

PEC 2 – ANÁLISIS DE REDES Y GRAFOS

Miguel Pérez Caro

1. Contexto

En el presente documento se va a hacer un análisis de las diferentes comunidades que se pueden formar en Twitter en torno a un partido de fútbol. Para ello se ha escogido el partido de liga que enfrentó al Sevilla FC y al FC Barcelona, el cuál tuvo lugar el pasado sábado 27 de febrero a las 16:15 en la ciudad de Sevilla. El partido enfrentaba a dos importantes equipos de la liga española, destacando el FC Barcelona que tiene un impacto mundial. Otro de los aspectos a destacar es que era un partido con alta tensión ya que los equipos se encontraban en la tercera y cuarta posición de la clasificación, superando el FC Barcelona al Sevilla por tan solo 2 puntos, por lo que una victoria del Sevilla les permitiría sumar 3 puntos y adelantarles en la clasificación. También añadía tensión al partido que ambos equipos se habían enfrentado en la ida de las semifinales de la Copa del Rey apenas 3 semanas antes, con resultado favorable para el Sevilla, y que 4 días después se volverían a enfrentarse para jugar la vuelta de las semifinales y así definir quién sería uno de los finalistas de la Copa del Rey. Todos los aspectos comentados hacían que existiera una gran tensión entre equipos y jugadores.

En este apartado del documento también es necesario comentar el transcurso del partido, ya que puede ayudar a entender mejor el análisis posterior, puesto que los eventos que transcurran en el partido seguramente tengan impacto en los tweets de los diferentes usuarios. El partido fue dominado por el FC Barcelona, que se adelantó en la primera parte con un gol de Ousmane Dembélé, y que en la segunda parte acabó afianzando su ventaja con un segundo gol, marcado por Leo Messi, que dejó el resultado final en un 0-2. Otros acontecimientos importantes durante el partido fueron la retirada por lesión del jugador del FC Barcelona Pedri, importante para su equipo, y la vuelta tras lesión del extremo del Sevilla Lucas Ocampos.

2. Objetivos

Una vez comentado el contexto del evento, es necesario mencionar los objetivos que se persiguen con dicho análisis:

- Obtener las diferentes comunidades de usuarios que se forman a través de los hashtags del partido. Observar las diferencias entre dichas comunidades.
- Analizar la importancia de cada equipo a través de la relevancia que tienen sus cuentas en las comunidades generadas.
- Analizar la relevancia que pueden llegar a tener los medios de comunicación. Comparar la importancia de cada uno.
- Por último, analizar si las casas de apuestas también adquieren un papel importante durante el partido de fútbol.

3. Metodología de Extracción

Para la extracción se hizo uso de la posibilidad que ofrece la herramienta Gephi mediante el plugin Twitter Streaming Importer que permite generar un grafo a partir de los tweets obtenidos en streaming según los hashtags definidos. Se escogió la modalidad *User Network* como lógica de red, de forma que los usuarios son representados como nodos, y las aristas dirigidas que los conectan representan a los usuarios que nombren en los tweets que se produzcan. Para dicha extracción se utilizaron los siguientes hashtags:

- #SevillaFCBarça
- #SevillaBarça

Para la elección de los hashtags, con el objetivo de que no se pueda cuestionar el por qué de su uso, se escogieron únicamente los hashtags usados por las cuentas oficiales de ambos clubes en español. El primero expuesto anteriormente fue el usado por la cuenta del Sevilla FC (@SevillaFC), y el segundo fue usado por la cuenta del FC Barcelona (@FCBarcelona_es).

De forma adicional, se revisaron los hashtags usados por otras importantes cuentas que ayudarán a entender los resultados y a establecer unas expectativas iniciales, con las siguientes conclusiones:

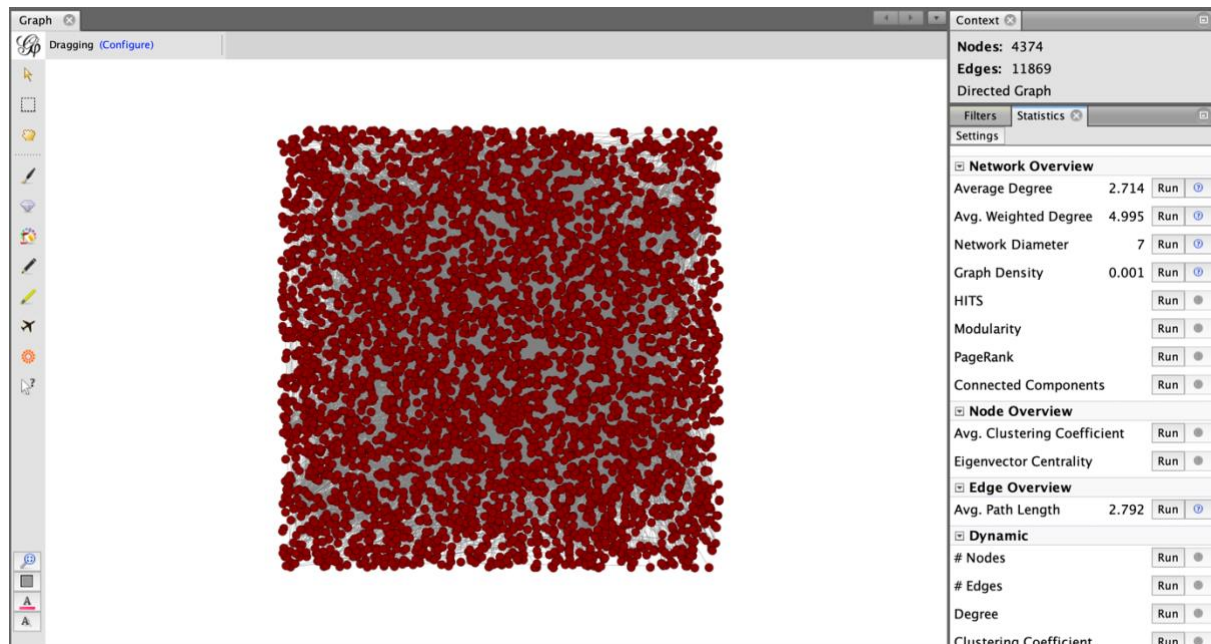
- Las cuentas de los clubes en otros idiomas usan los mismos hashtags que los usados por las cuentas en español. Por lo tanto, se espera que aparezcan estas cuentas y usuarios a nivel mundial.
- La televisión con los derechos en España para este partido es Movistar, cuyo hashtag era #LaCasadelFutbol, que es general para toda la jornada. Al no haberse incluido, no se espera que aparezca ninguna interacción relacionada con Movistar.
- Todos los periódicos y programas deportivos a nivel nacional que se revisaron usaban uno de los dos hashtags comentados, por lo que es lógico pensar que vayan a aparecer.

