



*ugr* | Universidad  
de **Granada**

## Grado en Ingeniería Informática. Cuarto.

### Práctica 4: Caso Práctico Análisis y Evaluación de Redes.

---

**Nombre de la asignatura:**

Redes y Sistemas Complejos. Lunes de 10:30 a 12:30.

**Realizado por:**

Néstor Rodríguez Vico. DNI: 75573052C.

email: [nrv23@correo.ugr.es](mailto:nrv23@correo.ugr.es)



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍAS  
INFORMÁTICA Y DE TELECOMUNICACIÓN.

---

Granada, 15 de diciembre de 2017.

# Índice

<b>1 Selección de un dominio, definición de una pregunta de investigación y obtención de un conjunto de datos estructural inicial asociado.</b>	<b>3</b>
<b>2 Construcción de la red compleja a analizar y visualizar.</b>	<b>3</b>
<b>3 Cálculo de los valores de las medidas de análisis.</b>	<b>3</b>
<b>4 Determinación de las propiedades de la red.</b>	<b>4</b>
<b>5 Calculo de los valores de las medidas de análisis de redes sociales.</b>	<b>7</b>
5.1 Grado. . . . .	7
5.2 Intermediación (betweeness). . . . .	8
5.3 Cercanía (closeness). . . . .	9
5.4 Vector propio. . . . .	10
<b>6 Descubrimiento de comunidades en la red.</b>	<b>11</b>
6.1 Método de Lovaina. . . . .	11
<b>7 Visualización de la red compleja.</b>	<b>12</b>
<b>8 Conclusión.</b>	<b>14</b>
<b>9 Bibliografía.</b>	<b>14</b>

# **1. Selección de un dominio, definición de una pregunta de investigación y obtención de un conjunto de datos estructural inicial asociado.**

El dominio con el que voy a trabajar es *Twitter*. Voy a tratar de ver que personas son las más influyentes en el debate que hay acerca de las elecciones al Gobierno de Cataluña del día 21 de Diciembre. Para ello, voy a descargar 3000 *tuits* (mezclando entre los más destacados y los más recientes) del *hashtag* #21D mediante la API de *Twitter* y usando la biblioteca *tweepy* para *Python*.

# **2. Construcción de la red compleja a analizar y visualizar.**

Una vez tengo los *tuits*, vamos a extraer todos los usuarios que ha publicado esos *tuits* y vamos a construir una red en la que los usuarios van a ser los nodos y va a haber un arco entre el nodo  $i$  y el nodo  $j$  si el usuario  $i$  sigue al usuario  $j$ . Debemos tener en cuenta que la red que vamos a obtener es una red dirigida, ya que puede ser que el usuario  $i$  siga al usuario  $j$  pero no al revés. Todo este proceso se ha realizado en *Python* para finalmente guardarla en un fichero *.graphml*, ya que si usaba *.net Gephi* no era capaz de leer los atributos que contenía el nodo (como puede ser el nombre de usuario para etiquetar el propio nodo).

# **3. Cálculo de los valores de las medidas de análisis.**

A continuación vamos a calcular ciertos valores de la red:

Número de nodos: N	2256
Número de enlaces: L	54403
Densidad: D	0.011
Grado medio: $\langle k \rangle$	24.115
Diametro: $d_{max}$	7
Distancia media: $\langle d \rangle$	2.872
Distancia media aleatoria $\langle d_{aleatoria} \rangle$	0.765
Coef. de clustering medio $\langle C \rangle$	0.427
Coef. de clustering medio aleatoria $\langle C_{aleatoria} \rangle$	0.0106
Número de componentes conexas	1
Número nodos componente gigante	2256
Porcentaje con respecto a red total	100 %
Número enlaces componente gigante	54403
Porcentaje con respecto a red total	100 %

Para calcular la distancia media de una red aleatoria equivalente, he aplicado la siguiente fórmula:

$$\langle d_{aleatoria} \rangle = \frac{\log N}{\log \langle k \rangle} = \frac{\log 2256}{\log 24,115} = 0,765$$

Para calcular el coeficiente de clustering medio de una red aleatoria equivalente, he aplicado la siguiente fórmula:

