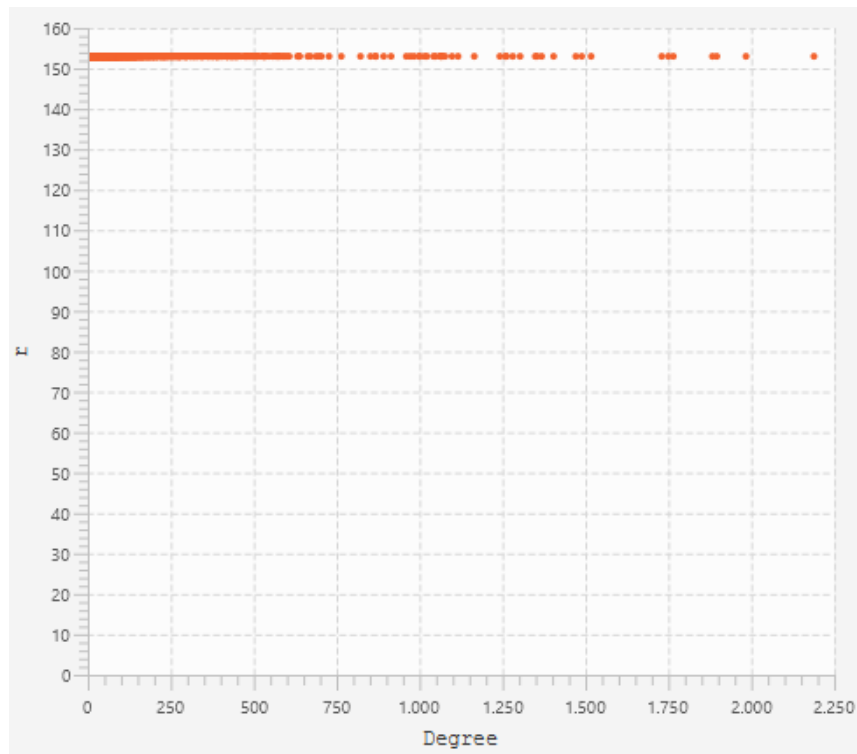
Summary Statistics	
Number of nodes	10469
Number of edges	178115
Avg. number of neighbors	34,109
Network diameter	7
Network radius	4
Characteristic path length	2,889
Clustering coefficient	0,457
Network density	0,003
Network heterogeneity	2,966
Network centralization	0,206
Connected components	8
Analysis time (sec)	85,821

Representación gráfica

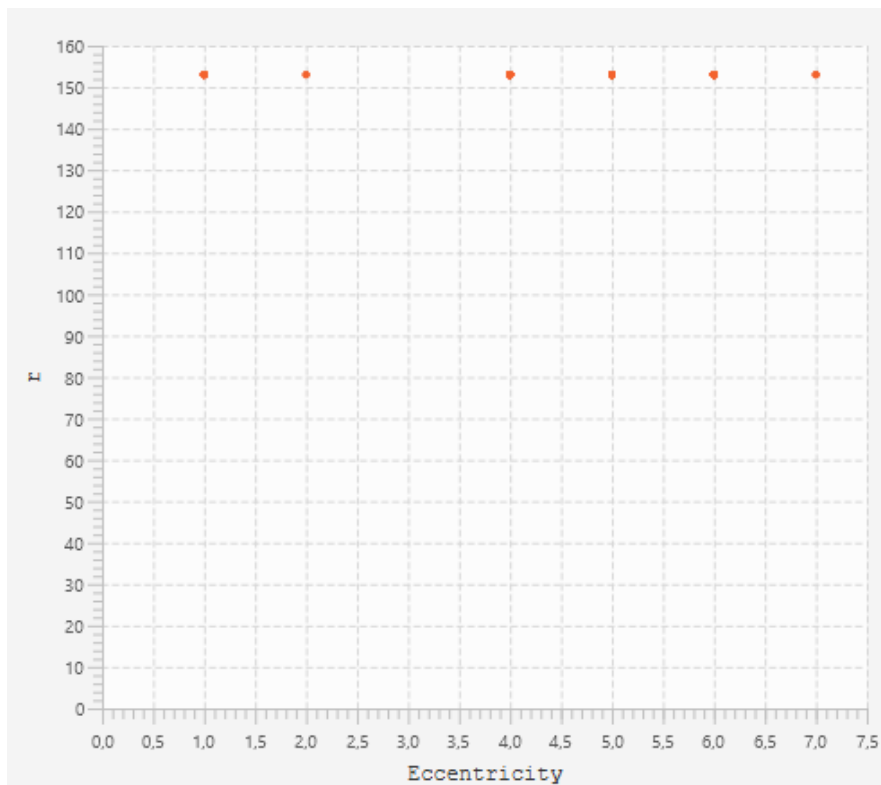


Gráficas de valores significativos (con cytoscape)