Nuevamente, se repite que el uso de los hashtags se limita a los escogidos por las cuentas oficiales en España de ambos equipos, y que las conclusiones anteriores son un mero análisis previo a los resultados para poder tener una idea inicial de lo que se podría obtener.

La extracción se realizó desde las 16:00, 15 minutos antes del comienzo del partido, hasta las 18:15, 10 minutos después de la finalización del mismo. De esta forma se espera haber capturado las interacciones que se realizaron en directo, para poder comenzar con el posterior análisis.

4. Análisis del Grafo

Al realizar la extracción mediante el plugin Twitter Streaming Importer, se genera el grafo directamente, cuyo resultado fue el siguiente:



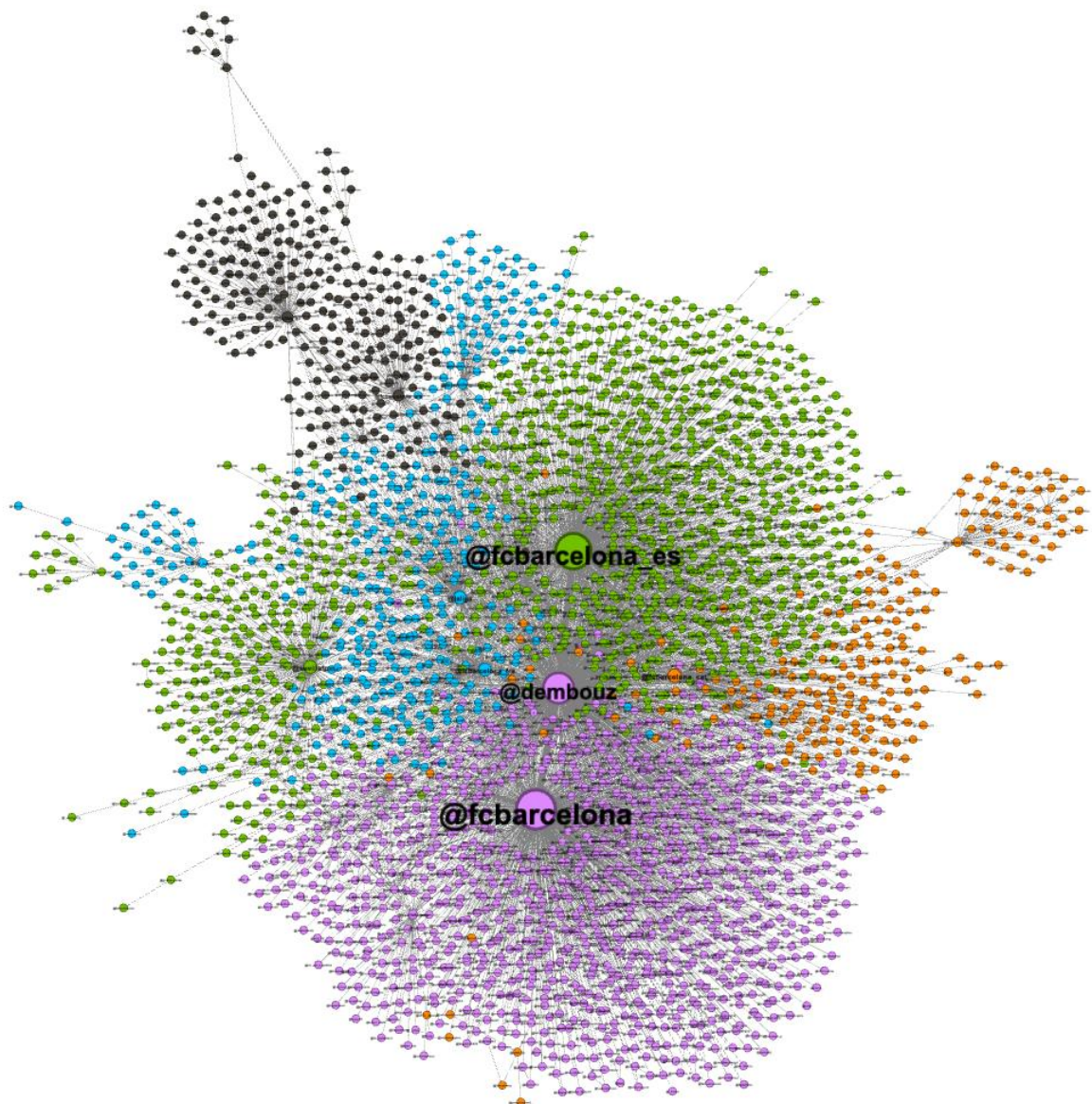
Como se puede observar en la imagen, el grafo consta de 4374 nodos y de 11869 aristas. Tiene un grado medio de 2.7 y de prácticamente 5 si se considera el grado con peso, por lo que no es muy bajo. El diámetro es 7, lo que indica que no hay que dar muchos saltos para llegar de un nodo a otro y la densidad es muy baja, lo que muestra que para el número de nodos hay pocas conexiones.

Para comenzar con el análisis del grafo, se crean las comunidades, con un resultado total de 95, aunque obviamente muchas no serán relevantes, y se colorean los nodos según a la partición a la que pertenecen. A continuación, se muestran las comunidades más importantes del grafo:

25	(24.12%)
0	(23.96%)
17	(8.3%)
2	(5.53%)
8	(5.14%)
22	(4.37%)
9	(4.09%)
26	(4.07%)
10	(2.79%)
36	(2.74%)
37	(2.49%)
41	(1.76%)
7	(1.67%)
39	(1.35%)
59	(0.75%)
5	(0.55%)
67	(0.5%)

