

# **Gestión de Información en la Web**

## **Master en Ingeniería Informática**

**Práctica 2: Análisis y Evaluación de Redes en Twitter**

**Luis Alberto Segura Delgado**

**DNI: 45922174-Y**

**segura2010@correo.ugr.es**

**Martes 5 de Abril de 2016**

# Índice

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introducción</b>   | <b>3</b>  |
| <b>2</b> | <b>Trabajo Realizado</b>  | <b>3</b>  |
| 2.1      | Descripción del Problema . . . . .  | 3         |
| 2.2      | Cálculo de los valores de las medidas de análisis . . . . .                   | 3         |
| 2.3      | Propiedades de la red . . . . .   | 5         |
| 2.3.1    | Distribuciones de Grados . . . . .  | 5         |
| 2.3.2    | Distribuciones de Distancias . . . . .  | 7         |
| 2.3.3    | Coefficiente de Clustering . . . . .  | 7         |
| 2.4      | Calculo de los valores de las medidas de análisis de redes sociales . . . . . | 9         |
| 2.4.1    | Grado . . . . .   | 10        |
| 2.4.2    | Cercanía . . . . .  | 11        |
| 2.4.3    | Intermediación . . . . .  | 12        |
| 2.4.4    | Vector Propio . . . . .   | 13        |
| 2.4.5    | PageRank . . . . .  | 13        |
| 2.4.6    | Medidas más interesantes . . . . .  | 14        |
| 2.5      | Descubrimiento de comunidades . . . . .                                       | 15        |
| 2.6      | Visualización de la red . . . . .   | 19        |
| <b>3</b> | <b>Discusión de los resultados y Conclusiones</b>                             | <b>20</b> |

# 1 Introducción

El objetivo de esta segunda práctica es formalizar todos los conocimientos adquiridos en el curso aplicándolos a un caso real de análisis de una red social online generada a partir de un medio social. Para ello, se ha seleccionado un medio social concreto (Twitter) y una pregunta de investigación. A partir del medio social elegido, se obtendrá el conjunto de datos y se construirá una red social, que será analizada con objetivo de responder a la pregunta de investigación planteada.

## 2 Trabajo Realizado

En esta sección se detalla el trabajo realizado en la práctica, indicando en primer lugar el problema concreto que se ha planteado y el conjunto de datos y la forma de obtenerlos para resolver dicho problema. A continuación se explicará el análisis realizado sobre los datos y la red social obtenida y finalmente las conclusiones obtenidas del estudio.

### 2.1 Descripción del Problema

El problema a estudiar es detectar cuales son los usuarios más relevantes en la discusión de Twitter sobre la emisión en **Periscope**<sup>1</sup> que tuvo lugar el día 25 de marzo, organizada por Gerard Piqué<sup>2</sup>.

Para abordar el problema, se han recopilado tweets publicados durante la emisión en los que se mencionaba a Piqué (@3gerardpique) y se incluía la palabra "Periscope". Y como la obtención de los datos se realizó unos días después, se han limitado la búsqueda a los tweets que se publicaron el día 25 de Marzo, día de la emisión<sup>3</sup>. Para obtener los tweets, se ha utilizado la herramienta NodeXL.

De cara a evaluar la red correctamente, se ha decidido eliminar el nodo de Piqué de la red, pues todos los tweets lo mencionan, por tanto se conecta con todos los usuarios, y esto dificulta el análisis de la red y su visualización al mismo tiempo que no resulta interesante.

### 2.2 Cálculo de los valores de las medidas de análisis

Para el análisis de la red se ha utilizado la herramienta **Gephi**.

Nuestra red social tiene los siguiente valores para las medidas de análisis:

- **Número de Nodos (N):** 1763
- **Número de Enlaces (L):** 1464
- **Densidad (D):** 0.001
- **Grado Medio ( $\langle k \rangle$ ):** 1.661
- **Diámetro ( $d_{max}$ ):** 2
- **Distancia Media ( $\langle d \rangle$ ):** 1.013
- **Distancia Media para la red aleatoria equivalente ( $\langle d_{aleatoria} \rangle = \frac{\log(N)}{\log(\langle k \rangle)}$ ):** 14.73
- **Coefficiente de Clustering Medio ( $\langle C \rangle$ ):** 0.05

---

<sup>1</sup><https://www.periscope.tv>

<sup>2</sup>[http://as.com/videos/2016/03/25/portada/1458916408\\_738738.html](http://as.com/videos/2016/03/25/portada/1458916408_738738.html)

<sup>3</sup>Búsqueda avanzada de Twitter: @3gerardpique periscope since:2016-03-25 until:2016-03-26 (<https://twitter.com/search?vertical=default&q=%403gerardpique%20periscope%20since%3A2016-03-25%20until%3A2016-03-26&src=typd>)

- **Coefficiente de Clustering Medio para la red aleatoria equivalente** ( $\langle C_{aleatoria} \rangle = \frac{\langle k \rangle}{N}$ ): 0.0009

El número de componentes conexas es de 929, mientras que 883 de los nodos no están conectados con ningún otro, ya que los usuarios mencionan principalmente a Piqué (eliminado de la red) y a Iker Casillas. En general los usuarios no se mencionan entre si, salvo excepciones. Es por esto que el grado medio sea tan bajo, pues aunque Casillas recibe muchas conexiones, también hay muchos usuarios que han quedado aislados al eliminar a Piqué de la red.

La componente gigante de nuestra red está en torno a Iker Casillas (@casillasworld), ya que recibe la mayor parte de menciones de los usuarios. La componente gigante contienen un total de 725 nodos (un 41.12% del total) y 1345 aristas (91.87% del total). En la figura 1 podemos ver una primera visualización de nuestra red, mientras que en la figura 2 podemos ver la parte de la red correspondiente a la componente gigante.

