



UNIVERSIDAD COMPLUTENSE
MADRID

Práctica 1. Análisis de una red

Grupo 5: Eric Heresi Medina, Alberto García Doménech



Introducción

Para realizar esta práctica hemos seleccionado la red “*The Marvel Social Network Networks of super heroes*” realizada por Cesc Rosselló, Ricardo Alberich, y Joe Miro de la Universidad de las Islas Baleares.

La red está compuesta de más de 10.000 nodos y más de 170.000 aristas.

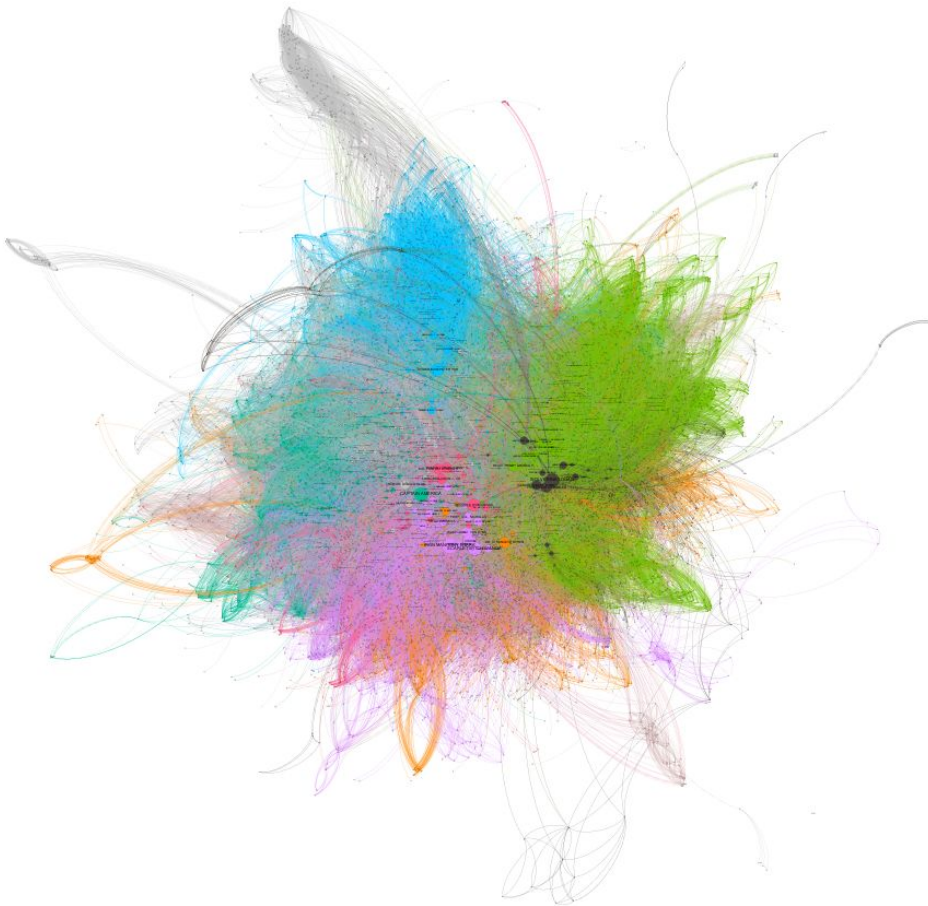
Los nodos son los personajes que aparecen en los cómics de Marvel hasta el año 2002 y las aristas el número de veces que coinciden dichos personajes (cuanta más veces coincidan mayor será el peso de la arista que comparten).

Para analizar y visualizar correctamente la red haremos uso de herramientas como Gephi, Python y CytoScape.

Tratamiento de datos en Gephi

La red estaba disponible en formato *.gephi* por lo que no ha sido necesario ningún preprocesamiento del archivo para poder importarlo en Gephi, donde hemos procedido al cálculo de diversos parámetros que nos serán útiles para poder filtrar, visualizar y en definitiva entender mejor la red.

Analizando inicialmente la red podemos observar que los nodos tienen un grado medio de 34.027 ($2 \times 178115 \div 10496$), una densidad de grafo de 0.003 ($178115 \div (10496 \times 10495 \div 2)$), una distancia media de 2.889, un diámetro de 7 y un total de 8 componentes conexas. Posteriormente hemos visualizado la red coloreando nodos mediante la métrica de *Modularity Class*, con las que hemos obtenido un total de 24 comunidades diferentes. El grosor de las aristas depende del peso de las mismas y el color depende del color de los nodos de los que originan. El layout escogido para esta primera visualización es *Force Atlas 2*.