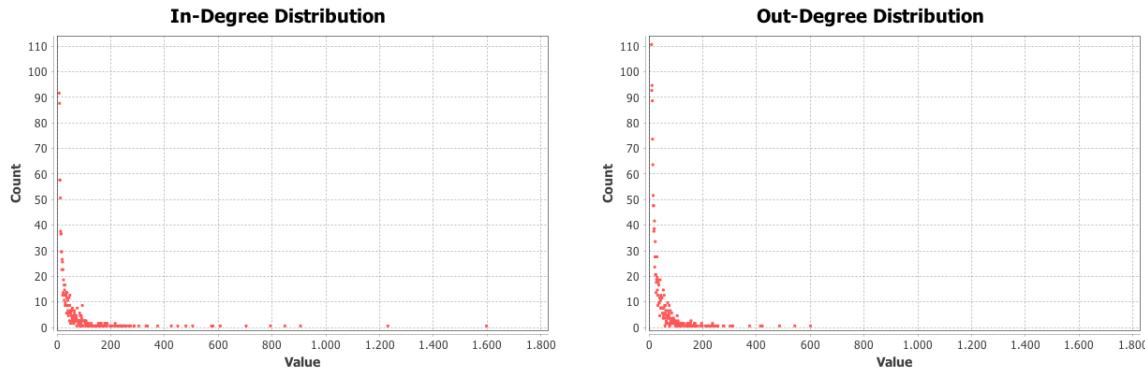
$$C = \frac{\langle k \rangle}{N} = \frac{24,115}{2256} = 0,0106$$

Podemos observar varias cosas interesantes. La primera de ellas es que tenemos una única componente conexa. Esto se debe a la propia idea de la red, y es que esta está construida sobre las propias relaciones de los usuarios. Para que hubiese más de una componente conexa deberíamos tener al menos un usuario de Twitter que no siguiese ni fuese seguido por ningún de los usuarios de nuestra red. Pero esto no puede suceder ya que nosotros construimos la red a partir de esas relaciones.

La siguiente cosa interesante es el alto valor del coeficiente de clustering que tenemos. Esto se debe a la idea que subyace en nuestra red, ya que se trata de una red social y lo normal es que los usuarios que yo sigo en *Twitter* se sigan entre ellos.

## 4. Determinación de las propiedades de la red.

Dado que se trata una red dirigida, vamos a ver las distribuciones de grado tanto de entrada como de salida:



Como podemos ver, las distribuciones de grados de nuestra red tienen la forma de las que hemos visto en clase, es decir, sigue la llamada “Ley de la Potencia”. Podemos ver la existencia de hubs, sobre todo si nos fijamos en la distribución del grado de entrada. Dada la idea que hay en nuestra red, podemos ver el grado de entrada como el número de seguidores (nodos que te apuntan) y el grado de salida como gente a la que sigues. El nodo con un mayor grado de salida es *Esquerra ERC* con un valor de 597 y el nodo

con una mayor grado de entrada es *KRLS* con un valor de 1594.<sup>1</sup>

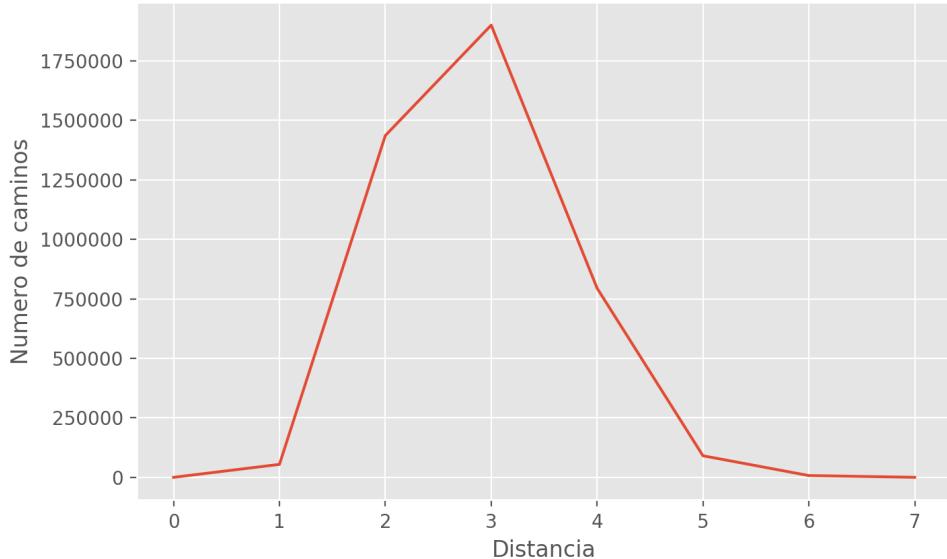
Viendo la distribución de grados que tiene, podemos decir que se trata de una red libre de escalas. Por lo tanto, vamos a calcular el exponente de grado  $\gamma$ . Para ello voy a aplicar la propia definición:

$$p_k \sim k^{-\gamma} \rightarrow \log p_k \sim -\gamma \log k \rightarrow \gamma \sim -\frac{\log p_k}{\log k}$$

Lo que vamos a hacer es contar el número de nodos que tienen grado  $k$  para tener  $p_k$  y  $k$ . Esto lo hacemos para cada nodo y sacamos la media. Como mi red es dirigida, he obtenido un valor de entrada y un valor de salida:

$$\gamma_{\text{entrada}} = 1,530899633110192 \quad \gamma_{\text{salida}} = 1,4934217312420166$$

Pasemos a ver ahora la gráfica asociada a la distancia. Con *Gephi* no se obtenía un resultado demasiado representativo, así que he usado la biblioteca *graph-tool* y *ggplot* ambos para *Python* para pintar una gráfica más representativa:

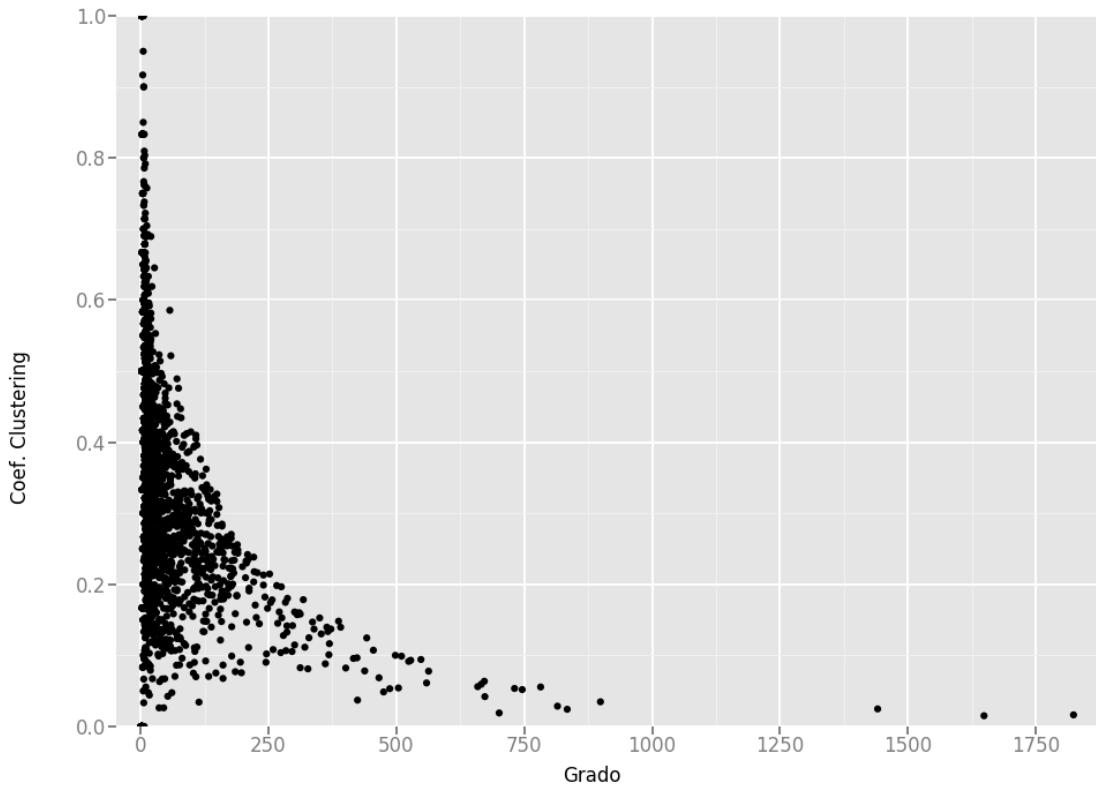


Como podemos ver, el mayor número de caminos lo tenemos para una distancia 3, lo cual es bastante lógico si tenemos en cuenta que la gráfica es más o menos simétrica y la distancia media es 2,872. Veamos también la gráfica del coeficiente de clustering. En este caso también he hecho uso de *graph-tool* y *ggplot* para representar el coeficiente de clustering frente al grado total (suma del grado de entrada más el grado de salida)<sup>2</sup>:

---

<sup>1</sup>No son valores de seguidores reales, son el número de seguidores dentro de nuestra red.