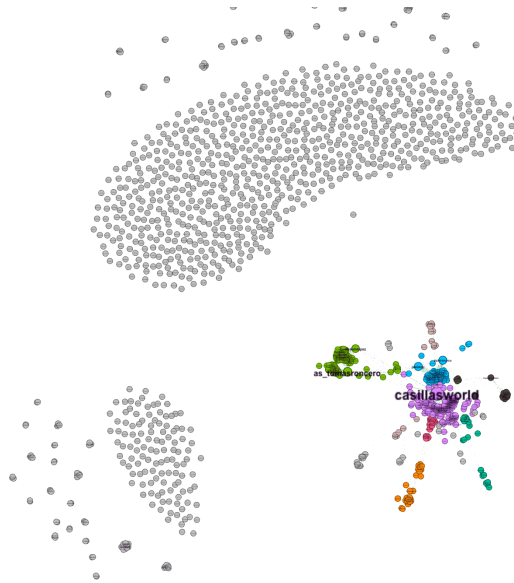


Figura 1: Red de estudio

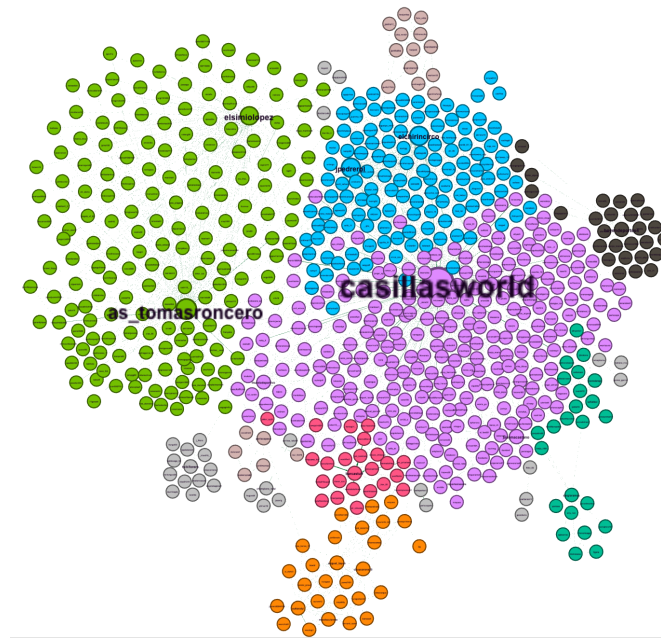


Figura 2: Componente Gigante de la red

Los nodos que hemos visto, cuyo grado es 0, no están conectados a ningún otro nodo de la red, pues mencionaban únicamente a Piqué (que ha sido eliminado de la red por ser el centro de la misma). Estos usuarios, no son de gran interés y en general son usuarios aislados que en algún momento han comentado la emisión. El resto de usuario que si están conectados a otros usuarios nos ayudarán a detectar comunidades y actores relevantes de nuestra red. En cuanto a la detección de comunidades, veremos más adelante exactamente que nos cuentan, pero a priori las comunidades de nuestra red deberían representar conversaciones entre diferentes usuarios sobre el evento.

## 2.3 Propiedades de la red

### 2.3.1 Distribuciones de Grados

En las figuras 3 y 4 podemos ver las distribuciones de grados de entrada y salida respectivamente, mientras que en la figura 5 podemos ver la distribución global de grados.

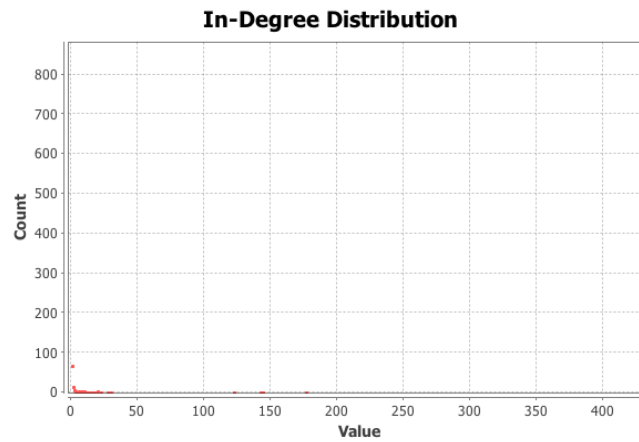


Figura 3: Distribución de grados de entrada

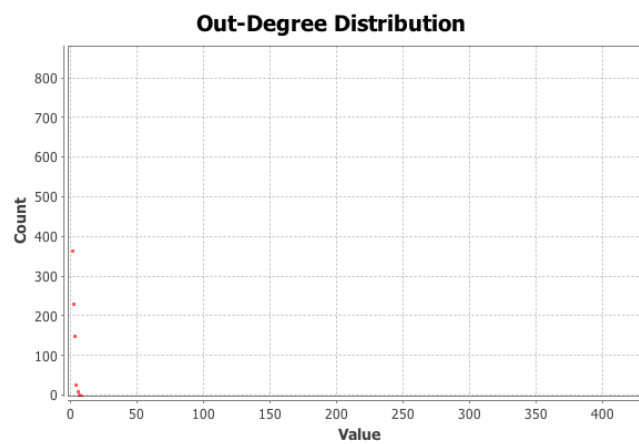


Figura 4: Distribución de grados de salida

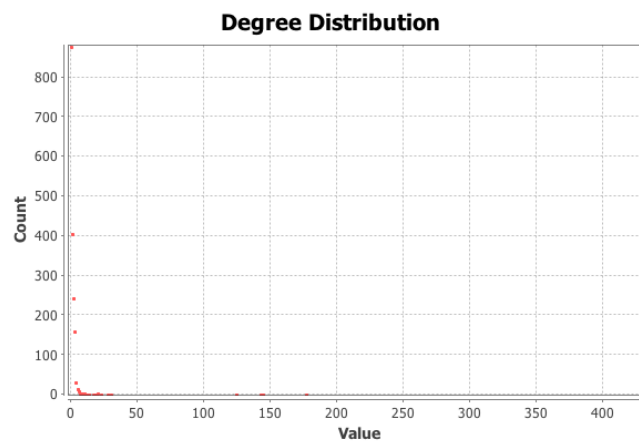


Figura 5: Distribución de grados

A partir de estas distribuciones de probabilidad de los grados de entrada y salida de los nodos de nuestra red

social, podemos determinar que nuestra red es libre de escala, y por tanto, que sigue la **Ley de la Potencia**. Las gráficas nos muestran una distribución de larga estela, como ya se ha mencionado en teoría, por ello, **podemos deducir que la red es libre de escala**. Tenemos una red con unos pocos nodos que concentran muchos enlaces hacia ellos (@casillasworld, @as\_tomasroncero, etc.). Estos nodos son los *hubs* de nuestra red. Los veremos más en detalle en las siguientes secciones.