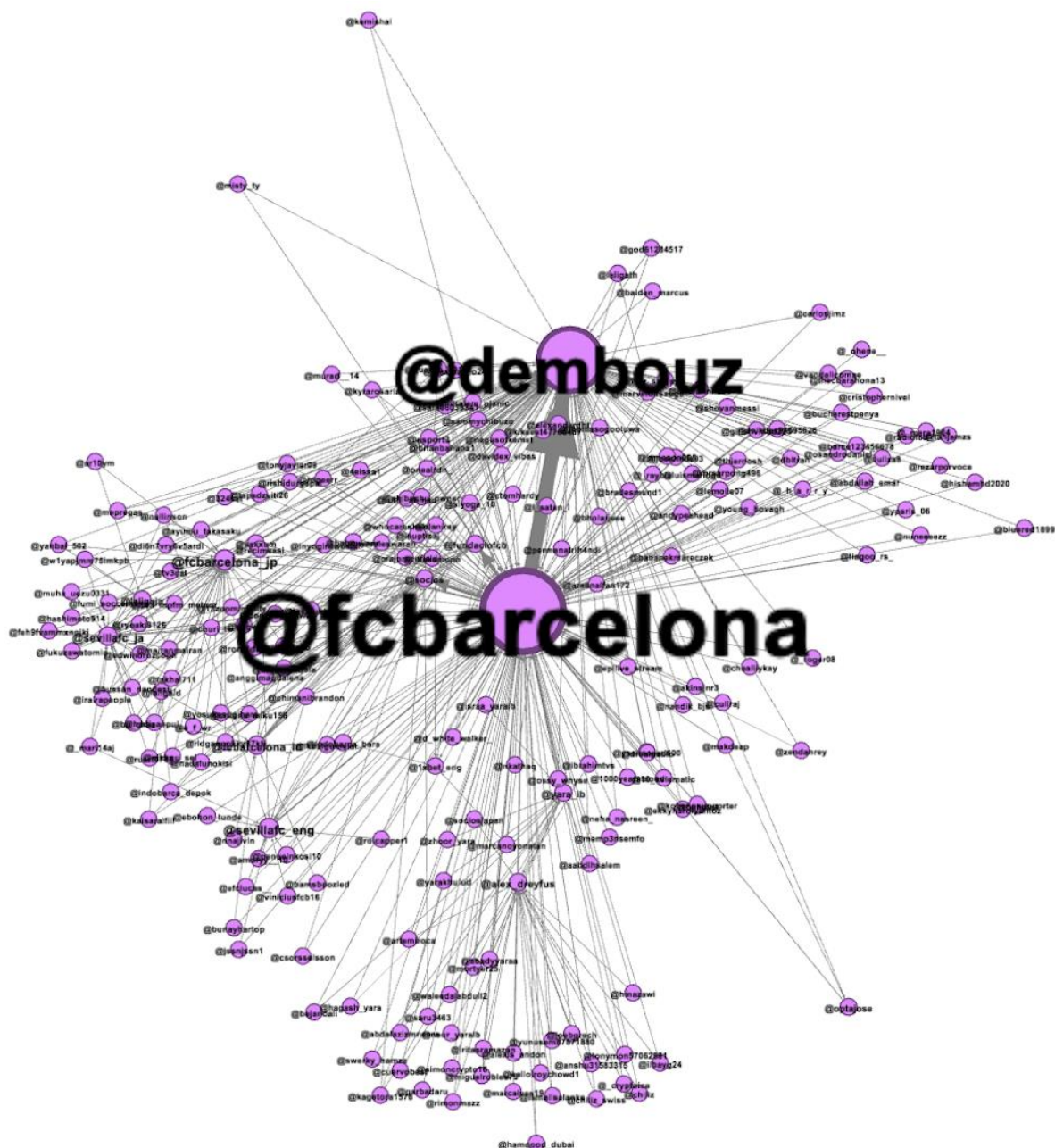
También se añaden las etiquetas, a las que se les establece un ranking por grado entrante para que destaquen aquellas que tienen un mayor resultado en esta estadística, y se hace lo propio con el tamaño de los nodos.

Llegados a este punto, se puede hacer uso de los algoritmos que se encuentran a nuestra disposición para mejorar la visualización y distribución del grafo. En primer lugar, se hace uso del algoritmo OpenOrd para hacer una primera separación que agrupe los nodos por comunidad. Posteriormente, se aplica el algoritmo ForceAtlas2, cuya idea es establecer como se comportan las fuerzas de atracción y resolución entre nodos, y que ayuda a conseguir una distribución de las comunidades en forma circular. También se hace uso de Noverlap para evitar que haya superposición de nodos y del algoritmo Expansión para generar un poco de separación entre los nodos, y que el grafo sea más visual. Finalmente, para realizar el análisis se decide filtrar las comunidades que representan un porcentaje superior al 5% porque se consideran más relevantes, obteniendo el siguiente grafo:



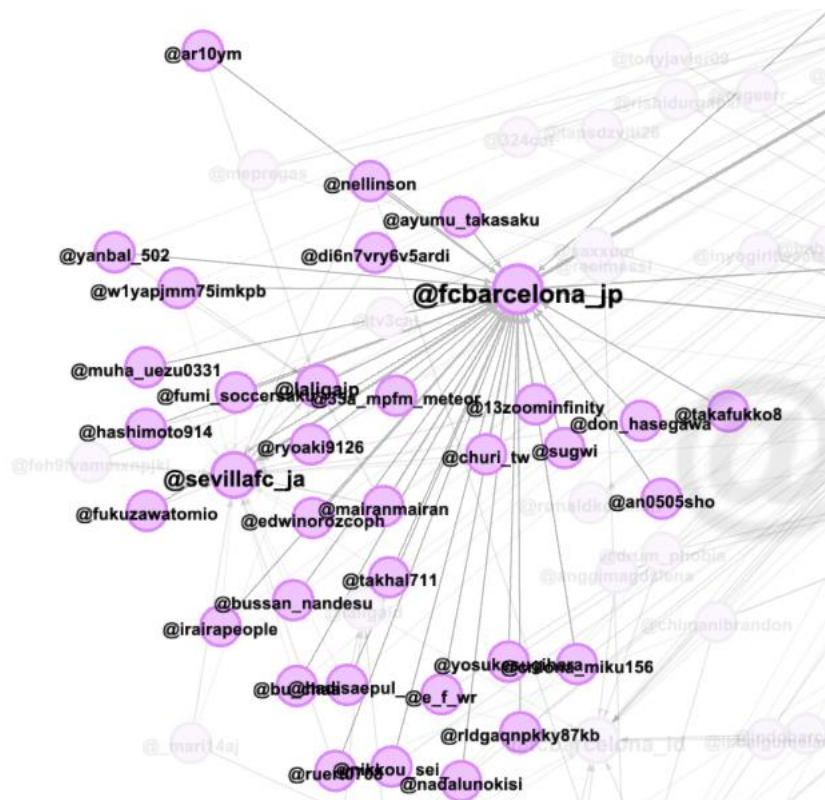
En esta primera imagen se aprecian las 5 comunidades principales, así como las diferencias en cuanto al porcentaje de usuarios que se representan en cada una, y también la posibilidad de que existan sub-comunidades dentro de las mismas. Se procede con el análisis de las comunidades. Para ello, se separan para obtener una mejor distribución y se filtran de forma individual.