<sup>2</sup>La idea de representar estas dos medidas juntas ha sido sacada del paper (figura 3) de *Renato Fabbri* nombrado en la bibliografía.



Como podemos ver, hay muchos nodos con un alto coeficiente de clustering y con un grado no demasiado alto (en torno al 5 por ciento del grado máximos: al rededor de grado 100 siendo el máximo de 1824). Esto es bastante normal, ya que en las redes sociales nuestros amigos suelen ser amigos entre ellos (en nuestro caso seguidores entre ellos).

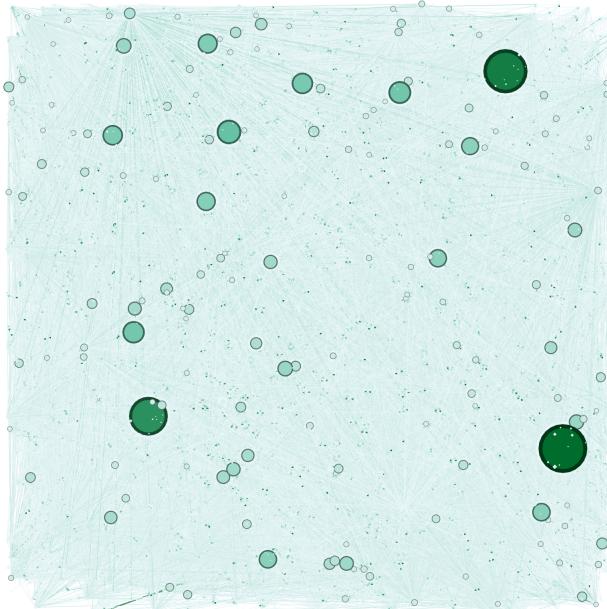
Podemos ver que se trata de una red de mundo pequeño observando la gráfica de distancia. Si nos fijamos, es bastante similar a la que podemos ver en la transparencia 70 del tema 2 de teoría. En las transparencias de teoría se representa la probabilidad de, dado un camino, ver cual será su distancia. En mi gráfica representamos lo mismo pero sin hacer uso de la probabilidad, sino representando el número de caminos como tal. Esta gráfica nos dice que, la mayoría de los caminos tienen una distancia pequeña, cercana a la distancia media. También vemos que el número de caminos que tiene una distancia grande son mínimos. Por lo tanto, efectivamente, estamos en una red de mundo pequeño.

Viendo el coeficiente de clustering de nuestra red y comparándolo con el coeficiente de clustering de la red aleatoria, podemos saber que nuestra red no es una red aleatoria.

## 5. Calculo de los valores de las medidas de análisis de redes sociales.

### 5.1. Grado.

Debemos tener cuidado con el grado, ya que como hemos visto en clase de teoría, se trata de una medida bastante local. Veamos aún así su representación:



He representado el grado tanto en el tamaño como en el color del nodo, de forma que, a mayor tamaño, mayor grado y a más oscuro el verde, más grado. Veamos cuales son los 5 nodos con mayor grado:

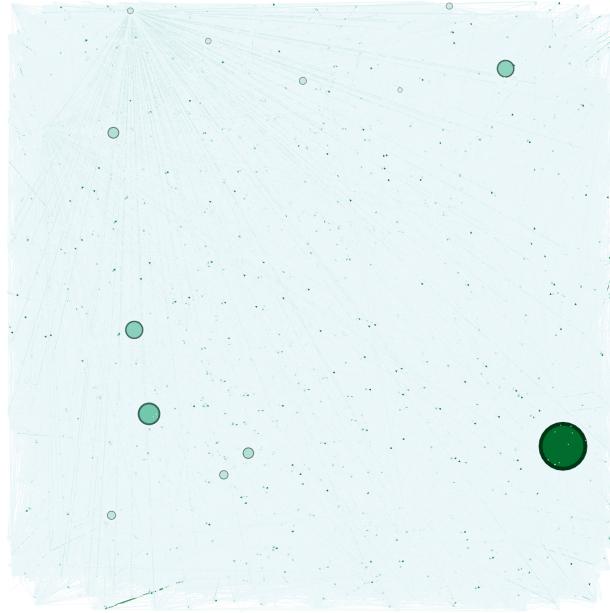
username	Grado	Grado entrada	Grado salida
Esquerra ERC	1824	1227	597
KRLS	1649	1594	55
AlfredBosch	1441	902	44
tonialba	899	333	482
CNICatalunya	834	371	411