### 2.3.2 Distribuciones de Distancias

En nuestra red pequeña, nos encontramos ante un caso de *mundo ultra-pequeño*, ya que como vimos anteriormente la distancia media ( $\langle d \rangle$ ) tenía un valor de 1.01, mientras que la distancia media para una red aleatoria equivalente era de 14.73. Para un mundo ultra-pequeño, la distancia media debe ser aún menor que  $\frac{\log(N)}{\log(\log(N))}$ , que en este caso obtiene un valor de 3.715, por lo que la distancia media de nuestra red tiene un valor más bajo incluso que la distancia media para un mundo ultra-pequeño. Por tanto, podemos concluir que **nuestra red social es un mundo ultra-pequeño**. En la figura 6 podemos ver la distribución de distancias.

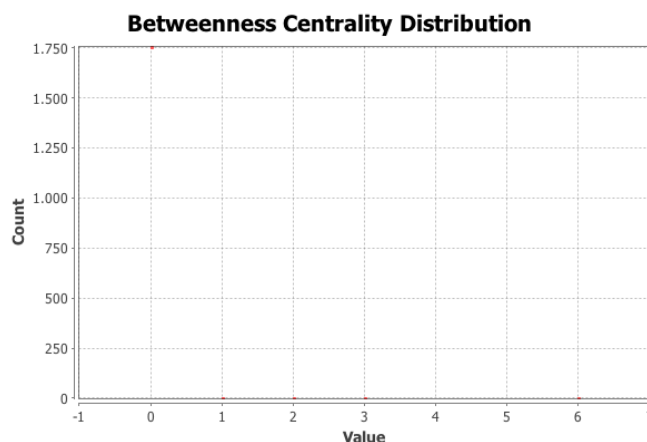


Figura 6: Distribución de Distancias

No se ve muy bien, pero si nos fijamos, cuando la distancia es 0, la frecuencia es muy alta. Sin embargo, conforme la distancia aumenta la frecuencia es mucho menor. Esto, como hemos visto, ocurre principalmente a causa de los nodos que se encuentran aislados, sin ninguna conexión (es un mundo ultra-pequeño, cada usuario es "su propio mundo").

### 2.3.3 Coeficiente de Clustering

El coeficiente de Clustering nos permite conocer la densidad local de la red, o dicho en otras palabras, la proporción de los vecinos de cada nodo que están conectados. En nuestra red, el coeficiente de clustering medio es de 0.05. Tenemos un coeficiente de clustering muy bajo, cosa que es lógica si nos fijamos en la gran cantidad de nodos que no están conectados con ningún otro nodo de la red. Como ya vimos antes, al eliminar a Piqué de la red, gran cantidad de nodos quedan totalmente desconectados y no tienen ninguna arista que los una a otros.

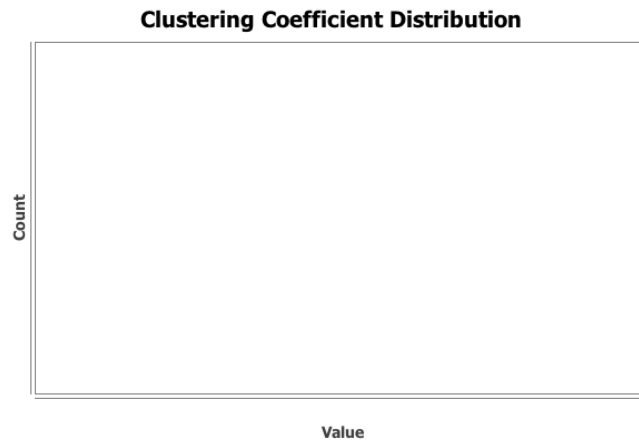


Figura 7: Distribución de Distancias

La figura 7 es el resultado que proporciona Gephi sobre el coeficiente de Clustering, como vemos, no es de utilidad porque no muestra nada en ella (he probado otras versiones por si fuese un problema de esta versión, pero el resultado es el mismo). Probablemente el problema es que no es capaz de representar valores tan bajos, ya que excepto unos cuantos nodos, la gran mayoría obtienen un coeficiente de Clustering igual a 0. Estos nodos son, evidentemente, los nodos que están aislados. Aún así, podemos visualizar la red de forma que, tanto el color de los nodos como su tamaño, se dibuje en función del coeficiente de Clustering de cada nodo. El resultado lo podemos ver en la figura 8.

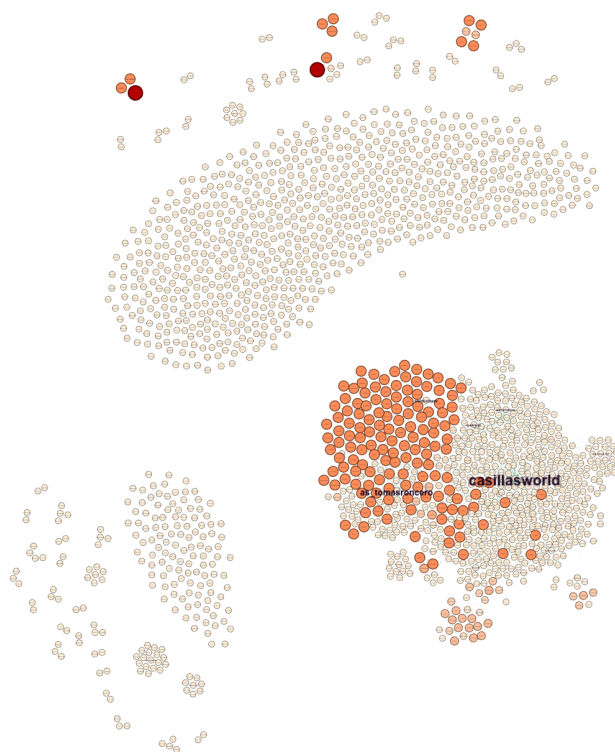


Figura 8: Visualización de la red, donde el color y tamaño de los nodos representan su coeficiente de Clustering



Como vemos, los nodos de la parte superior obtienen un coeficiente de clustering mayor, producido por que se conectan con los nodos que vemos cerca de ellos y que también tienen un coeficiente alto. Esto ocurre porque son pequeños grupos de usuarios que se mencionan entre ellos, creando pequeños grupos. Los nodos de la componente gigante que aparecen con un coeficiente de clustering mayor, son los nodos que se conectan (mencionan) a @as\_tomasrncero y a @elsimiolopez. Al tener estas dos conexiones en común, obtienen un coeficiente de clustering mayor. También podemos ver un pequeño grupo en la parte inferior de la componente gigante, en el que los usuarios se conectan entre si y por ello obtienen un coeficiente de clustering mayor.

