



**UNIVERSIDAD  
DE GRANADA**

**Gestión de Información en la Web**

**Práctica 2. Caso Práctico de Análisis y Evaluación  
de Redes en Twitter**

**2019-2020**

**Máster en Ingeniería Informática**

Lidia Sánchez Mérida

[lidiasm96@correo.ugr.es](mailto:lidiasm96@correo.ugr.es)

# Índice de contenidos

Temática de investigación y medio social	3
Análisis de la red obtenida	4
Propiedades de la red	6
3.1. Distribución de grados	6
3.2. Análisis de la distancia media	7
3.3. Análisis del coeficiente de clustering medio y reciprocidad	7
Medidas de análisis de redes sociales	8
3.1. Intermediación	8
3.2. Cercanía	9
3.3. Vector propio	9
Comunidades en la red	10
Visualización	12
6.1. Centralidad intermedia	12
6.2. Centralidad de cercanía	13
6.3. Vector propio	15
6.4. Comunidades de la red y subredes	16
6.5. Grado bidireccional, de entrada y salida	19
Conclusiones	21
Bibliografía	22

## 1. Temática de investigación y medio social

El día del libro comienza a celebrarse en el siglo XX como homenaje a las obras literarias en España. Se asoció el día 23 de abril como fecha oficial debido a que en ese mismo día de 1616 fallecían tres de los principales escritores de la literatura española. Sin embargo, este mismo día también fue testigo del nacimiento y fallecimiento de otra serie de escritores durante algunos años, de manera que esta fiesta se ha extendido a nivel internacional. Es tradición regalar un libro entre los seres queridos para conmemorar esta simbólica fiesta de la literatura.

La comunidad del día del libro es internacional, pero la interacción entre los lectores en este día tan especial es más bien local. De modo que la primera pregunta a plantear es la siguiente: **¿la conectividad de la comunidad del día del libro en Twitter será igual de reducida que en la realidad?**. Asimismo, el grado de influencia de cualquier red social es sumamente considerable, llegando a alcanzar millones de usuarios pero en un ambiente más emotivo, **¿cuántos usuarios existirán con tal poder de influencia? ¿serán escritores o usuarios amantes de la lectura?**.

Para responder a las preguntas descritas anteriormente he escogido como medio social **Twitter**, por lo que he utilizado el *plugin Twitter Streaming Importer* del que dispone la herramienta Gephi para descargar una red de datos asociados a esta temática. Previo a este paso me he creado una cuenta de desarrollo y he registrado mi aplicación para obtener los credenciales necesarios con los que realizar la conexión con la API de Twitter a través de dicho *plugin*. A continuación he añadido los siguientes hashtags **#diadellibro**, **#diadellibroencasa**, **#diadellibro2020** y he seleccionado la opción **User Network** para obtener información sobre la interacción que realizan los usuarios que han participado en esta temática. Algunos de los campos descargados por este *plugin* se detallan a continuación:

- *Id*. Es el identificador del usuario, el cual generalmente es su *username*.
- *Label*. Es la etiqueta que representa al nodo. Como mi red está compuesta de usuarios, en mi caso se corresponde con el nombre de usuario.
- *Timeset*. Es la fecha y hora a la que se produjo la interacción.
- *Twitter\_type*. Es el tipo de dato descargado. En mi caso todos los datos son de tipo usuario.
- *Created\_at*. Es la fecha en la que se creó la cuenta el usuario. En algunos registros este dato no se encuentra disponible y por tanto no tiene valor.
- *Description*. Representa el texto del tweet que ha escrito el usuario en cuestión.
- *Friends\_count*. El número de usuarios a los que sigue un usuario particular.
- *Followers\_count*. El número de seguidores que tiene el usuario.
- *Real\_name*. El nombre completo del usuario, en caso de que de que lo haya especificado. Si no, este campo no tiene valor.

## 2. Análisis de la red obtenida

Como hemos podido observar anteriormente en la descripción de algunos de los campos de la red, esta **no es ponderada** por lo que cada enlace tiene la misma importancia que el resto.

Por otro lado, sí que disponemos del orden en el que se ha realizado la interacción entre los usuarios, por lo que conocemos quién de los dos ha mencionado, citado o retuiteado el tweet del otro. Por lo tanto podemos determinar que es una **red dirigida**, y habrá que tomarlo en consideración para los futuros análisis.