Reducción de la red

Como podemos observar debido al gran tamaño de la red no podemos sacar conclusiones demasiado detalladas de la misma, por lo que vamos a proceder a reducirla utilizando como baremo el grado de los nodos ya que la gran mayoría de personajes que aparecen son muy poco conocidos y por tanto tienen menor grado. Esto se hace más visible si comparamos como de conocidos son los 5 nodos con mayor grado (Iron Man, Lobezno, Capitán América, Bruja Escarlata, Visión) con 5 de los nodos con menos grado (Doctor Dredd, Dominic Fortune, Mack Killburn, Steele, Chain).

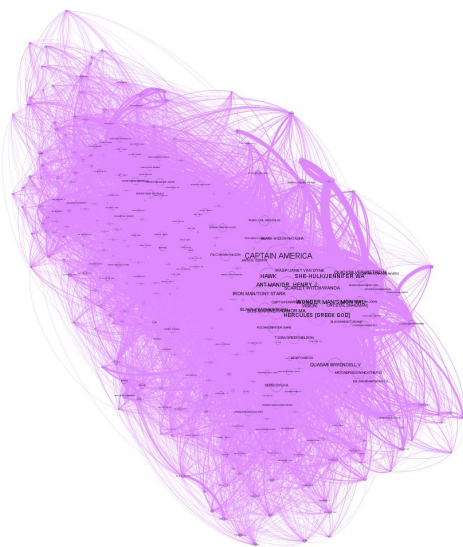
En primer lugar reduciremos el tamaño de la red aplicando el filtro de *Degree Range* > 70 obteniendo 1052 nodos y 69758 aristas.

Con la red más reducida volvemos a calcular los distintos parámetros de la red, obteniendo en este caso un grado medio de 132.62, una distancia media de 1.896, un diámetro de 4, una densidad de grafo de 0.126 y solamente 1 componente conexa.

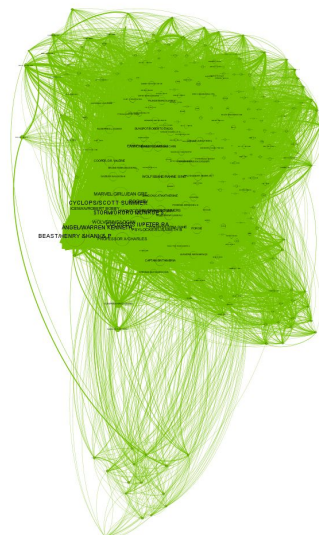
Volvemos a calcular las comunidades mediante *Modularity Class* obteniendo en este caso 9 comunidades y usando de nuevo el layout *Force Atlas*.



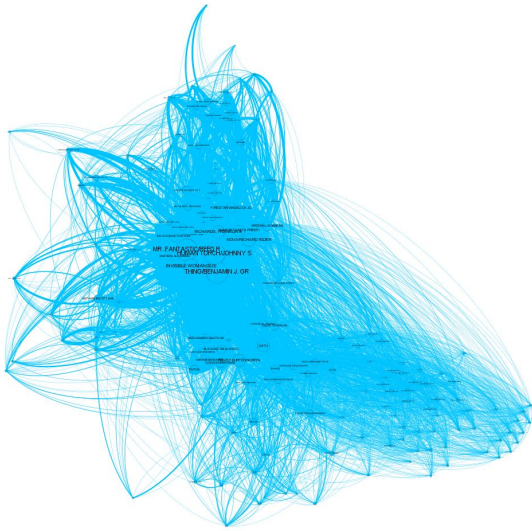
Vamos a analizar las diferentes comunidades para ver cómo funciona el algoritmo de Gephi:



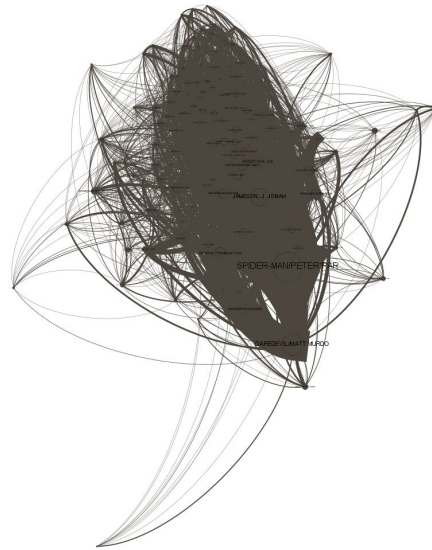
1.Los Vengadores



2.X-Men



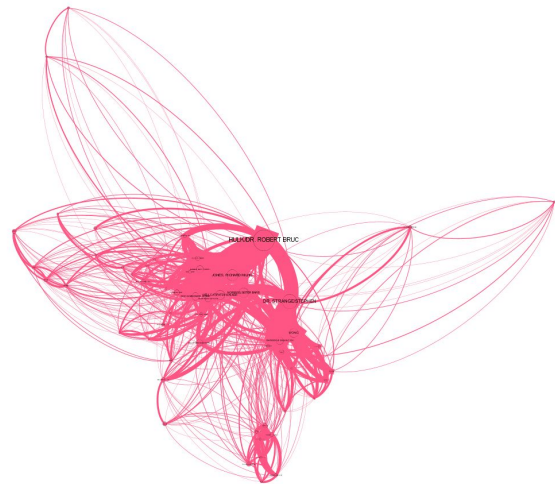
3.Los 4 Fantásticos



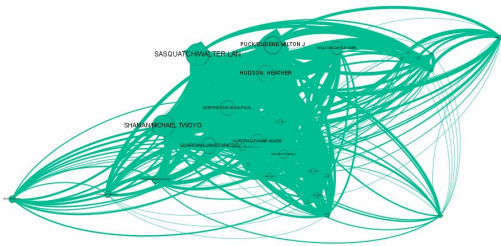
4.Spider-man



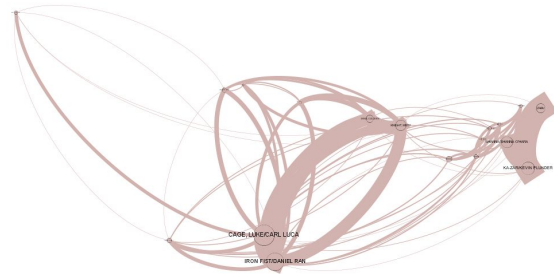
5.Thor



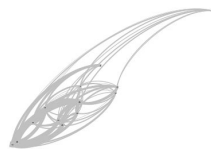
6.Hulk y Doctor Strange



7.Alpha Flight



8.Luke Cage y Iron Fist



9.Mutantes de otros universos

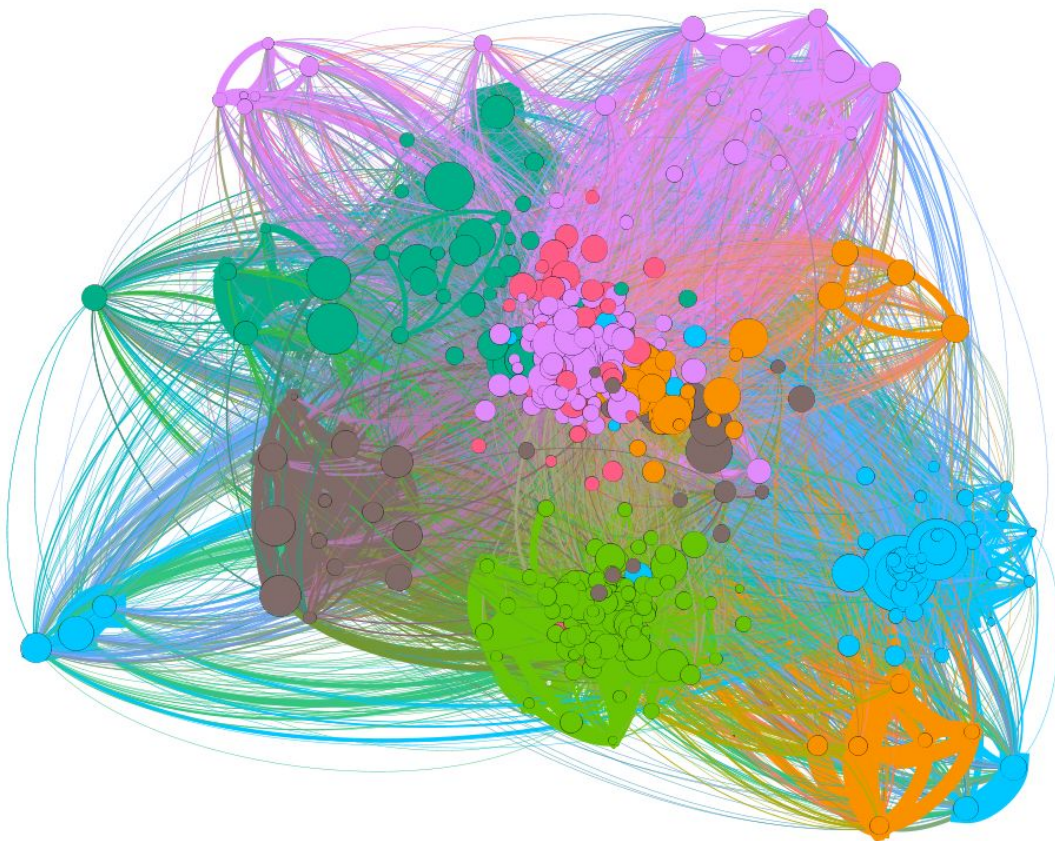
1. En la primera comunidad podemos ver nodos como Capitán América, Iron Man, Ant-man, Avispa, Quicksilver, Hawkeye, etc. que suelen salir en los números de los Vengadores. Esta comunidad es la más grande(28.61%) y la que tiene el nodo con más grado y con con más centralidad, Capitán América
2. En la segunda comunidad vemos nodos como Beast, Angel, Cíclope, Lobezno, Tormenta, etc. que conforman el grupo de los X-Men, uno de los más numerosos de la red.
3. En la tercera vemos nodos como Mister Fantastico, La Cosa, La Mujer Invisible y la Antorcha Humana que conforman Los 4 Fantásticos.
4. En la cuarta vemos a personajes como Spider-Man, Daredevil, J.Jonah Jameson, etc. del entorno de Spider-Man.
5. En la quinta vemos nodos del entorno de Thor como el propio Thor o Loki.
6. En la sexta tenemos a Doctor Strange y Hulk y personajes relacionados con los dos.
7. En la séptima vemos a Puck, Sasquatch, Northstar, etc. miembros del equipo canadiense Alpha Flight.