Como hemos visto, el hecho de que el coeficiente medio de clustering sea tan bajo, nos indica que no hay un grado significativo de clustering local, cosa que hemos podido ver en la figura 8, en la que hemos visto que solamente hay unos pocos grupos.

## 2.4 Cálculo de los valores de las medidas de análisis de redes sociales

Ya hemos visto el valor medio de los grados, ahora vamos a ver para cada usuario el grado concreto y a tratar de analizar, a partir de esta y otras de las medidas de centralidad, cuales son los usuarios más importantes de nuestra red social. Para empezar, vamos a analizar los actores más interesantes de nuestra red en base al grado. Al trabajar con una red dirigida, trabajaremos con dos grados, el de entrada (que nos indica el prestigio de un usuario a la hora de ser citado/mencionado) y el de salida (que nos indica el alcance de la influencia de un usuario). En nuestro problema, nos interesa más el grado de entrada, pues nos indica los usuarios que han sido más mencionados, y por tanto, los más conocidos y/o interesantes para el resto de usuarios. Como no podría ser de otra manera, el usuario más mencionado (con mayor grado de entrada) es Iker Casillas. Esto ya lo habíamos visto antes, y como decíamos, es lógico al ser el protagonista de la emisión. Veamos que otros usuarios son también mencionados de forma frecuente.

### 2.4.1 Grado

| Usuario         | Grado Entrada | Grado Salida |
|-----------------|---------------|--------------|
| casillasworld   | 434           | 0            |
| as_tomasroncero | 177           | 0            |
| elchirincirco   | 144           | 0            |
| jpgedrrol       | 143           | 0            |
| elsimiolopez    | 123           | 1            |
| juanmacastano   | 30            | 0            |
| hoyendeportes4  | 28            | 0            |
| barcastuff      | 23            | 0            |
| chuycorona25    | 20            | 0            |
| miguel_layun    | 20            | 0            |
| mundodeportivo  | 18            | 1            |
| abc_deportes    | 17            | 0            |
| eukarolyi       | 15            | 3            |
| miseleccionmx   | 14            | 0            |
| txikiforero     | 13            | 0            |
| sefutbol        | 12            | 0            |
| sergioramos     | 10            | 0            |
| sientelaroja    | 10            | 0            |
| jordialba       | 9             | 0            |
| marchbartra     | 9             | 0            |

Tabla 1: Usuarios ordenados por grado de entrada

En la tabla 1 podemos ver la lista de usuarios más mencionados. Como podemos ver, a parte de Casillas, el usuario más importante de nuestra red es Tomás Roncero (@as\_tomasroncero), seguido por @elchirincirco y @jpgedrrol. Estos resultados me parecen interesantes y curiosos, pues muestran algunos detalles interesantes en nuestro problema. Es curioso que los usuarios más mencionados, y por tanto más importantes de nuestra red desde el punto de vista del grado de entrada, no sean jugadores de fútbol que se encontrasen convocados con la selección española, ya que en el momento de la emisión los jugadores estaban convocados y concentrados en el hotel para afrontar un serie de partidos amistosos. De hecho, me resulta curioso que haya usuarios mas mencionados que Cesc Fabregas, que aparecía junto a Piqué y Casillas en la emisión (aunque con mucho menos protagonismo). Pero de estas curiosidades que nos muestra la red, hablaremos más tarde en la sección dedicada a conclusiones.

### 2.4.2 Cercanía



Figura 9: Visualización de la parte central de la red. El color de los nodos indica, de menos rojo a más rojo, el valor de la medida de centralidad, de menor a mayor respectivamente. El tamaño de los nodos representa el grado

En la figura 9 podemos ver como queda la visualización de la parte central de la red, en la que el color de los nodos nos indica el valor de la medida de centralidad (Harmonic Closeness Centrality en Gephi). Si nos fijamos, esta medida no nos da demasiada información en nuestra red, ya que al estar gran parte de los nodos conectados a 'hubs', casi todos los nodos (usuarios) son importantes desde el punto de vista de esta medida. Como sabemos, el planteamiento de esta medida es diferente a la de intermediación (que veremos a continuación) y al grado, la centralidad trata de evaluar la cercanía de un nodo al centro de la red, a otros nodos centrales, a los hubs. Y como hemos visto en la imagen de nuestra red de la práctica, se puede ver fácilmente esto, como los nodos cercanos y conectados a los nodos centrales de la red obtienen un valor mayor de centralidad que el resto. Y por como es nuestra red, en la que muchos están conectados directamente a los hubs (Casillas, Roncero, ...), muchos de los nodos tienen una centralidad cuyo valor es 1. En la figura 9 no se ha incluido toda la red, sino una parte central para que se vea bien la medida de centralidad. Sin embargo, hay muchos nodos que han quedado totalmente desconectados al eliminar a Piqué de la red, como ya se ha comentado anteriormente. Con lo cual, estos nodos tienen un valor de centralidad igual a 0, cosa que, con lo que sabemos, es lógico. Por tanto, esta medida, al igual que el grado, nos permite ver cuales son los hubs, con respecto a los centrales que son. Y al igual que el grado, los nodos centrales (con un valor menor de centralidad, pero que permiten que los que se conecta a ellos tengan un valor alto), son prácticamente los mismo que veíamos según el grado.

### 2.4.3 Intermediación

| Usuario         | Intermediación |
|-----------------|----------------|
| eukarolyi       | 6              |
| pepvergeli      | 6              |
| jorditenerife95 | 3              |
| franarteaga     | 2              |
| culealcalaino   | 1              |
| damosasa        | 1              |
| real_oscarinn   | 1              |

Tabla 2: Usuarios ordenados por Intermediación

Como vemos en la tabla 2, tenemos usuarios menos conocidos que son los que obtienen los mejores valores para la medida de intermediación. Como sabemos, la medida de intermediación trata de reflejar en que medida un nodo se encuentra en el centro, en la zona intermedia entre diferentes conjuntos de nodos, de forma que la única forma de que la información fluya de un grupo a otro sea a través de él, permitiéndole controlar lo que pasa de un lado a otro. Estos nodos conectan grupos de nodos de forma que pueden controlar la información que se intercambia entre los grupos, por ello, desde el punto de vista de esta medida, estos son los usuarios más importantes.