Como podemos ver, sucede lo esperable. Tenemos como personas más grandes (entendiendo por grandes el número de seguidores) a Carles Puigdemont (*KRLS*) y a Esquerra Republicana (*Esquerra ERC*). Podemos ver también que hay nodos que tienen un grado alto pero no por tener muchos seguidores, si no por seguir a mucha gente, como sucede con el usuario *tonialba*.

El nodo más vistoso es el que se encuentra abajo a la derecha, el cual es Esquerra Republicana (*Esquerra-ERC*). A continuación, arriba a la derecha, el cual es Carles Puigdemont (*KRLS*) y finalmente abajo a la izquierda, el cual es AlfredBosch (*Alfred-Bosch*).

## 5.2. Intermediación (betweenness).

Al igual que antes, he representado la intermediación tanto en el tamaño como en el color del nodo, de forma que, a mayor tamaño, mayor intermediación y a más oscuro el verde, más intermediación:



Veamos cuales son los 5 nodos con mayor intermediación (ya normalizado):

username	Intermediación
Esquerra ERC	0.220874
AlfredBosch	0.0999
_CarmenLopez	0.079955
KRLS	0.07753
xavier_torrens	0.048703

Como hemos visto en teoría, esta medida capta la correduría de la información por la estructura de la red. Obtiene un mayor valor para los nodos por los que pasen más caminos mínimos por él. Estos nodos son los que se encargan de hacer de puente entre comunidades. En nuestro caso, nos sale que el usuario con un mayor valor de intermediación es Esquerra Republicana (*Esquerra-ERC*), bastante por delante de Carles Puigdemont (*KRLS*), lo cual nos lleva a pensar que (*KRLS*) no desea estar muy relacionado

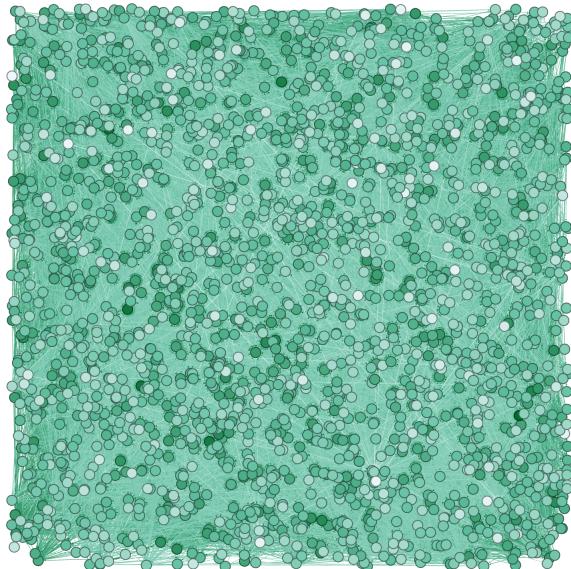
con las demás comunidades en comparación con Esquerra Republicana (*Esquerra-ERC*).

En este caso, el nodo más vistoso es el que se encuentra abajo a la derecha, el cual es Esquerra Republicana (*Esquerra-ERC*). A continuación, abajo a la izquierda, el cual es AlfredBosch (*AlfredBosch*). Carles Puigdemont (*KRLS*) sigue en el mismo sitio para que podamos reconocerlo fácilmente, arriba a la derecha.

### 5.3. Cercanía (closeness).

Como hemos visto en las transparencias de teoría, la cercanía es una forma de medir la centralidad, la cual plantea que el hecho de que puede no ser tan importante tener muchos amigos directos ni estar situado “entre” otros actores. En este caso, se le da importancia a “estar en medio de las cosas”, no demasiado lejos del centro, para lo cual no es necesario estar en una posición de correduría. La suma de las distancias geodésicas (distancias de los caminos mínimos) para cada actor es la lejanía de dicho actor al resto. La inversa de dicha suma es la medida de cercanía.