En primer lugar, se va a analizar la comunidad de color rosa, que es la que contiene un número mayor de usuarios, y se va a aplicar el filtro k-core con un valor de 4, lo que devuelve únicamente los nodos que están conectados a otros 4, simplificando el gráfico y facilitando el análisis:



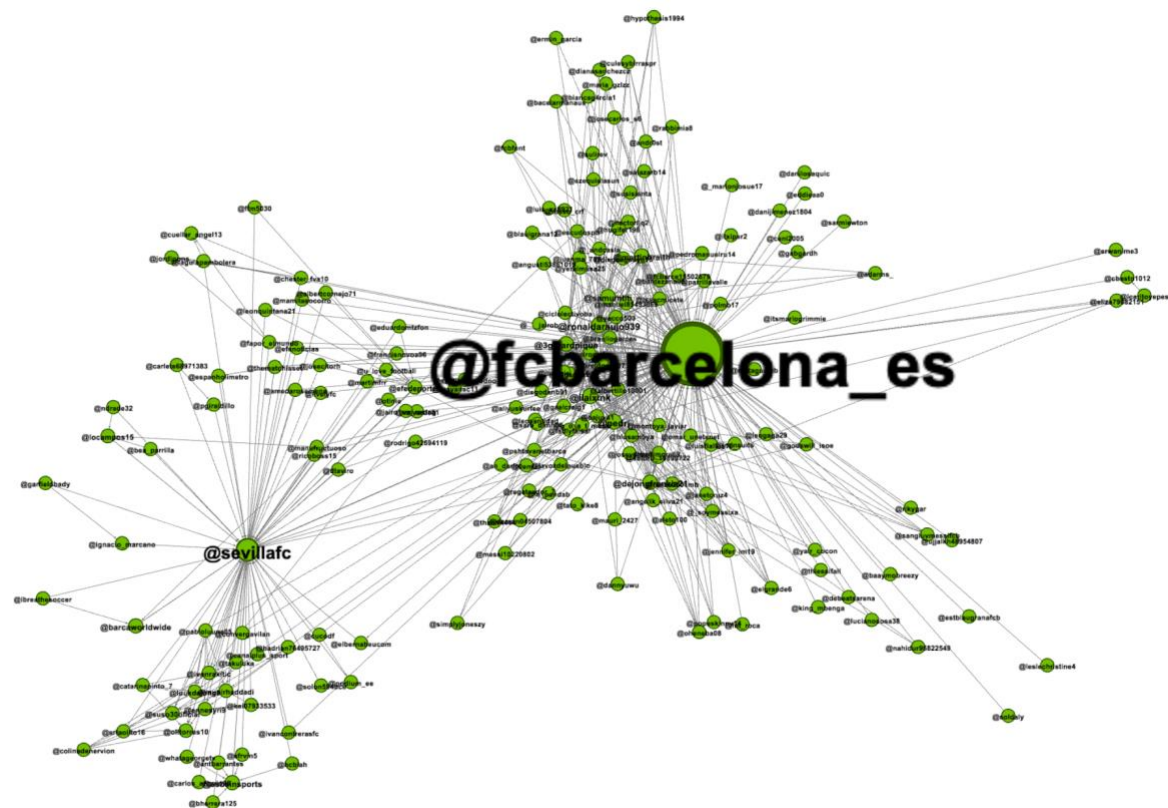
Se aprecian dos de los tres nodos que destacan por encima del resto debido al ranking por grado entrante establecido. Estos son los correspondientes a la cuenta internacional en inglés

del FC Barcelona y a la cuenta del jugador Ousmane Dembelé (@dembouz), que como se comentó en el contexto de este documento, fue el autor del primer gol, por lo que su relevancia en este grafo se debe a la gran cantidad de menciones que recibió dicho usuario, obviamente derivados del gol que marcó. Respecto a la cuenta del jugador, también es destacable que no realiza ninguna interacción desde hace más de 2 años, pero aún así recibió una gran cantidad de menciones.



En la imagen anterior se pueden apreciar las cuentas de twitter de ambos clubes para la comunidad japonesa, así como la de la liga, siendo necesario remarcar que el mercado asiático y, en especial el japonés, es uno de los que más está intentando explotar la liga española para promocionar su producto. En esta comunidad también se encuentra la cuenta de twitter internaciones del Sevilla FC (@SevillaFC_ENG), por lo que parece claro que es una comunidad que aglutina usuarios internacionales que interactuaron con las cuentas internacionales del FC Barcelona y del Sevilla FC y también con el autor del primer gol, así como una sub-comunidad japonesa.

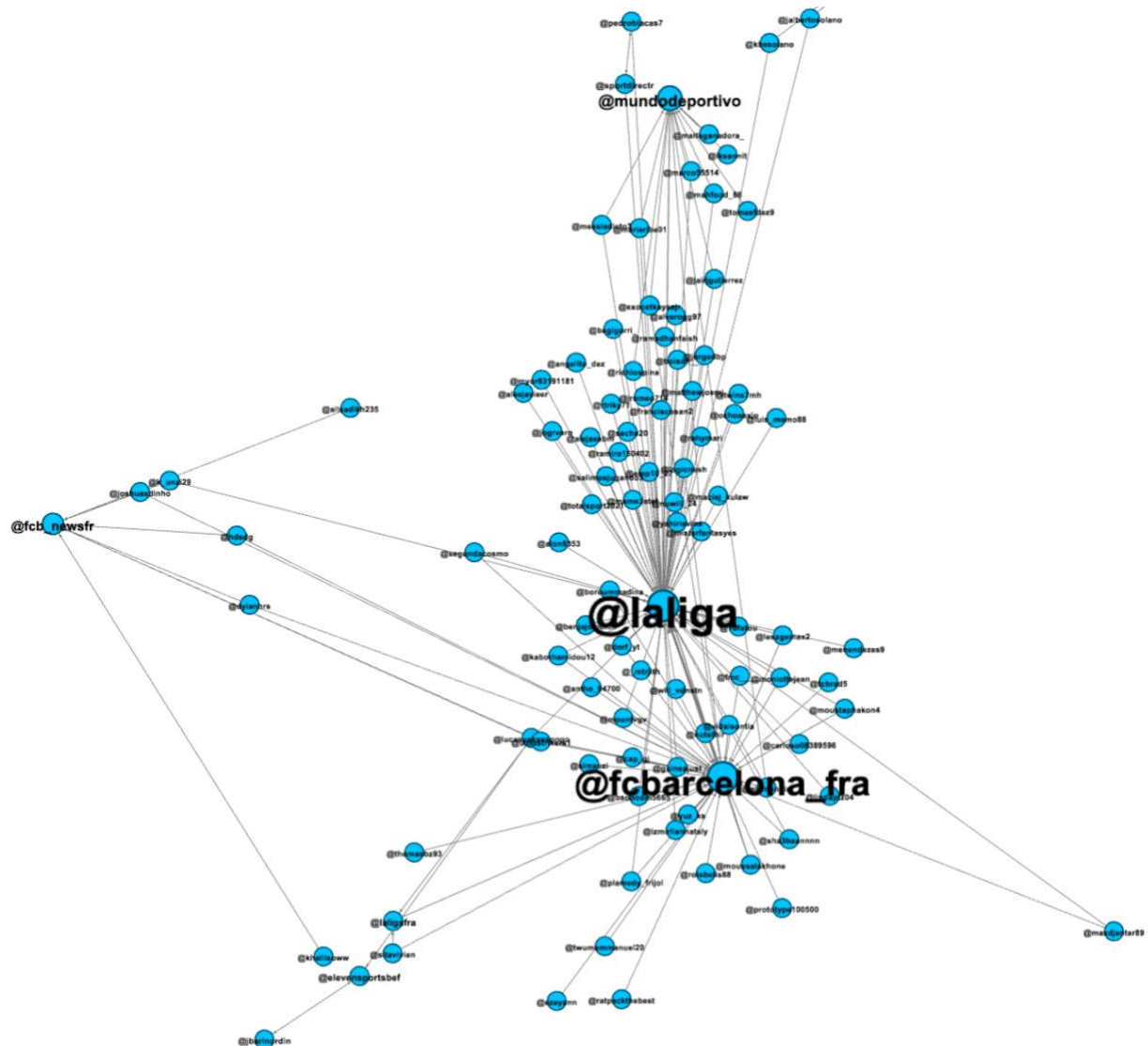
Se procede al análisis de la segunda comunidad, que representa un porcentaje alto del grafo, muy cercano al de la primera, volviéndose a aplicar el mismo filtro k-core:



En esta segunda comunidad destaca principalmente la cuenta del FC Barcelona en español, aunque en este caso la cuenta del Sevilla FC también parece relevante, a diferencia de la comunidad anterior, pero su importancia en el grafo sigue siendo menor que las del FC Barcelona. Además, ambos nodos se encuentran separados y parece que forman dos subcomunidades dentro de la misma, algo que no sucedía anteriormente.

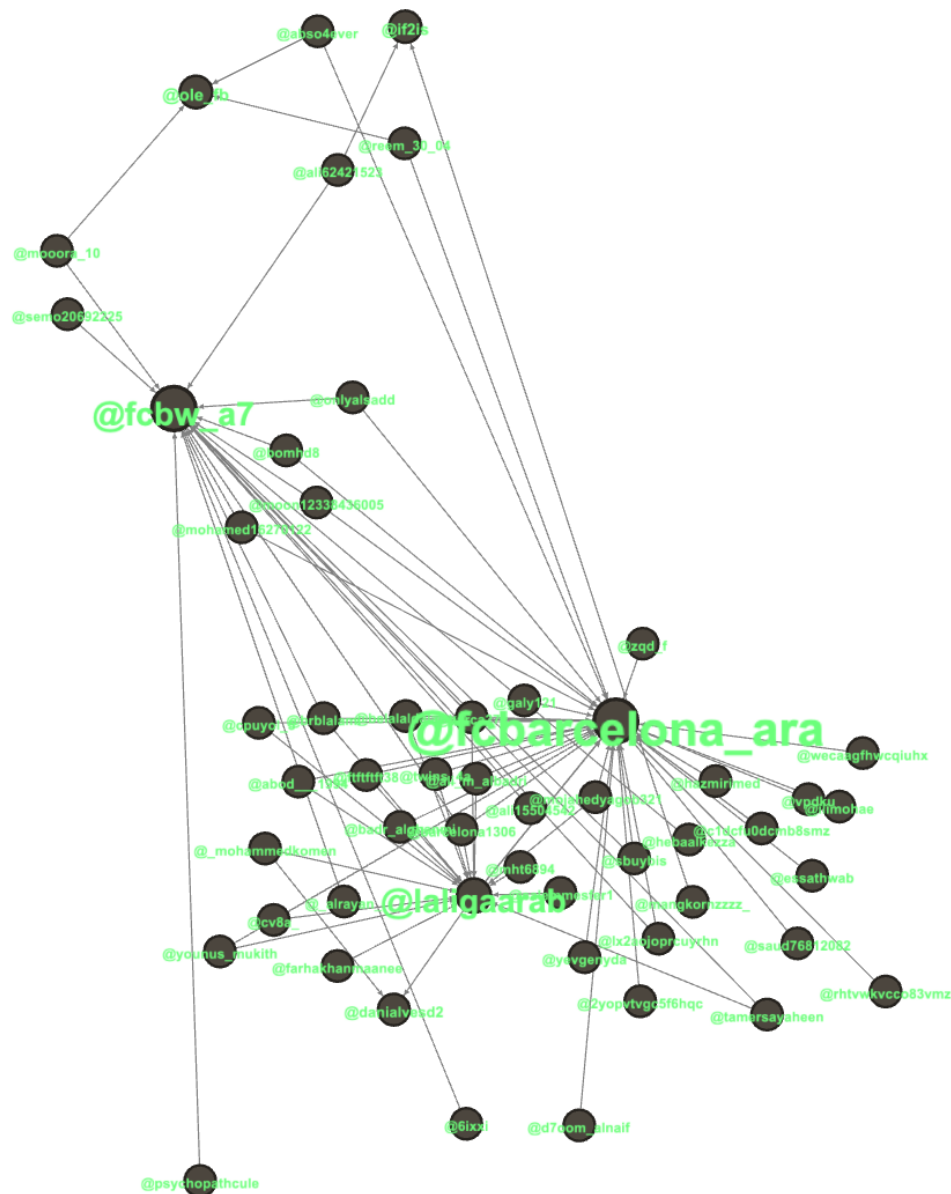
Otro aspecto para destacar es nuevamente la inclusión de jugadores de fútbol. El único que destaca sobre los demás es Pedri (@pedri), lo cual se puede deber a la lesión que sufrió, aunque también se incluyen otros jugadores como Gerard Piqué, Ronald Araujo o Frenkie De Jong, todos ellos debidos a las menciones que se sucedieron hacia sus perfiles durante el transcurso del partido.

La tercera comunidad para analizar representa ya un porcentaje mucho menor que las otras dos, ya que en este caso es solo un 8.3%:



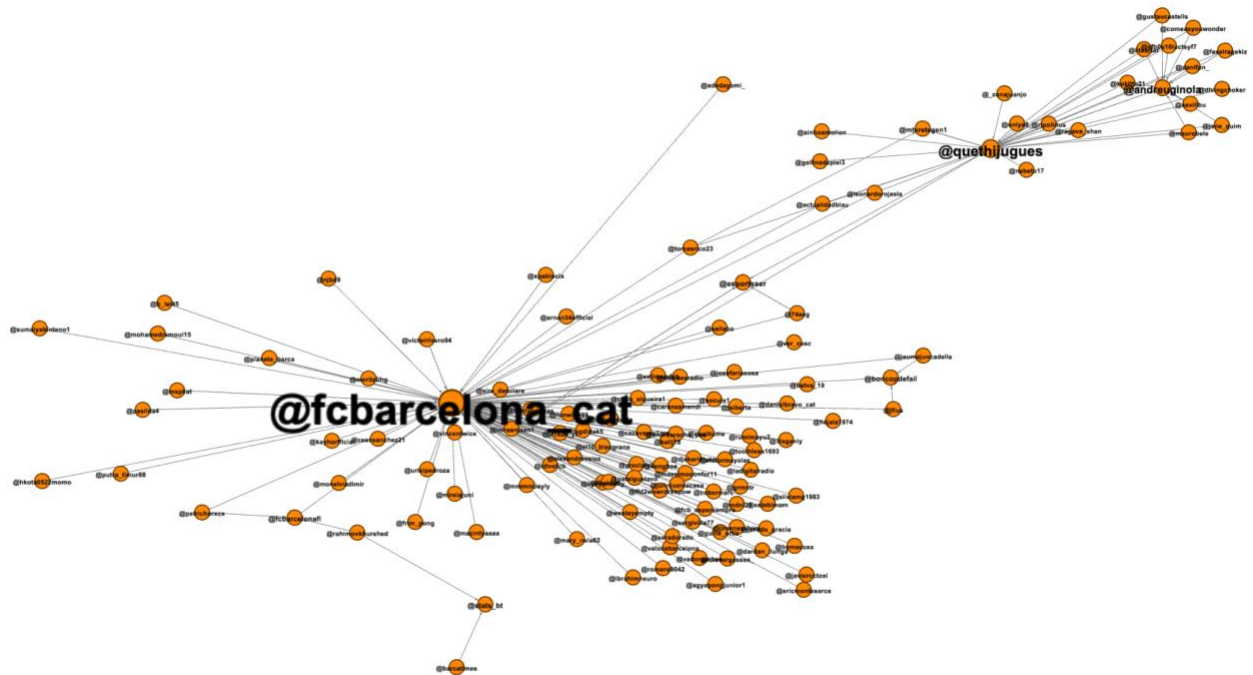
Nuevamente se aplica un k-core igual a 4, y ya se aprecia la menor cantidad de nodos. En esta comunidad destacan 3 posibles sub-comunidades. La primera de ellas en la parte inferior parece una comunidad francesa en torno a la cuenta del FC Barcelona en francés, donde también aparece la cuenta de la liga en francés, pero no la del Sevilla, y esto se debe a que el filtro del k-core la ha eliminado ya que dicho usuario si que se encuentra en el grafo, pero su grado es menor a 4. Existe una segunda sub-comunidad en el centro de la imagen en torno a la cuenta de la liga, y finalmente se puede apreciar la primera comunidad en torno a un medio de comunicación en la parte superior, en este caso es mundo deportivo, periódico que se centra en la actualidad futbolística del FC Barcelona.

La cuarta comunidad a analizar apenas representa un 5.5% del total del grafo, y en este caso hay que destacar que el k-core se ha reducido a 3 debido a la menor cantidad de nodos, y que se ha modificado el color de las etiquetas para una mejor visualización.



Esta comunidad parece que aglutina los usuarios árabes, o que interactuaron con las cuentas destinadas al mundo árabe del FC Barcelona (@fcbarcelona_ara) y de la liga (@laligaarab). También destaca otro nodo en la parte superior izquierda, @fcbw_a7, que parece ser una cuenta que cubre la información del FC Barcelona. Nuevamente destaca que para una comunidad extranjera aparece la cuenta del FC Barcelona, pero no la del Sevilla FC. En el caso de la comunidad francesa se debía al bajo grado de la cuenta del Sevilla en francés, pero en este caso, a pesar de que el Sevilla FC tiene una cuenta en árabe, esta no aparece en el grafo, por lo que no fue mencionada en ningún tweet que contuviera el hashtag. También es necesario añadir que es una cuenta que no interactúa desde hace meses. Otro detalle que es importante mencionar es nuevamente el interés de la liga en explotar un mercado extranjero, en este caso el árabe.

Finalmente, la última comunidad a analizar es de dimensiones similares a la anterior por lo que también se aplica un k-core igual a 3:



Este último grupo a analizar parece que aglutina a la comunidad catalana. Destaca en una posición centrada de la imagen la cuenta del FC Barcelona en catalán, por lo que se asume que las interacciones en torno a esta cuenta serán en catalán, y en la parte derecha de la imagen destaca el usuario @quetehijugues, programa de Ser Catalunya, que cubre la actualidad deportiva en Cataluña, centrada en el FC Barcelona y en catalán, lo que deja claro que esta es la comunidad catalana en este grafo.

Una vez analizadas más a fondo las comunidades que se consideraron inicialmente principales debido al porcentaje de nodos que incluían, también puede ser interesante analizar por encima comunidades más pequeñas para ver si los nodos principales pueden ser relevante para las conclusiones que se puedan obtener de este análisis.

- Las dos siguientes comunidades más relevantes representan un 4.37% y un 4.09% respectivamente, y la primera destaca la cuenta de la liga en inglés, mientras que la segunda incluye las comunidades brasileña y turca del FC Barcelona.
- Las tres siguientes comunidades representan un 4.07%, 2.77% y 2.74% respectivamente. En la primera destaca la presencia de beinsportusa, cadena con los derechos del partido en EEUU, la segunda incluye varios medios de comunicación como los diarios deportivos marca y as, y el programa deportivo de la cadena cope, Tiempo de Juego. Finalmente, la tercera comunidad de este grupo incluye el programa deportivo de la cadena ser, Carrusel Deportivo, y una sub-comunidad en torno a una periodista deportiva que cubre el FC Barcelona, Carme Barceló.

Además de hacer uso de la visualización de los grafos para el análisis, también se pueden obtener conclusiones interesantes a partir del laboratorio de datos.