8. En la octava encontramos a Luke Cage, Iron Fist y diferentes personajes relacionados con ellos,
9. Por último en la novena y la comunidad más pequeña (1.24%) encontramos diversos personajes en su modalidad mutante como Reed Richards o Capitán América.

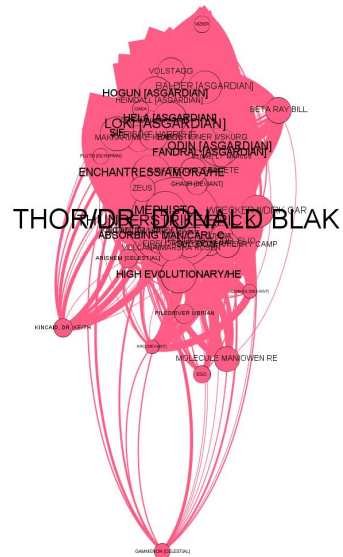
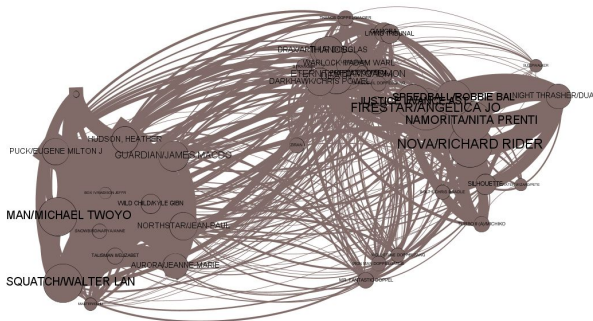
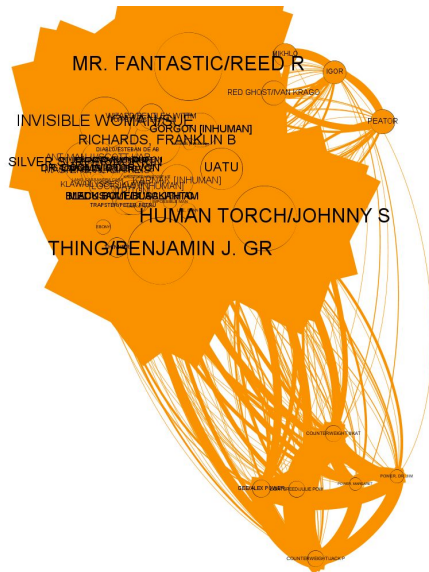
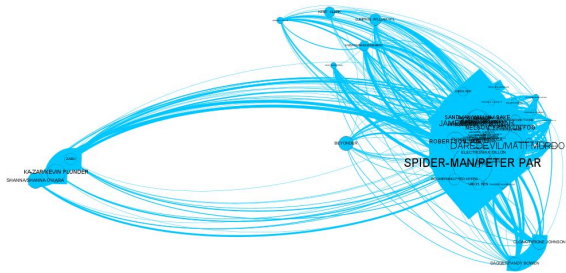
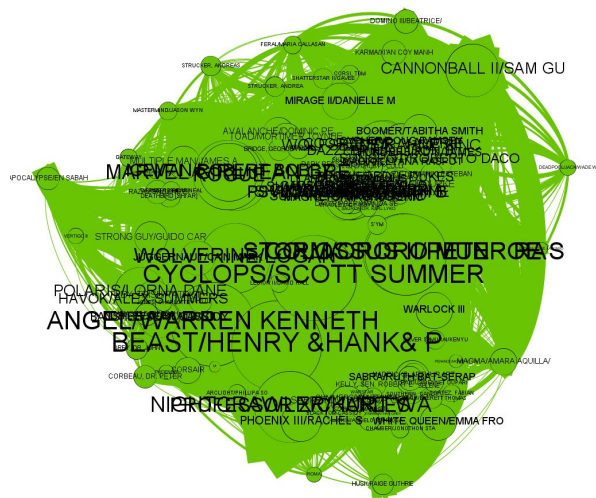
Aunque hemos reducido diez veces el tamaño de la red sigue teniendo nodos muy poco importantes por lo que vamos a reducir por última vez, esta vez aplicando el filtro de *Degree Range* > 130 (en la red original), reduciendo su tamaño a 522 nodos y 38496 aristas.