A diferencia de antes, he representado la cercanía sólo la intensidad del verde, ya que si lo hacía en el tamaño del nodo, muchos quedaban superpuestos. El resultado es el siguiente:



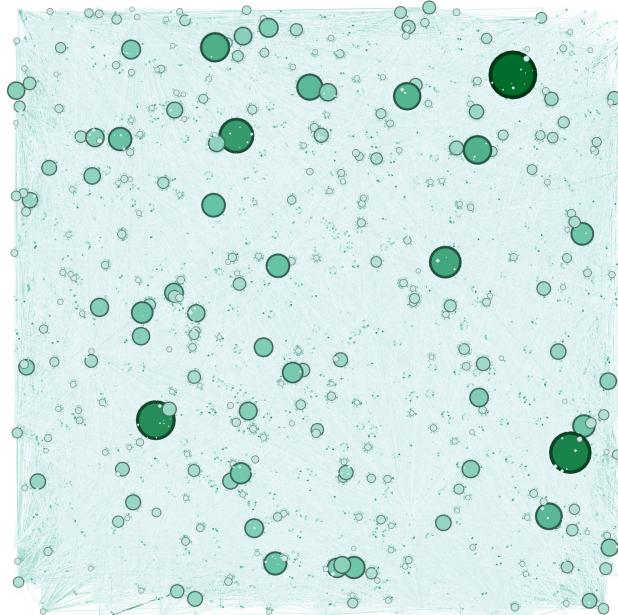
Veamos cuales son los 5 nodos con mayor cercanía (ya normalizada):

username	Cercanía
KRLS	0.747927
Esquerra ERC	0.6523
AlfredBosch	0.578947
tonialba	0.561224
ErnestoEkaizer	0.555419

Podemos ver que Carles Puigdemont (*KRLS*) y Esquerra Republicana (*Esquerra ERC*) siguen siendo los actores más relevantes.

#### 5.4. Vector propio.

La centralidad de vector propio tiene como base la idea de que la centralidad de un nodo depende de cómo de centrales sean sus nodos vecinos. La idea es que el poder y el status de un actor (ego) se define recursivamente a partir del poder y el status de sus vecinos (alters). Es una versión más elaborada de la centralidad de grado al asumir que no todas las conexiones tienen la misma importancia. Es como dice el dicho, “más vale calidad que cantidad”. Veamos una representación de la red en la cual he representado el valor de vector propio tanto en el tamaño como en el color del nodo, de forma que, a mayor tamaño, mayor valor de vector propio y a más oscuro el verde, más valor de vector propio:



Veamos cuales son los 5 nodos con mayor valor de vector propio:

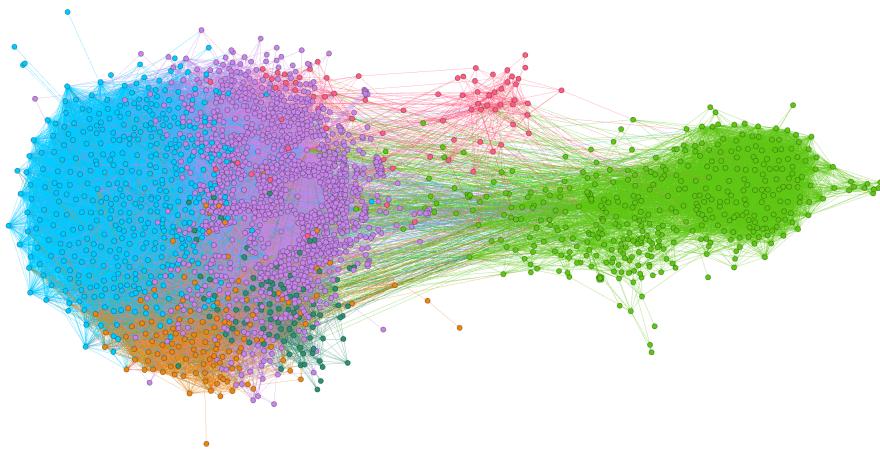
username	Vector propio
KRLS	0.194283
Esquerra_ERC	0.174494
AlfredBosch	0.171341
tonialba	0.153661
lizcastro	0.14591

Una vez más, podemos ver a Carles Puigdemont (*KRLS*) y Esquerra Republicana (*Esquerra\_ERC*) en la cima de la tabla. En este caso con bastante más sentido ya que los grandes actores de la red se conectan entre ellos en la vida real (en este caso en *Twitter*) y de esta forma sube su valor de vector propio.