La red social descargada dispone de un total de **8.354 nodos**, los cuales representan información acerca de los tweets que han postado los usuarios con respecto a los *hashtags* establecidos anteriormente. Asimismo, este grafo dispone de **13.780 enlaces**, que representan la interacción de otros usuarios con los tweets asociados a esta temática, como los retuits, las menciones, entre otros. Al ser un grafo dirigido podemos calcular el **número de enlaces máximo** como se explica en esta fuente [1] dando un total de **69.780.962**. Si comparamos este valor con el número de enlaces reales en la red, podemos explicar que la **densidad sea tan baja**, ya que este atributo mide el número de interacciones existentes entre el total posible. De hecho Gephi ni siquiera ha podido calcularla por ser tan cercana a 0 por lo que la he calculado manualmente con la siguiente fórmula para grafos dirigidos [2]:

$$\frac{|E|}{|V| \cdot (|V| - 1)}$$

Por lo tanto, podemos determinar que la mayoría de los nodos no dispondrán de un número alto de enlaces. Esta teoría acompaña el bajo valor del **grado medio**, el cual nos indica que cada nodo, de media, tendrá aproximadamente 2 conexiones. Por lo tanto, cada tweet habrá sido mencionado, citado o retuiteado unas dos veces.

En relación al **diámetro** de la red podemos determinar que solo hacen falta seis nodos para alcanzar el nodo más lejano, lo que nos indica que el camino mínimo para alcanzar al tweet más aislado es relativamente corto. De igual modo, el bajo valor de la **distancia media** determina que, en esta red, el camino mínimo para alcanzar un nodo está formado por aproximadamente dos nodos. Esto nos indica que, en este tipo de red, el grado de difusión es muy rápido puesto que apenas supone un esfuerzo visualizar el tweet de cualquier usuario.

Si comparamos el valor de la distancia media real con el de la **distancia media teórica** podemos observar que este último es mucho mayor, puesto que en una red equivalente aleatoria no se considera la existencia de nodos *hubs* que facilitan el alcance hacia los nodos que conecta.

Medida	Valor
Número de nodos	8.354
Número de enlaces	13.780
Número máximo de enlaces	69.780.962
Densidad del grafo	0,000197475
Grado medio	1,650
Diámetro	6
Distancia media	1,6505565964950952
Distancia media para una red aleatoria equivalente	18,033
Coeficiente de Clustering medio	0,110
Coeficiente de Clustering medio para la red aleatoria equivalente	0,00019751
Número de componentes conexas	1.382
Número de nodos componente gigante (%)	2.206 (26,41%)
Número de aristas componente gigante (%)	4.184 (30,36%)

Como podemos apreciar el **coeficiente medio de clustering** es bastante bajo, lo cual sigue indicando que la conectividad local también es baja. Este hecho lo hemos podido percibir por la densidad y el grado medio ya que apuntan a que cada usuario ha interaccionando con 2 personas aproximadamente. Sin embargo, el **coeficiente teórico para una red aleatoria** es mucho más pequeño aún puesto que en este cálculo no se ha considerado que la cantidad de relaciones locales entre los usuarios es más alta debido a la facilidad proporcionada en Twitter para reaccionar a un tweet de uno de nuestro seguidos.

En relación a la conectividad de la red podemos visualizar que esta se puede dividir en **1.382 componentes conexas**, lo que indica que en realidad existen 1.382 subgrafos. Este hecho también está provocado por la existencia de muchas subredes con pocos nodos, es decir, existen muchos tweets con los que interaccionan un bajo número de usuarios. El hecho de que existan tantos subgrafos diferentes, me lleva a reflexionar que las comunidades existentes serán muy dispersas y, por tanto, habrá que reducir la red para intentar agruparlas lo máximo posible y así poderlas representar gráficamente.

### 3. Propiedades de la red

En esta sección vamos a analizar los distintos gráficos obtenidos de las distribuciones de las medidas explicadas anteriormente con el objetivo de determinar las propiedades más características de la red.

#### 3.1. Distribución de grados

Comenzamos estudiando la **distribución de grados de entrada**, en la cual podemos apreciar que la mayoría de los tweets de los usuarios apenas han recibido unas cuantas interacciones y por lo tanto se concentran en los valores más pequeños del gráfico. No obstante, existen algunas excepciones, especialmente a partir del valor 100 que son los llamados **hubs**, usuarios que sí que han recibido un mayor número de interacciones puesto que tienen más seguidores que los mencionan, citan o retuitean sus posts.