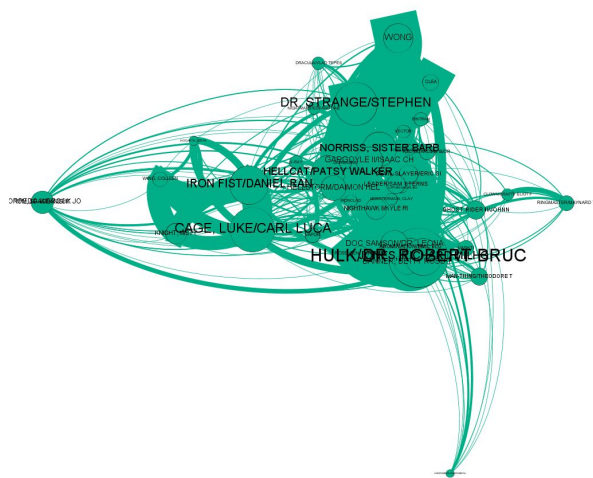
Debido a que hemos ido eliminando los nodos menos conectados, esta última red tiene una única componente conexas, el grado medio de los nodos es aún mayor que anteriormente siendo 147.494, la distancia media baja a 1.718, el diámetro desciende a 3 y la densidad del grafo es de 0.283.

Aplicando la métrica *Modularity Class* conseguimos 7 comunidades distintas. Esta vez, para visualizar el grafo utilizamos el layout OpenOrd.



Si comparamos las comunidades obtenidas con 522 nodos con las obtenidas con 1052 podemos ver que no varían demasiado:





Aunque ha desaparecido la comunidad de los mutantes de otros universos, las demás han seguido igual salvo la comunidad de Luke Cage y Iron Fist que se ha unido a la comunidad de Hulk y el Doctor Strange.

Tratamiento de datos en Python

Para poder utilizar la información sobre la red en Python hemos exportado la red inicial a dos archivos .csv, un archivo conteniendo la información de los nodos y otro con la de las aristas. Vamos a utilizar la librería Networkx para construir y estudiar el grafo.

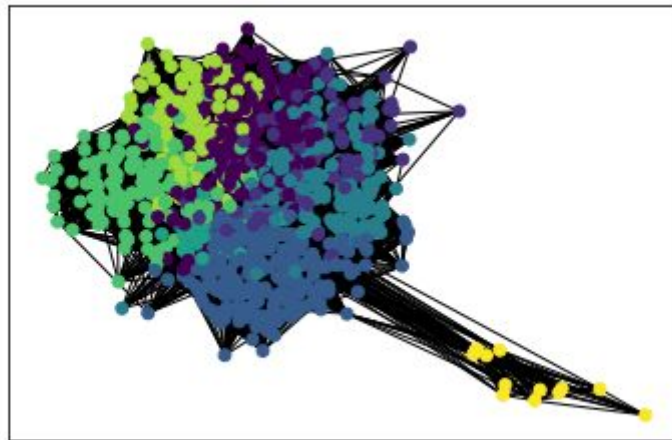
Procedemos a calcular distintas medidas de análisis y comprobamos que coinciden con las que proporciona Gephi. Como hemos visto anteriormente la red entera es demasiado grande por lo que procedemos a la reducción de su tamaño eliminando los nodos con grado menor a 70.