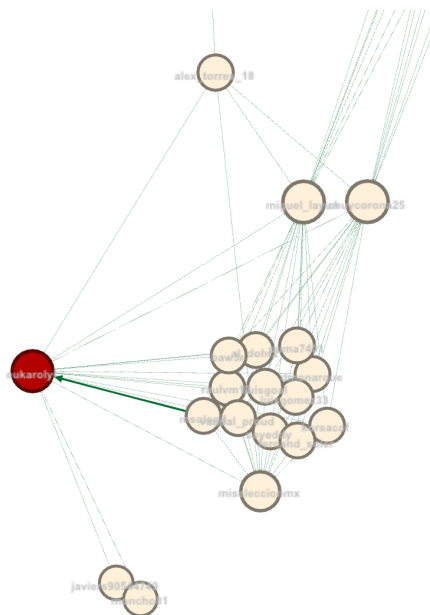


Figura 10: Visualización de la parte de la red unida por el nodo con mayor intermediación. El color de los nodos indica, de menos rojo a más rojo, el valor de la medida de intermediación, de menor a mayor respectivamente. El tamaño de los nodos representa el grado

En la figura 10 podemos ver lo que hemos visto anteriormente. El usuario *eukarolyi*, es el encargado de conectar los nodos de ese grupo (a simple vista esta es una especie de comunidad cuyo centro y el que une a todos los nodos es este usuario, veremos más tarde que nos dice el algoritmo de comunidades de Lovaina).

### 2.4.4 Vector Propio

| Usuario         | Vector Propio |
|-----------------|---------------|
| casillasworld   | 1             |
| as_tomasroncero | 0.467376      |
| elchirincirco   | 0.327179      |
| jpgedrrol       | 0.32491       |
| elsimiolopez    | 0.279027      |
| juanmacastano   | 0.068055      |
| hoyendeportes4  | 0.063518      |

Tabla 3: Usuarios ordenados por Vector Propio

La centralidad de Vector Propio es una versión más elaborada de la centralidad de grado, por ello, los resultados son prácticamente iguales, y como podemos ver en la tabla 3, los usuarios más importantes de nuestra red desde el punto de vista de esta medida, son los mismos que nos indicaba la centralidad de grado. De hecho, si nos fijamos en la figura 11, podemos ver como también los colores de los nodos y el tamaño (vector propio y grado) coinciden al representar a los nodos más importantes.

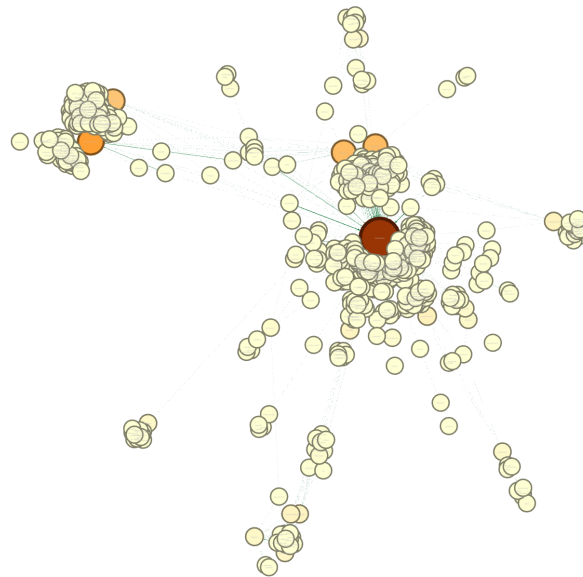


Figura 11: Visualización de la parte central de la red. El color de los nodos indica, de menos rojo a más rojo, el valor de la medida de Vector Propio, de menor a mayor respectivamente. El tamaño de los nodos representa el grado

### 2.4.5 PageRank

El PageRank es una forma diferente de calcular el vector propio, basandose en colocarse de forma aleatoria en un nodo e ir recorriendo recursivamente los enlaces de un nodo a otro, y contabilizando cuantas veces llegamos a los nodos a través de otros. De esta forma se reajusta el PageRank para cada nodo en función de las veces que llegamos a él a través de otros y el peso que tengan en ese momento los nodos que conectan con él. Personalmente es de las medidas que más me gusta, ya que su planteamiento me parece muy interesante. En nuestro problema, los actores más importantes siguen siendo los casi los mismos que nos encontrabamos

con el grado y con el vector propio de las subsecciones anteriores, ya que nuestro problema es muy concreto y esta muy claro quienes son los actores importantes.

| Usuario         | PageRank |
|-----------------|----------|
| casillasworld   | 0.095674 |
| as_tomasroncero | 0.054631 |
| elsimiolopez    | 0.020393 |
| jpedrerol       | 0.016179 |
| elchirincirco   | 0.016122 |
| hoyendeportes4  | 0.008424 |
| abc.deportes    | 0.006155 |
| juanmacastano   | 0.005478 |

Tabla 4: Usuarios ordenados por PageRank

Con el PageRank, como podemos ver en la tabla 4, aparecen usuarios que estaban posicionados peor según otras medidas. Por ejemplo, *@elsimiolopez* es más importante según el PageRank, de lo que lo era según el grado o según el vector propio. Aunque básicamente, todos los usuarios destacados por el PageRank, también eran importantes (algo más o algo menos) según el resto de medidas.

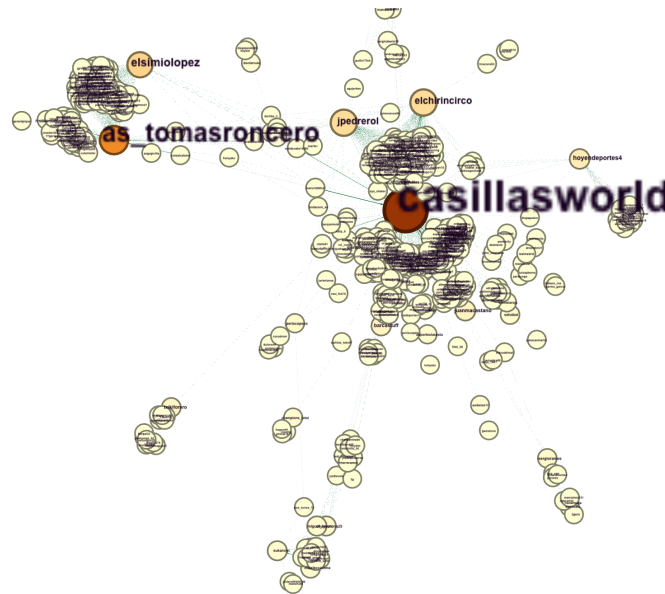


Figura 12: Visualización de la parte central de la red. El color de los nodos indica, de menos rojo a más rojo, el valor de la medida de PageRank, de menor a mayor respectivamente. El tamaño de los nodos representa el grado

En la figura 12 podemos ver una representación en la que los nodos representan el PageRank. Como vemos, es muy parecida a la visualización que obteníamos con el grado y con el vector propio.