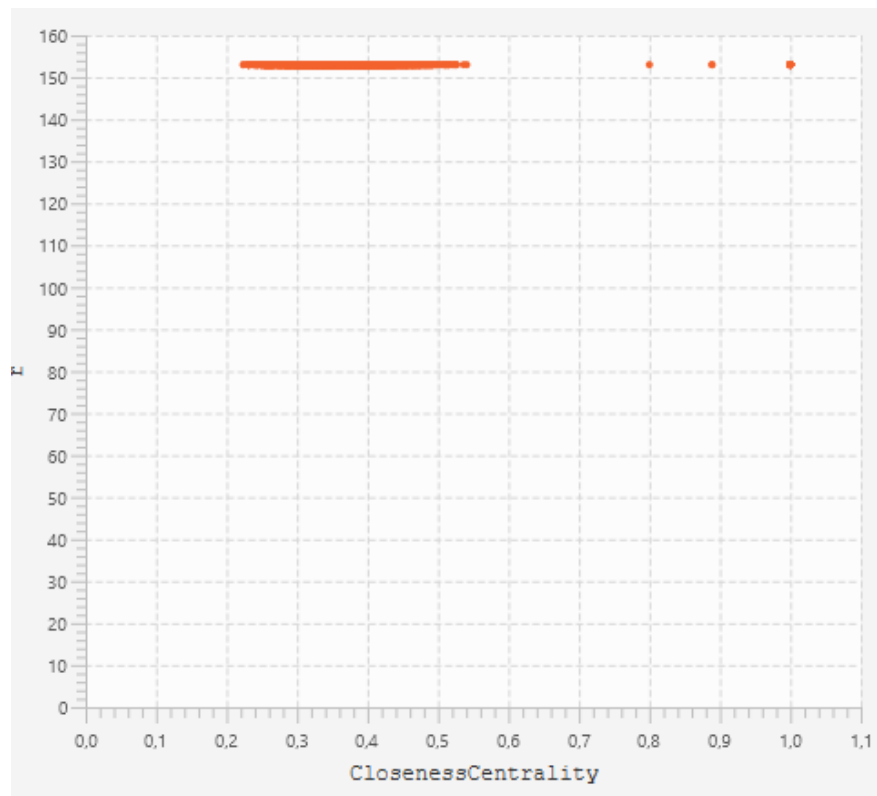
Distribución de Grado



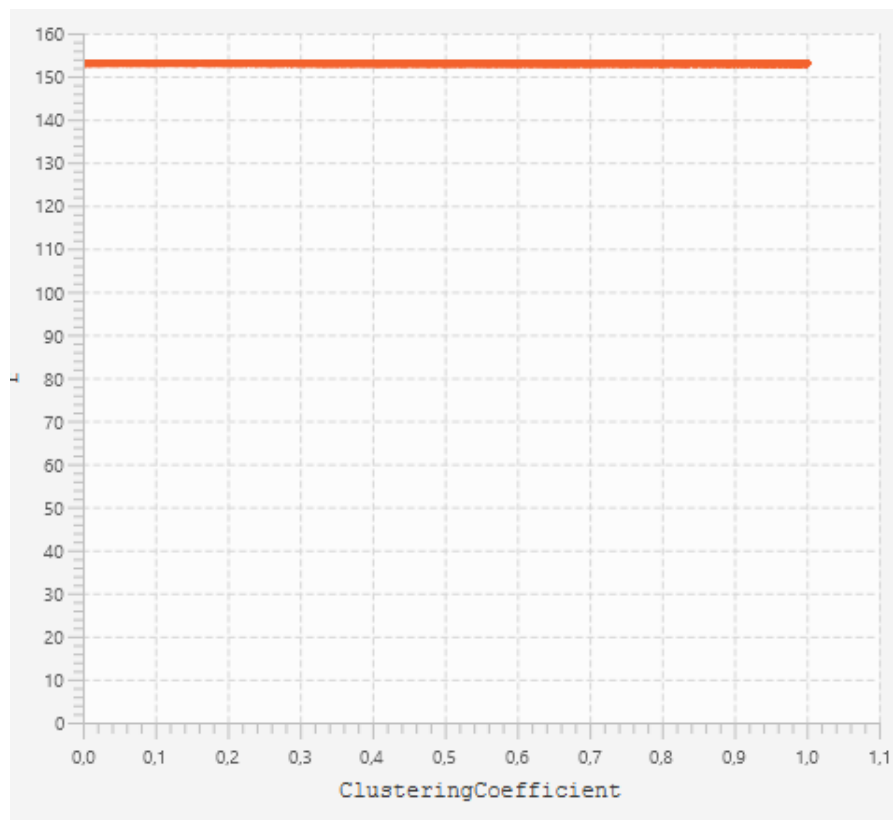
Excentricidad



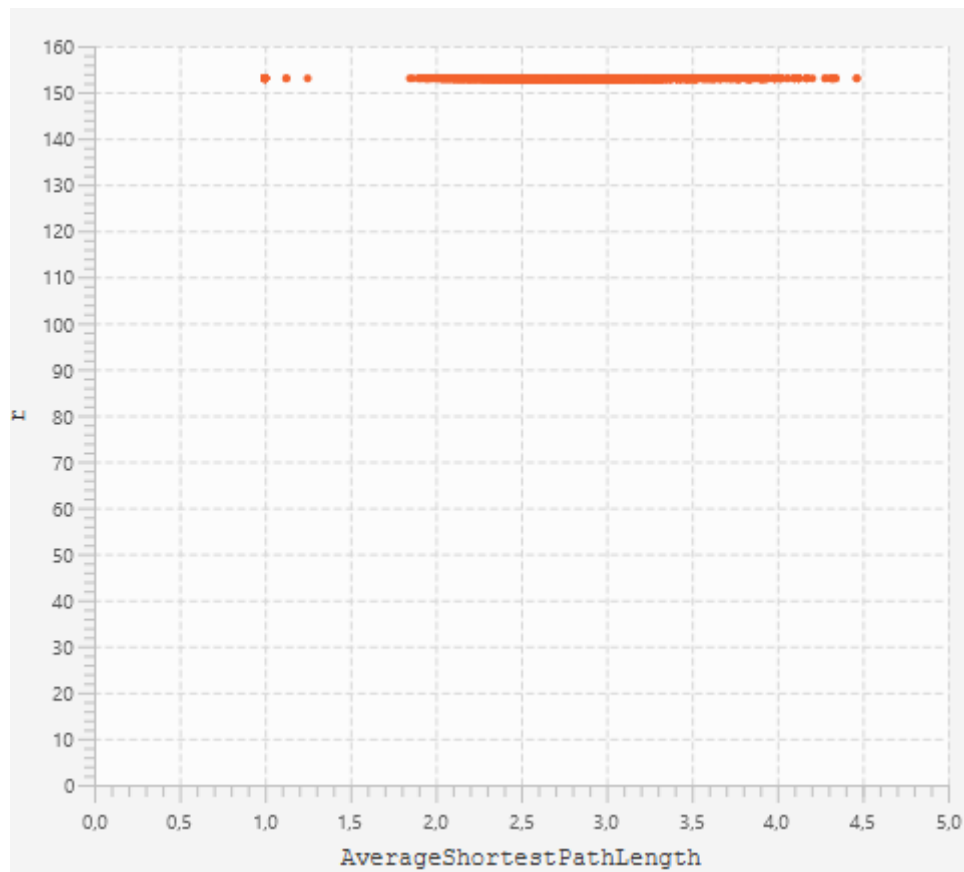
Centralidad



Agrupamiento



Longitud media del recorrido más corto



6.- Análisis de la red y conclusión

Analizando la topología de nuestro grafo con los valores que hemos obtenido, podemos sacar varias conclusiones sobre los personajes de Marvel y sus correlaciones.

A grandes rasgos podemos decir que es un grafo con nodos bastante relacionados entre sí, Este dato podemos comprobarlo mirando el grado medio que es 34,027, lo que implica que cada personaje de marvel de media aparece con otros 34 personajes distintos ya sea en películas, series o cómics.

El coeficiente de clustering medio es 0,530 lo que implica que de media cada personaje de Marvel comparte 5 películas/cómics/series con otro.

El personaje que ha aparecido con más personajes es Iron-man, ya que es el nodo de mayor grado con 2189 personajes relacionados.

El grafo cuenta con un alto nivel de modularidad, esto implica que las comunidades que existen están muy correlacionadas por lo que hay personajes que suelen compartir participaciones en series películas o cómics, esto lo comprobamos mirando que el número de comunidades débiles con poca correlación que únicamente son 8 (Size distribution table).

Por último podemos comprobar que los valores son muy irregulares y escasos en los grados más altos, ya que siendo el rango de grados desde 1 a 2189 la media de grados es 34,06.

7.- Referencias

- <https://github.com/gephi/gephi/wiki/Datasets>
- <https://www.youtube.com/playlist?list=PLlvIcfwy1T6IDiW3K10TpIK3rvdwMLOb2>
- <https://www.barabasilab.com/>