Data Table											
Nodes Edges Configuration Add node Add edge Search/Replace Import Spreadsheet Export table More actions Filter: Id											
Id	Label	twitter_type	friends_count	followers_co...	real_name	In-Degree	Weighted In-Degree	Closeness Centrality	Betweenness Centrality	Modularity Class	
@fcb Barcelona	@fcb Barcelona	User	93	35675368	FC Barcelona	2046	3621.0	0.695652	26387.146398	25	
@fcb Barcelona...	@fcb Barcelona...	User	84	16342269	FC Barcelona	1802	3259.0	0.430769	28389.751587	0	
@fcb Barcelona...	@fcb Barcelona...	User	84	5991188	FC Barcelona	361	665.0	0.432836	4643.387302	8	
@laliga	@laliga	User	965	5738279	LaLiga	328	534.0	0.518519	25180.892063	17	
@marca	@marca	User	1654	5445868	MARCA	147	179.0	1.0	435.0	10	
@yara_lb	@yara_lb	User	167	4117530	Yara	28	38.0	0.425	0.0	25	
@fcb Barcelona...	@fcb Barcelona...	User	66	3262007	نادي برشلونة	200	280.0	1.0	749.543651	2	
@2010miste...	@2010miste...	User	649	3177133	MisterChip (...)	1	1.0	0.0	0.0	57	
@diarios	@diarios	User	778	2755135	Diario AS	21	23.0	0.0	0.0	10	
@mundodep...	@mundodep...	User	1046	2683884	Mundo Depo...	150	152.0	0.442857	2057.150794	17	
@usmnt	@usmnt	User	266	2138311	U.S. Soccer ...	0	0.0	1.0	0.0	26	
@goal	@goal	User	320	2105333	Goal	66	66.0	0.0	0.0	22	
@efenoticias	@efenoticias	User	64	1888682	EFE Noticias	0	0.0	0.379747	0.0	0	
@fcb Barcelona...	@fcb Barcelona...	User	67	1874368	FC Barcelona	327	558.0	0.464789	4277.807143	17	
@sport	@sport	User	118	1737588	Diario SPORT	102	144.0	1.0	57.0	7	
@footmercato	@footmercato	User	445	1314955	Foot Mercato	2	2.0	0.0	0.0	17	
@la_ser	@la_ser	User	927	1297307	Cadena SER	0	0.0	0.443038	0.0	37	
@pedrerol	@pedrerol	User	1690	1226337	Josep Pedrerol	2	2.0	0.0	0.0	41	
@fcb Barcelona...	@fcb Barcelona...	User	67	1212715	FC Barcelona	260	337.0	1.0	1193.038889	9	
@as_tomasro...	@as_tomasro...	User	2000	1184071	Tomás Ronc...	0	0.0	0.324074	0.0	37	
@leomessifa...	@leomessifa...	User	305	1091382	Leo Messi Fa...	0	0.0	0.308511	0.0	0	
@laligaarab	@laligaarab	User	108	1044710	LaLiga	45	57.0	0.484375	504.9	2	
@sevillaafc	@sevillaafc	User	106	981990	Sevilla Fútbol...	339	661.0	0.4	15808.216667	0	
@laligaen	@laligaen	User	240	807066	LaLiga English	351	468.0	0.571429	12173.664652	22	
@fcbw_a7	@fcbw_a7	User	0	708398	FCB World	206	286.0	0.0	0.0	2	
@324cat	@324cat	User	334	660946	324.cat	0	0.0	0.441176	0.0	25	
@tv3cat	@tv3cat	User	428	602342	TV3.cat	0	0.0	0.441176	0.0	25	
@as_futbol	@as_futbol	User	184	571733	AS Fútbol	0	0.0	1.0	0.0	10	

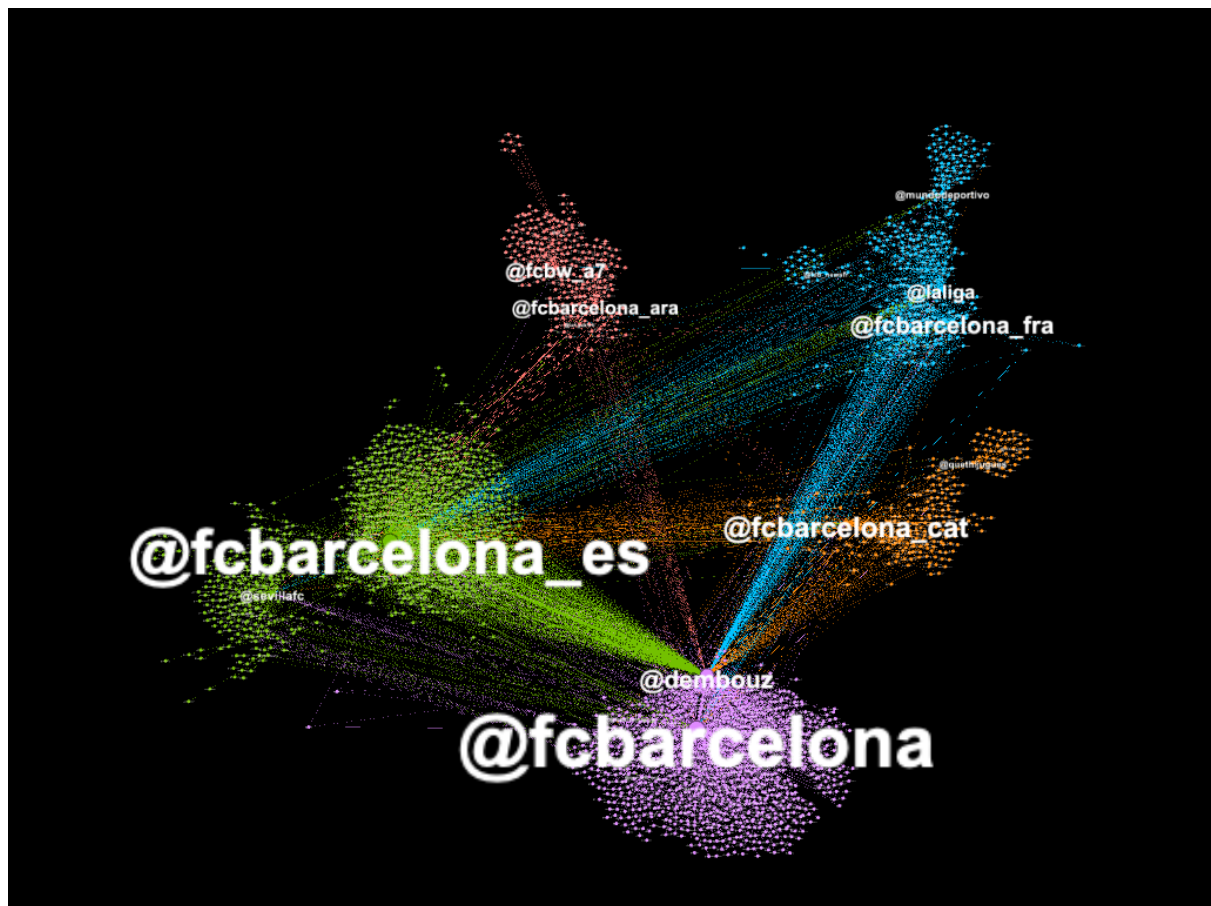
En la imagen anterior se encuentran los usuarios ordenados por el número de followers y se demuestra como muchas de estas cuentas, la mayoría relacionadas con el mundo periodístico, a pesar de tener muchos followers no tienen mucha relevancia en el grafo, ya que no fueron muy mencionadas.