## 6. Descubrimiento de comunidades en la red.

### 6.1. Método de Lovaina.

Este método es el que implementa Gephi implementado de base. Es un método de clustering jerárquico. Tiene un enfoque greedy que va optimizando la medida de la modularidad. En mi caso, he obtenido un valor de modularidad,  $Q$ , de 0,383. Si visualizamos la red representando la comunidad en el color del nodo, obtenemos lo siguiente:



En mi caso, hemos encontrado 6 comunidades. Veamos el tamaño de ellas:

ID	Color	Porcentaje
0	Morado	50.44 %
1	Naranja	5.36 %
2	Verde Oscuro	4.08
3	Azul	18.31 %
4	Rosa	3.28 %
5	Verde	18.53 %

Para que nos hagamos una idea, Carles Puigdemont (*KRLS*) y Esquerra Republicana (*EsquerraERC*) están en la misma comunidad, la 0. Sin embargo, Albert Rivera (*AlbertRivera*) está en otra comunidad, en la 5. Si pensamos en la vida real, esto es así, sabemos que tal y como está el tema ahora mismo, ambos se encuentran en situaciones distintas.

## 7. Visualización de la red compleja.

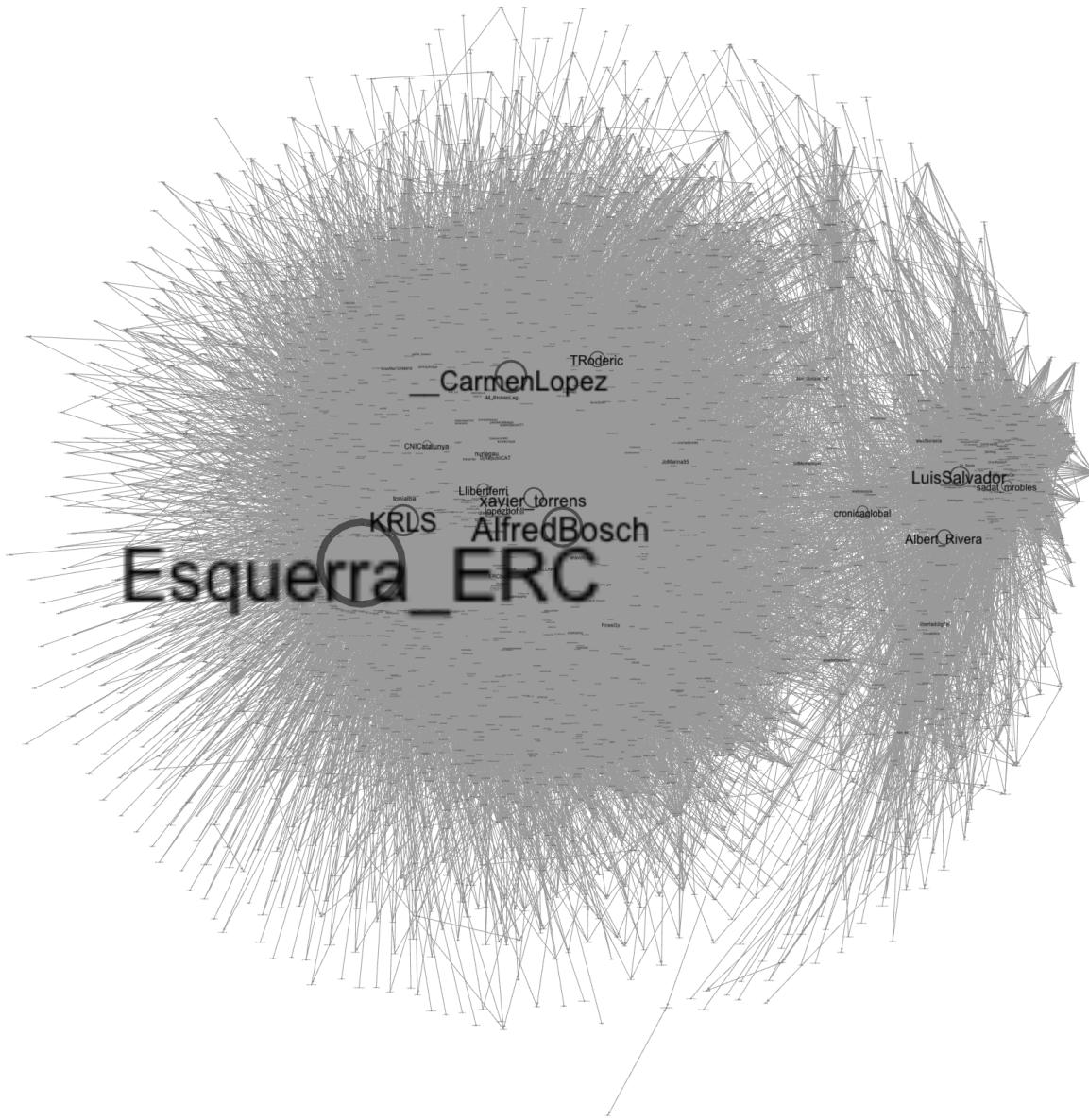
En las visualizaciones que hemos visto a lo largo de la práctica no se ha aplicado ningún algoritmo de visualización, simplemente es la visualización que carga *Gephi* por defecto. En la detención de comunidades se ha usado el algoritmo de visualización *Force Atlas 2* (*Kamada Kawai*). Veamos las mismas comunidades pero con el algoritmo *Fruchterman Reynold*:



La visualización es totalmente distinta de la que ya hemos visto, pero la idea es la

misma. Podemos ver 4 comunidades más juntas y la comunidad de color verde algo más alejada. Para esta representación he usado un valor de  $\text{área} = 5000$ ,  $\text{gravedad} = 10$  y  $\text{velocidad} = 150$ .

Vamos a visualizar ahora, usando *Frutchermand Reynold* con un valor de  $\text{área} = 2000$ ,  $\text{gravedad} = 50$  y  $\text{velocidad} = 150$  nuestra red y luego un algoritmos de *ajuste de etiquetas*, para ver bien las etiquetas. Vamos a ver todos los nodos del mismo color y el tamaño del nodo va a representar el valor de intermediación de los nodos. Veamos el resultado:



Podemos ver cosas bastante curiosas. Visualmente, podemos ver el significado que tiene la medida de intermediación. Como ya hemos comentado anteriormente, esta medida

representa la capacidad de estar en medio del flujo de información. En nuestra red, podemos ver como los actores con un mayor de intermediación (*Esquerra-ERC*, *KRLS*, *CarmenLopez*, *AlfredBosch*) están en el centro de todo. Si miramos más la parte derecha de nuestra red, podemos ver que en esa parte tenemos a *Albert\_Rivera* y a *LuisSalvador* como los actores más relevantes de esa zona. Si nos fijamos, estos dos actores también están en medio de los demás actores.

## 8. Conclusión.

Como podemos ver, a lo largo de este estudio hemos sido capaces de responder a la pregunta que surgió al principio del mismo. Hemos sido capaces de identificar los actores más importantes y relevantes en la discusión sobre las elecciones del 21 de Diciembre en *Twitter*. Como ya hemos comentado a lo largo de este estudio, nos encontramos ante una red libre de escala y de mundo pequeño.

También hemos calculado el exponente de grado de nuestra red, obteniendo un resultado de  $\gamma_{entrada} = 1,530899633110192$  y  $\gamma_{salida} = 1,4934217312420166$ . También hemos visto que, en una red libre de escala se cumple:

$$k_{max} \approx k_{min} N^{\frac{1}{\gamma-1}}$$

En mi caso, ambos  $\gamma$  están por encima de 1 pero no mucho, por lo tanto el exponente de N es mayor que 1 y por lo tanto el grado máximo de nuestra red crece según crece el número de nodos, N. Esto nos hace ver que estamos en una red de mundos ultra pequeños.

## 9. Bibliografía.

- tweepy
- networkx
- graph-tool
- On the evolution of interaction networks: primitive typology of vertex and prominence of measures - Renato Fabbri, Vilson Vieira da Silva Junior, Ricardo Fabbri y Osvaldo N Oliveira
- Gephi.