De un modo similar ocurre en la **distribución de grados de salida**, en la que la mayoría de usuarios no realizan demasiadas interacciones con los tweets de otros. En este caso es aún más evidente puesto que hay un menor número de nodos **hubs** y una mayor concentración de usuarios en los primeros valores del intervalo. Por lo tanto estos resultados me llevan a pensar que la temática del día del libro está más enfocada a un ámbito participativo individual, es decir, cada usuario propone sus libros favoritos o recomendaciones pero la mayoría no reciben demasiadas interacciones, además de que tampoco las realizan.

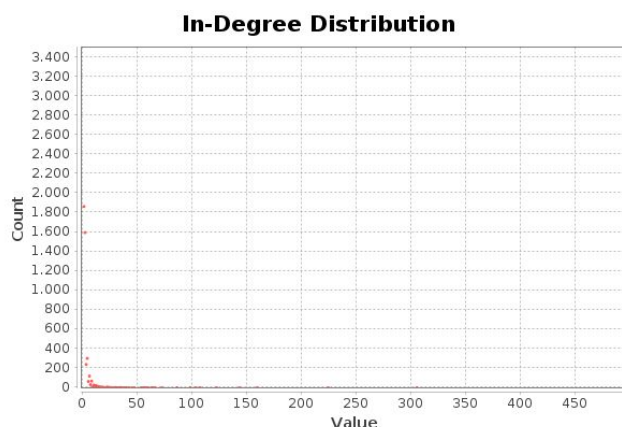


Figura 1. Distribución de grados de entrada.

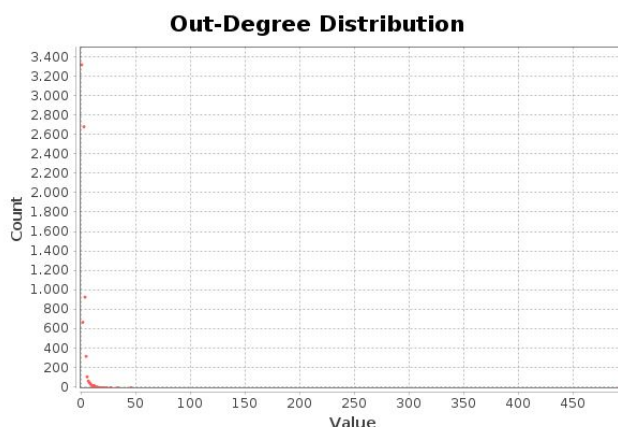


Figura 2. Distribución de grados de salida.

Observando la distribución de grados de entrada podemos apreciar una larga lista de nodos que tienen un mayor número de interacciones en sus tweets. Esta estela es característica de aquellas redes cuyas distribuciones de grados siguen la **ley de la potencia**, y por tanto la red de entrada es **libre de escala**. Esto simboliza que existe la probabilidad de que haya nodos que son capaces de disponer de un mayor número de conexiones de entrada que el resto de nodos de la red.

Para la red de salida podemos observar una situación similar aunque no tan pronunciada pues la cola es de menor longitud ya que, en este caso, los usuarios que han mencionado, citado y/o retuiteado más que la media no son muy numerosos (no hay muchos *hubs*).

### 3.2. Análisis de la distancia media

La distancia media real de la red obtenida es de 1,65 mientras que la distancia teórica calculada es 18,033. Como podemos observar, claramente la distancia real es más inferior a la teórica calculada para comprobar si es un *mundo pequeño*. Mientras que para comprobar si se trata de un *mundo ultra-pequeño* el resultado es de 4,103 por lo que podemos determinar que **mi red social está más cerca de ser un mundo ultra-pequeño**. La razón de ser reside en la existencia de nodos *hubs* que facilitan el acceso a los nodos más alejados y por lo tanto reducen increíblemente la distancia media. Como los cálculos anteriores están pensados para realizarlos sobre redes aleatorias y en estas no existen tal tipo de nodos, las distancias son bastante mayores que las de una red social real.