Dividimos en comunidades para comparar el algoritmo frente al usado por Gephi. Networkx nos devuelve 8 comunidades en vez de las 9 proporcionadas por Gephi:

	Id	Label	Degree	Group
7	10477	IRON MAN IV/JAMES R.	462	0
143	10613	HOGAN, VIRGINIA PEPP	255	0
144	10614	HOGAN, HAROLD J. HAP	214	0
673	11143	MANDARIN	213	0
771	11241	SITWELL, JASPER	201	0
	Id	Label	Degree	Group
6	10476	IRON MAN/TONY STARK	2189	1
86	10556	WOLVERINE/LOGAN	1984	1
71	10541	SCARLET WITCH/WANDA	1883	1
88	10558	VISION	1765	1
101	10571	INVISIBLE WOMAN/SUE	1750	1
	Id	Label	Degree	Group
155	10625	THOR/DR. DONALD BLAK	1281	2
152	10622	LOKI [ASGARDIAN]	420	2
440	10910	ODIN [ASGARDIAN]	416	2
438	10908	FANDRAL [ASGARDIAN]	390	2
542	11012	MEPHISTO	387	2
	Id	Label	Degree	Group
105	10575	BEAST/HENRY & HANK & P	1259	3
107	10577	CYCLOPS/SCOTT SUMMER	1070	3
96	10566	ANGEL/WARREN KENNETH	1064	3
91	10561	STORM/ORORO MUNROE S	1062	3
304	10774	COLOSSUS II/PETER RA	1015	3
	Id	Label	Degree	Group
62	10532	SPIDER-MAN/PETER PAR	1730	4
228	10698	WATSON-PARKER, MARY	1116	4
483	10953	DAREDEVIL/MATT MURDO	960	4
236	10706	JAMESON, J. JONAH	914	4
219	10689	ROBERTSON, JOE	638	4
	Id	Label	Degree	Group
100	10570	THING/BENJAMIN J. GR	1404	5
95	10565	MR. FANTASTIC/REED R	1368	5
277	10747	HUMAN TORCH/JOHNNY S	1349	5
60	10530	HULK/DR. ROBERT BRUC	1048	5
51	10521	SUB-MARINER/NAMOR MA	972	5
	Id	Label	Degree	Group
69	10539	HAWK	1165	6
359	10829	ANT-MAN/DR. HENRY J.	1077	6
87	10557	SHE-HULK/JENNIFER WA	1064	6
83	10553	WONDER MAN/SIMON WIL	1022	6
373	10843	HERCULES [GREEK GOD]	983	6
	Id	Label	Degree	Group
81	10551	CAPTAIN AMERICA	1896	7
1204	11674	FALCON/SAM WILSON	537	7
1017	11487	DUGAN, TIMOTHY ALOYI	489	7
1893	12363	SCREAMING MIMI/MELIS	430	7
1625	12095	MOONSTONE II/KARLA S	424	7

Aunque el algoritmo sigue agrupando a casi todos los mismos nodos, vemos como desaparece la comunidad de Luke Cage y como el Capitán América en vez de estar con los Vengadores tiene su propia comunidad con nodos como Falcon y Dugan propios de sus cómics.

Aunque no nos da tanta información como la visualización de Gephi podemos mostrar por pantalla la visualización que ofrece networkx utilizando el *spring layout* para conseguir un *Force-directed* grafo además de colorear las distintas comunidades con diferentes colores.



Para finalizar con el uso de Python vamos a reducir de nuevo la red para obtener 522 nodos como en Gephi, usando el mismo filtro de eliminación de nodos con grado menor a 130.

Como era de esperar los dos programas coinciden en las medidas de grado medio, densidad, distancia media etc.

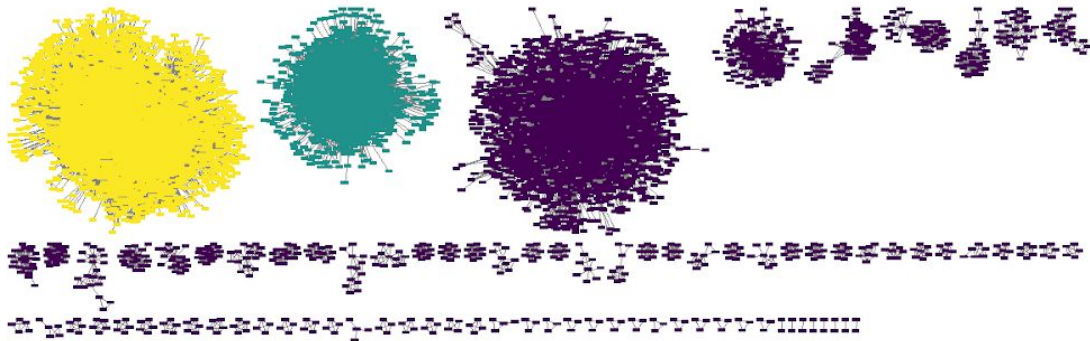
En cambio, el algoritmo usado por networkx solo diferencia en 4 comunidades frente a las 7 de Gephi. Vemos a quien agrupa:

	Id	Label	Degree	Group
6	10476	IRON MAN/TONY STARK	2189	0.0
81	10551	CAPTAIN AMERICA	1896	0.0
62	10532	SPIDER-MAN/PETER PAR	1730	0.0
65	10535	WASP/JANET VAN DYNE	1517	0.0
100	10570	THING/BENJAMIN J. GR	1404	0.0
	Id	Label	Degree	Group
86	10556	WOLVERINE/LOGAN	1984	1.0
209	10679	PROFESSOR X/CHARLES	1471	1.0
241	10711	ICEMAN/ROBERT BOBBY	1354	1.0
105	10575	BEAST/HENRY & HANK & P	1259	1.0
299	10769	HAVOK/ALEX SUMMERS	1098	1.0
	Id	Label	Degree	Group
71	10541	SCARLET WITCH/WANDA	1883	2.0
101	10571	INVISIBLE WOMAN/SUE	1750	2.0
695	11165	FURY, COL. NICHOLAS	1262	2.0
374	10844	BLACK WIDOW/NATASHA	1242	2.0
51	10521	SUB-MARINER/NAMOR MA	972	2.0
	Id	Label	Degree	Group
88	10558	VISION	1765	3.0
168	10638	DR. STRANGE/STEPHEN	1489	3.0
60	10530	HULK/DR. ROBERT BRUC	1048	3.0
1087	11557	BLACK KNIGHT V/DANE	1042	3.0
83	10553	WONDER MAN/SIMON WIL	1022	3.0

Esta vez agrupa a los nombres más conocidos como Iron Man, Capitán América y Spider-man juntos, separando a los 4 Fantásticos en varios grupos. Los otros grupos son de nuevo los X-Men, Hulk y Doctor Strange juntos otra vez, y gente variada relacionada con S.H.I.E.L.D como Nick Furia y Viuda Negra.

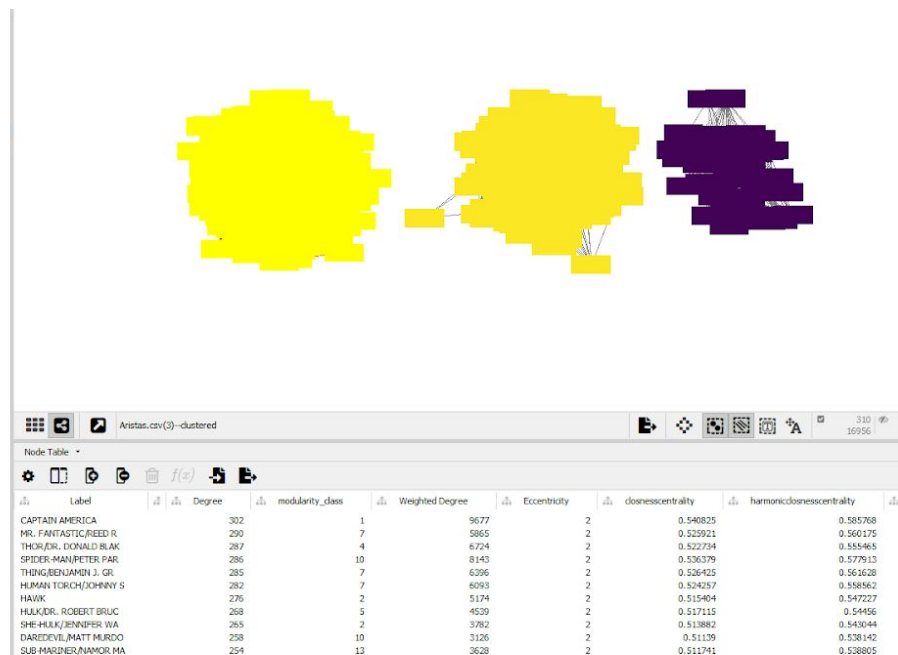
Tratamiento de datos en Cytoscape

Usaremos el programa Cytoscape para comparar el algoritmo de comunidades frente al de Gephi o al de Networkx.