#### 2.4.6 Medidas más interesantes

En esta sección hemos estudiado las medidas de centralidad en nuestra red, pero no todas son igual de interesantes de cara a obtener información en nuestro problema concreto. Entre todas las medidas que

hemos visto, bajo mi punto de vista y en vista de los resultados obtenidos por cada una de ellas de cara a responder la pregunta que nos hemos planteado, las más interesantes son **PageRank**, **Vector Propio y grado**. La intermediación puede ser interesante, pero para nuestro estudio no lo es, ya que solo nos permite descubrir usuario muy concretos que en pequeños grupos aparecen como intermediarios. Estos pequeños grupos no son de mucho interés en nuestro estudio. Tampoco es de gran interés la cercanía, aunque esta claro que de cara a utilizarla en la visualización para detectar los hubs/nodos centrales, puede ser interesante; aunque para ello ya tenemos las medidas que hemos dicho que si son interesante, como el PageRank o el vector propio.

## 2.5 Descubrimiento de comunidades

A continuación, vamos a tratar de detectar comunidades en nuestra red. Anteriormente ya nos habíamos imaginado, a partir de la visualización de la red, que podrían existir algunas comunidades, y ahora es el momento de comprobarlo utilizando un algoritmo específico para tal efecto. Debido a las limitaciones de la última versión de Gephi por las cuales no están disponibles todos los algoritmos de detección de comunidades, solamente se ha utilizado el algoritmo de Lovaina.

En un primer intento, utilizando como valor para el parámetro de resolución 1.0, la separación en comunidades, desde el punto de vista de la medida de Modularidad, si hay comunidades, ya que se obtiene un valor  $Q=0,647$ . El problema es que, aunque ese valor es muy bueno, implica que se detecten 942 comunidades, que vienen principalmente de los nodos inconexos de nuestra red (a causa de eliminar a Piqué). Sin embargo, como podemos ver en la figura 13, las comunidades que ha detectado parecen correctas y muy acertadas.

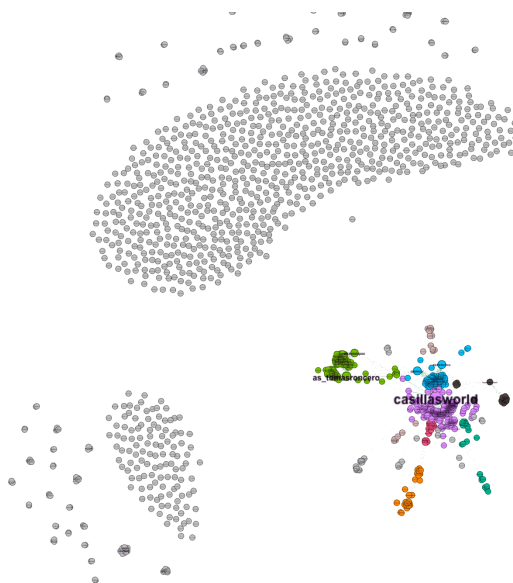


Figura 13: Comunidades de la red detectadas por el algoritmo de Lovaina

En la figura 13 podemos ver que las 942 comunidades, son realmente comunidades de nodos individuales inconexos. Realmente tenemos 8 comunidades importantes, como podemos ver en la figura 15, mientras que el resto son comunidades que ha detectado el algoritmo correspondientes a nodos aislados, y que no nos interesan.

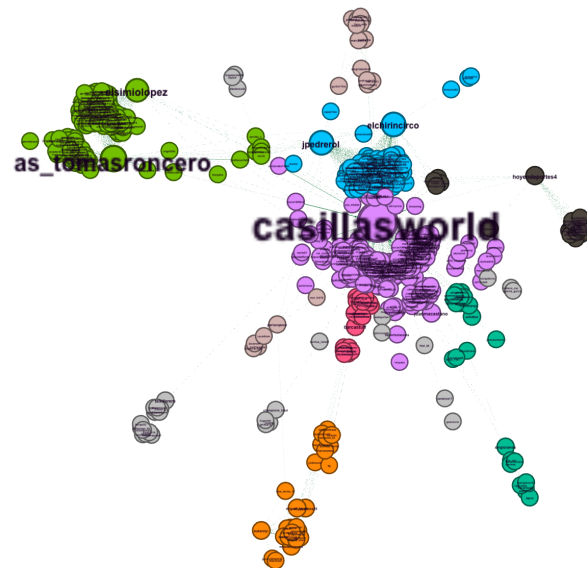


Figura 14: Comunidades interesantes de la red detectadas por el algoritmo de Lovaina

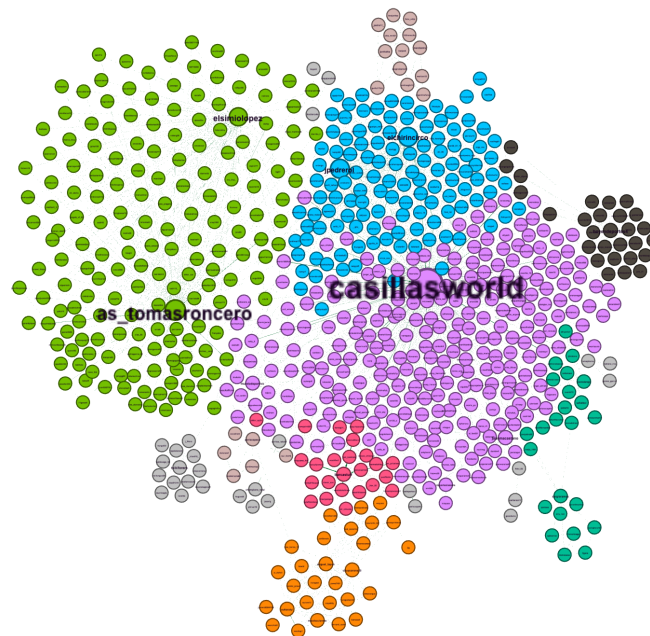


Figura 15: Comunidades interesantes de la red detectadas por el algoritmo de Lovaina sin solapar

De las 8 comunidades que si que son interesantes, podemos ver las siguientes:

- **casillasworld**: Comunidad formada, principalmente, por los usuarios que han mencionado a Casillas. No todos los usuarios que han mencionado a Casillas forman parte de esta red, hay algunos otros que al haber mencionado en el mismo tweet a otros usuarios de otras comunidades han quedado asignados a esas otras comunidades (Ver figura 16).
- **as\_tomasroncero y elsimiolopez**: Comunidad formada, principalmente, por los usuarios que han mencionado a Roncero y a elsimiolopez. Aparentemente buscando hacer comentarios graciosos sobre Ron-



cero y mencionando a elsimiolopez, que es una cuenta de humor. Básicamente podríamos decir que vemos a los alumnos enseñándole lo aprendido a su profesor de humor, aunque con los tweets recogidos no les hace ningún caso.