En cambio, si ordenamos por la centralidad de intermediación, se observa que aparecen muchos de los nodos que se han mencionado:

Data Table											
Nodes Edges Configuration Add node Add edge Search/Replace Import Spreadsheet Export table More actions Filter: Id											
Id	Label	twitter_type	friends_count	followers_count	real_name	In-Degree	Out-Degree	Degree	Weighted In-Degree	Closeness Centrality	Betweenne...
@fcb Barcelona_es	@fcb Barcel...	User	84	16342269	FC Barcelona	1802	10	1812	3259.0	0.430769	28389.751587
@fcb Barcelona	@fcb Barcel...	User	93	35675368	FC Barcelona	2046	9	2055	3621.0	0.695652	26387.146398
@laliga	@laliga	User	965	5738279	LaLiga	328	9	337	534.0	0.518519	25180.892063
@sevillaafc	@sevillaafc	User	106	981990	Sevilla Fútb...	339	9	348	661.0	0.4	15808.216667
@laligaen	@laligaen	User	240	807066	LaLiga Eng...	351	4	355	468.0	0.571429	12173.664652
@sevillaafc_eng	@sevillaafc...	User	178	160150	Sevilla FC	99	7	106	134.0	0.615385	7912.197619
@fcb Barcelona_cat	@fcb Barcel...	User	84	5991188	FC Barcelona	361	10	371	665.0	0.432836	4643.387302
@fcb Barcelona_fra	@fcb Barcel...	User	67	1874368	FC Barcelona	327	11	338	558.0	0.464789	4277.807143
@carrusel	@carrusel	User	990	337869	Carrusel D...	96	8	104	155.0	0.465753	2993.345238
@quethijugues	@quethiju...	User	1360	28464	Què T'hi J...	90	4	94	110.0	0.357143	2095.333333
@mundodeportivo	@mundod...	User	1046	2683884	Mundo De...	150	4	154	152.0	0.442857	2057.150794
@fcb Barcelona_br	@fcb Barcel...	User	67	1212715	FC Barcelona	260	9	269	337.0	1.0	1193.038889
@sevillaafc_ja	@sevillaafc_ja	User	7644	10431	セビージャ...	67	11	78	150.0	0.655172	1082.024237
@fundaciofcb	@fundacio...	User	157	84301	Barça Fou...	20	1	21	222.0	0.421053	1003.038462
@tjcope	@tjcope	User	2109	504690	Tiempo d...	48	8	56	71.0	0.473684	828.545238
@fcb Barcelona_ara	@fcb Barcel...	User	66	3262007	نادي برشلونة	200	7	207	280.0	1.0	749.543651
@josemarifc	@josemarifc	User	287	26023	Sevilla FC [...]	85	1	86	103.0	0.425	706.0
@fcb Barcelona_jp	@fcb Barcel...	User	133	310191	FC /バルセ...	116	5	121	200.0	0.463415	635.623443
@laligaarab	@laligaarab	User	108	1044710	LaLiga	45	8	53	57.0	0.484375	504.9
@marvntreyhd1	@marvntrey...	User	234	544	MarvNTre...	52	5	57	52.0	0.459459	457.833333
@esportsser	@esportsser	User	2258	14403	Esports SE...	10	8	18	56.0	1.0	452.0
@marca	@marca	User	1654	5445868	MARCA	147	4	151	179.0	1.0	435.0
@esbeinsports	@esbeinsp...	User	584	261322	beIN SPOR...	37	2	39	40.0	0.426471	331.928571
@fcb Barcelona_id	@fcb Barcel...	User	66	429046	FC Barcelona	65	3	68	83.0	0.333333	329.333333
@bugaluiscf	@bugaluiscf	User	333	606	BugaLuis	21	4	25	31.0	0.326316	228.666667
@laligabrar	@laligabrar	User	100	218135	LaLiga	49	2	51	58.0	0.5625	167.816667
@laligatvbar	@laligatvbar	User	371	158815	LaLigaTV ...	6	5	11	10.0	0.392405	167.769048
@defensalibero	@defensali...	User	14	452	DEFENS...	30	2	32	35.0	0.6	138.0

Esto se puede deber a que esos nodos actúan como centros de sus comunidades conectando a muchos nodos a través de ellos, por lo que se encuentran en muchos caminos críticos.

Finalmente se muestra en la siguiente imagen las 5 comunidades principales separadas:



5. Conclusiones

Las principales conclusiones que se han obtenido en base a los objetivos marcados inicialmente son:

- Hay dos comunidades que destacan sobre las demás en cuanto al tamaño de estas, pero prácticamente todas las comunidades principales tienen un nodo central relacionado con el FC Barcelona. No se aprecian comunidades dedicadas al Sevilla FC con la excepción de una sub-comunidad que se forma en torno a la cuenta del Sevilla FC en español, y que aún así, está claramente eclipsada en la misma comunidad por la cuenta del FC Barcelona en español. Por lo tanto, en muchos casos parece que las comunidades se forman por las interacciones de los usuarios con las diferentes cuentas del FC Barcelona según el idioma, es decir, parecen comunidades por localización geográfica. También es destacable la inmensa diferencia que se produce entre la importancia de las cuentas del Sevilla FC y las del Barcelona FC, que destaca sobremanera con las cuentas internacionales.
- Un aspecto para destacar, que no se había tenido en cuenta en los objetivos iniciales, es la presencia de jugadores, no por la importancia de algún tweet que pudieran haber publicado sino por la cantidad de menciones que recibieron. Se observa que la cuenta @dembouz tiene una gran influencia en el grafo ya que debido a su gol fue mencionado por multitud de usuarios, lo que se puede apreciar en la imagen de la página anterior, donde no solo está relacionada con su comunidad, sino que hay aristas hacia todas las comunidades. El resto de jugadores, es lógico que aparezcan con una menor relevancia, pero sorprende el caso de Lionel Messi, capitán del FC Barcelona, jugador más conocido y relevante de su equipo, y que también anotó un gol durante el partido. El motivo de que Messi no adquiriera importancia en este grafo es bastante simple, no tiene cuenta de twitter.
- Los medios de comunicación tienen una importancia ínfima en comparación con los clubes de fútbol o los jugadores relevantes del partido. Se aprecia una sub-comunidad en torno a un periódico de fútbol catalán, que es mundo deportivo. pero el resto de sub-comunidades se encuentran en grupos muy reducidos, de dimensiones similares, y con poca interacción en comparación con muchas de las cuentas internacionales del FC Barcelona.
- Finalmente, destaca la ausencia de casas de apuestas en este grafo. Se ha comprobado en el laboratorio de datos si las casas de apuestas más importantes a nivel nacional estaban presentes en los datos como, por ejemplo, Sportium, Codere, Bet365, Bwin, WilliamHill, etc. Solo se ha encontrado Sportium, y con un grado muy bajo, por lo que su relevancia es insignificante. Haciendo una búsqueda en twitter de dichas casas de apuestas y sus tweets a la hora del partido, se comprueba que Sportium es la única que estuvo twitteando los acontecimientos que se sucedieron durante el transcurso del mismo. Esto hace pensar que ningún usuario twitteó sobre dichas cuentas incluyendo el hashtag del partido.