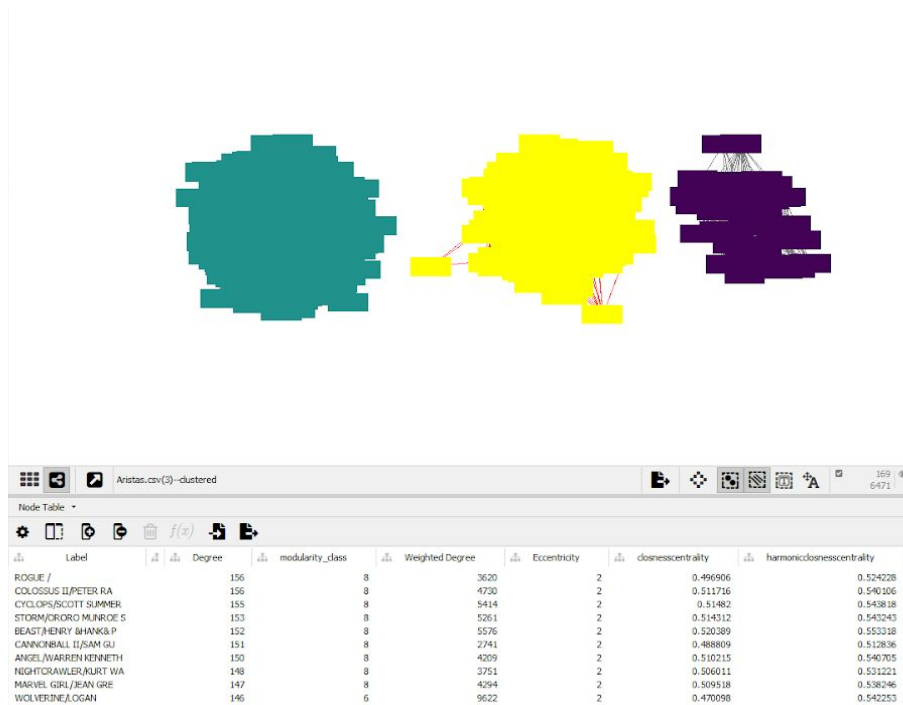


Podemos observar como con la red completa, el programa divide en demasiadas comunidades para que sean significativas cada una de ellas por lo que como en el resto de programas reduciremos el tamaño a 500 nodos.

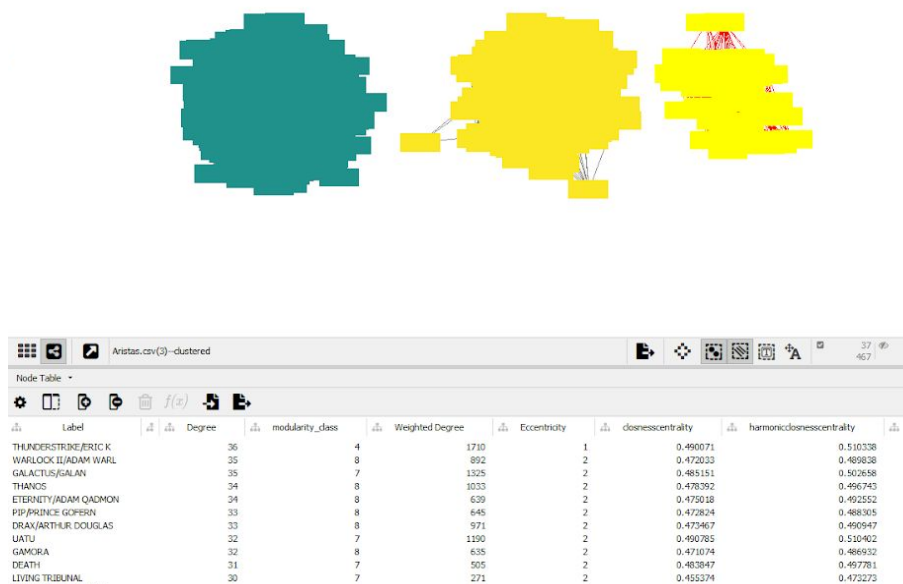
Con este número de nodos, obtenemos 3 comunidades:



La primera comunidad consta de nombres como Capitán América, Reed Richards, Spider-man, Thor,... por lo que la podríamos considerarla como el grupo con los nombres más “famosos”.



La segunda consta de nombres como Rogue, Lobezno, Cíclope, Tormenta, Jean Grey,... es decir, el grupo de los X-Men.



La última comunidad está formada por un conjunto de personajes que pertenecen al ámbito espacial o galáctico como Thanos,Uatu,Galactus,Gamora,etc.

Podemos ver que el algoritmo de clasificación en comunidades en Cytoscape es mucho más “permisivo” juntando en menos grupos a más nodos que no tienen tanto que ver, mientras que Gephi no disminuye tan drásticamente el número de comunidades, clasificando en grupos más compactos a los nodos.

Conclusión

Como conclusión podemos afirmar que los personajes de Marvel están fuertemente relacionados y que clasificando con herramientas como Gephi y Cytoscape conseguimos resultados fieles a la realidad teniendo en cuenta la alta complejidad de las relaciones entre ellos. Por último nos gustaría recalcar que si se volviese a hacer esta red a día de hoy habría mucha variación debido a que desde 2002 hasta hoy marvel ha realizado multitud de eventos relacionando a sus distintos universos.