- **elchirincirco y jpedrerol**: Comunidad formada, principalmente, por los usuarios que han mencionado a elchirincirco y a jpedrerol. Esta comunidad refleja exactamente lo mismo que la anterior, usuarios mencionando a jpedrerol para hacer la gracia y que elchirincirco las lea y les retweetee. Esto es algo muy común en Twitter.
- **hoyendeportes4**: Comunidad formada, principalmente, por los usuarios que han mencionado a hoyendeportes4. La intención de estos usuarios parece ser avisar a la cuenta de Deportes Cuatro que Piqué está haciendo una emisión.
- **sefutbol y sergioramos**: Comunidad formada, principalmente, por los usuarios que han mencionado a sefutbol y a sergioramos (Ver figura 17).
- **barcastuff**: Comunidad formada, principalmente, por los usuarios que han mencionado a barcastuff.
- **miguel\_layun y chuycorona25**: Comunidad formada, principalmente, por los usuarios que han mencionado a miguel\_layun y a chuycorona25 (Ver figura 18).
- **\_pedro17-, jordialba, marcbartra y sergiroberto**: Comunidad formada, principalmente, por los usuarios que han mencionado a \_pedro17-, jordialba, marcbartra y sergiroberto (Ver figura 19).

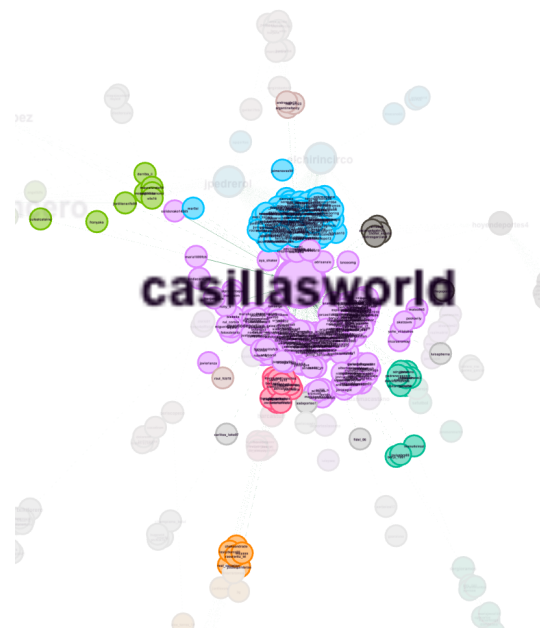
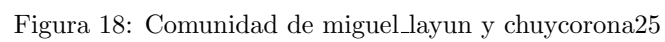
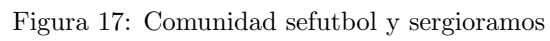


Figura 16: Comunidad Casillas



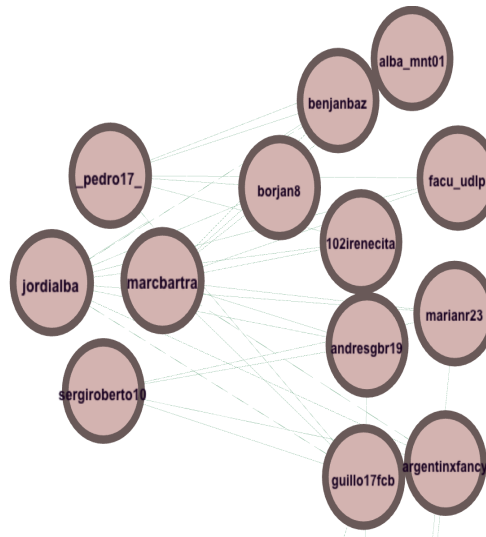


Figura 19: Comunidad de \_pedro17\_, jordialba, marcbartra y sergiroberto

Como hemos visto, estas 8 son las comunidades principales detectadas por el algoritmo, las restantes corresponden a nodos aislados. Estas 8 comunidades tienen sentido, principalmente las comunidades de Casillas, Roncero y elsimiolopez, elchirincirco y jpedrerol, y la de hoyendeportes4. Estas son las principales comunidades y las más claras, como hemos podido ver más fácilmente en las figuras de la visualización de la red.

Digamos que no son "comunidades reales" por así decirlo, ya que no se agrupan de forma real, pero sí que bajo las menciones entre los usuarios que forman parte de ellas, se pueden ver fácilmente su agrupación. Y dicha agrupación tiene sentido y esta clara. Tenemos un grupo cuya característica principal es que mencionan a Casillas durante la emisión, otro grupo que menciona a elsimiolopez y a Roncero en un mismo tweet, lo que refleja que este grupo de usuarios lo que buscan es hacer una serie de chistes que involucran a Roncero y tratan de ser retwitteados por una cuenta de humor como la de @elsimiolopez. Igual ocurre con el grupo que centra sus chistes en Pedrerol y que busca la atención de la cuenta de humor @elchirincirco. Esto último tienen sentido, pues Pedrerol es el presentador del programa de televisión "El Chiringuito", y la cuenta @elchirincirco es una cuenta parodia de este programa. Por otro lado, tenemos un grupo de usuarios que parecen tratar de avisar a los encargados de la cuenta de los deportes de la cadena de televisión Cuatro (@hoyendeportes4). Y finalmente, el resto de comunidades son más pequeñas y menos importantes en nuestra red, pero representan grupos de personas con otros intereses. Por ejemplo, decir algo sobre Sergio Ramos o, como veíamos con la comunidad en torno a Deportes Cuatro, tratan de avisar a los encargados de @sefutbol, que es la cuenta de Twitter oficial de la Selección Española. También es curiosa la última comunidad que hemos visto que representa a un grupo de usuarios que menciona a varios jugadores de FC Barcelona (\_pedro17\_, jordialba, marcbartra y sergiroberto).

## 2.6 Visualización de la red

A lo largo de esta documentación, se ha hecho uso en varias ocasiones de las utilidades de visualización que ofrece Gephi, de forma que la información obtenida y las conclusiones se puedan explicar mejor gracias a dichas visualizaciones. Como hemos visto, las visualizaciones han sido de gran ayuda para comprender y explicar mejor la red que teníamos entre manos.

Para la visualización de la red se ha usado principalmente como algoritmo de distribución **ForceAtlas 2**, que es el algoritmo que mejor representa (al menos a mi modo de ver) la red con la que hemos trabajado,

ya que agrupa los nodos que se relacionan entre sí, lo que permite ver mejor las comunidades. De hecho, desde el principio y sin haber utilizado aún el algoritmo de Lovaina, ya se podían ver e intuir las comunidades que encontraríamos más tarde en nuestra red.

Para las visualizaciones que hemos visto en las que se destaca el grado de los nodos, el PageRank y algunas otras medidas de análisis de redes, se han utilizado las opciones disponibles para ello en Gephi. Como por ejemplo, que el tamaño o el color del nodo se apliquen en función de alguna de estas medidas.

### 3 Discusión de los resultados y Conclusiones

Como hemos visto a lo largo de este trabajo, hemos sido capaces de responder a la pregunta de investigación que nos hacíamos al comienzo. Como hemos ido viendo en las distintas secciones de esta documentación, los usuarios más importantes en nuestra red estaban relativamente claros independientemente de las medidas de análisis que usásemos. Todas nos conducían de un modo u otro a los mismos usuarios.

El usuario más importante de toda la red es Iker Casillas (@casillasworld), que como hemos visto es el usuario con mayor número de menciones, y por tanto, con mayor grado de entrada. El resto de medidas, como hemos visto, también destacaban a este usuario y le daban el mayor valor de toda la red. Esto es algo totalmente lógico si tenemos en cuenta que fue el protagonista de la emisión junto a Piqué, que actuó mas bien como una especie de entrevistador.

Por otro lado, tenemos que Tomás Roncero (@as\_tomasroncero) es, según casi todas las medidas que hemos visto, el segundo personaje más importante de nuestra red. Pero no solo es el segundo más importante, sino que hemos visto gráficamente y gracias a la detección de comunidades, que de su mano va también el usuario @elsimiolopez (cuenta de humor y aparentemente parodia de un compañero de Roncero, Siro Lopez). Hemos sido capaces de detectar que uno de los protagonistas en Twitter durante la emisión fue Tomás Roncero, y que el objetivo de los usuarios que lo mencionaban parece ser (sin mirar los tweets concretos) hacer chistes sobre el y buscar la atención de una cuenta de humor como la de @elsimiolopez. De esta forma, los usuarios probablemente buscaban ser retwitteados por esta cuenta humorística para ser más populares y que más usuarios les conozcan. Como he dicho, todo esto sin entrar a ver los tweets concretos y sin comprobar si realmente estas suposiciones son correctas. Aunque como usuario de Twitter, se que este es el comportamiento habitual, los usuario aprovechan este tipo de acontecimientos para llevarse su minuto de gloria haciendo algún tweet gracioso y recibiendo algún retweet por parte de alguna cuenta de humor con muchos más seguidores.

De igual manera que hemos visto con Roncero y @elsimiolopez, también tenemos un caso similar con las cuentas de Pedrerol (@jpedrerol), presentador del programa de televisión sobre fútbol, "El Chiringuito", y @elchirincirco. En este caso, tenemos que este comportamiento parece lógico, ya que al ser Pedrerol el presentador del programa es normal que pueda recibir gran parte de las críticas y los chistes. Por tanto, la conclusión para este par de usuarios es la misma que para los anteriores, un grupo de usuarios que mencionan a estos dos con el objetivo de hacer chistes que tienen como protagonista a Pedrerol y como objetivo ser retwitteados por @elchirincirco.

Finalmente, a parte de los protagonistas que hemos visto, tenemos las cuentas de @hoyendeportes4, @abc\_deportes y @juanmacastano, que como hemos visto en algunas medidas de análisis como el PageRank, el Vector Propio y el Grado, eran los siguientes usuarios más importantes de nuestra red. Algunos de ellos, aún siendo usuarios importantes, no "lideraban" una comunidad por si sola (como habíamos visto con Casillas, Pedrerol o Roncero), como hemos visto, sino que formaban parte de comunidades más grandes. Esto ocurre porque se han mencionado junto a usuarios más importantes, por lo que a la hora de descubrir comunidades tienden a

formar parte de otras comunidades, como es el caso de @juanmacastano, que es mencionando generalmente junto a Casillas, por lo que él y los usuarios que los mencionan forman parte de la comunidad en torno a Casillas.

Estos han sido principalmente los usuarios más importantes de la red y las conclusiones que podemos sacar de ello. Pero en torno a las comunidades no solo hemos sido capaces de sacar conclusiones como las anteriores, sino que también hemos detectado algunas cosas curiosas. Por ejemplo, parece interesante que tengamos una comunidad en la que se incluyen 4 jugadores del FC Barcelona (-pedro17-, jordialba, marchbartra y sergiroberto) y los usuarios que los mencionan.

También me ha parecido curioso que jugadores como Sergio Ramos, Jordi Alba o Marc Bartra hayan sido más mencionados y aparezcan con valores superiores de vector propio que otros jugadores, como por ejemplo Cesc Fabregas que estuvo presente en la emisión junto a Piqué y Casillas. Si bien es cierto que uno de los espectadores de Periscope hizo una pregunta relativa a Bartra y se habló unos segundos de él, pero bajo mi punto de vista es curioso que se mencionase más a Bartra que al propio Cesc.

## Notas Finales sobre la Entrega

Junto a esta documentación en formato PDF se entrega **el fichero generado por Gephi con la red**, de forma que se pueda visualizar correctamente. Por si este fichero no funcionase correctamente también se incluye **el fichero CSV original** que se obtuvo a través de NodeXL para que simplemente sea necesario abrirlo con Gephi, aunque no estarán las medidas calculadas, ni la visualización ni Piqué estará eliminado, por lo que habría que hacerlo. En mi caso uso Mac, por lo que es posible que el fichero de Gephi no funcione. Yo obtuve el fichero CSV con NodeXL y lo abrí en Windows con Gephi, hice algunos cálculos y lo pase a Mac, pero no conseguí abrir el fichero de Gephi que genere en Windows y tuve que empezar desde el principio en Mac.