### 3.3. Análisis del coeficiente de clustering medio y reciprocidad

Como podemos apreciar en la siguiente imagen, la mayoría de usuarios apenas reciben interacciones en sus tweets ya que muchos de los nodos se concentran en torno a un coeficiente de clustering 0. Esto nos indica que la **conectividad local es bastante baja**, lo cual ya habíamos intuido debido a la densidad y el grado medio. Sin embargo, también podemos visualizar una concentración de nodos entre los valores 0,4 y 0,5, superiores a la media, que se corresponden con algunos nodos cuya conectividad local es más alta. En términos de mi red, podemos decir que se corresponden con aquellos usuarios que tienen una red de amigos con la que están conectados. Estos nodos no actúan como nodos *hubs* puesto que estos se caracterizan por tener un coeficiente de clustering más bajo ya que su objetivo es conectar nodos aislados, y por lo general no suelen tener una red de vecinos amplia.

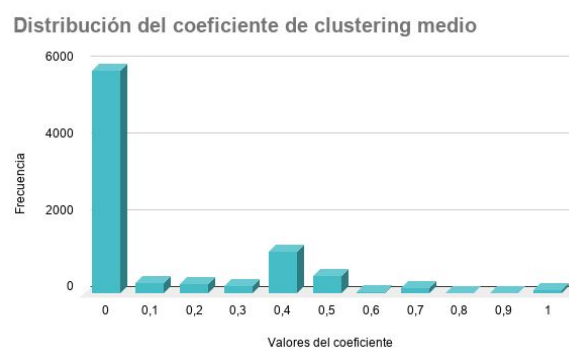


Figura 3. Distribución del coeficiente de clustering medio.

En lugar de calcular la transitividad de la red debemos optar por la **reciprocidad** puesto que se trata de una red dirigida. De este modo podemos obtener la probabilidad de que dos usuarios hayan interactuado mutuamente. Para ello es necesario aplicar la siguiente fórmula según se establece en los apuntes de la asignatura:

$$R = \frac{\#mut}{\#mut + \#asim}, R \in [0, 1]$$

donde en el denominador se representan el número de conexiones mutuas (**díadas**) y el denominador representa el número total de enlaces de la red. Para obtener el número de díadas he investigado acerca de cómo filtrar la red con *Gephi*. Según este artículo [4] dentro de la opción *Aristas* existe una en particular denominada *Aristas mutuas* que es la que calcula las conexiones bidireccionales. En mi caso, la red dispone de **438 díadas**, lo que supone un 3,18% del total de enlaces. Por lo tanto, si realizamos los cálculos con los datos que disponemos:

$$R = \frac{438}{13.780} = 0,0317$$

Como podemos apreciar la **reciprocidad de la red es bastante baja** puesto que su valor se aproxima bastante al límite inferior del intervalo. Esto nos indica que la mayoría de conexiones no son recíprocas, es decir, la mayoría de usuarios no han mencionado, citado y/o retuiteado los tweets entre ellos. De este modo, en la red obtenida, un usuario ha podido interactuar con el tweet de otro usuario pero este no ha participado en ninguno de los posts del primero.

## 4. Medidas de análisis de redes sociales

### 3.1. Intermediación

A continuación vamos a analizar el número de nodos que actúan como nexo de unión. Para ello vamos a analizar el siguiente gráfico en el que se muestran los valores de **intermediación** de los usuarios de la red. Como podemos apreciar la mayoría se concentra en el primer intervalo del gráfico donde los valores son más bajos, por lo que podemos determinar que la mayoría no tienen un papel crucial en los caminos en los que participa. Sin embargo, existen algunas excepciones que sí disponen de valores más altos. Estos usuarios son los que conectan a una mayor cantidad de nodos debido a que han recibido y/o realizado muchas interacciones, por lo que tienen el poder de controlar qué información les llega a cada uno de los usuarios con los que están enlazados.

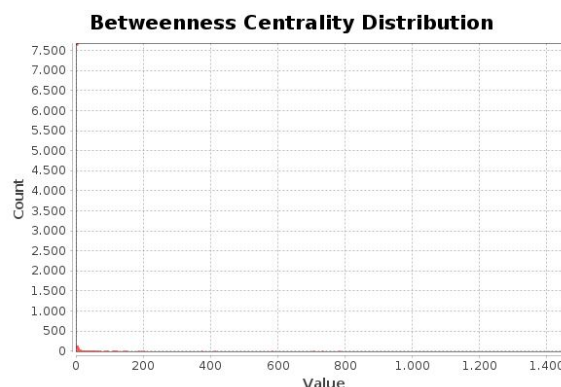


Figura 4. Distribución de la centralidad de intermediación.



### 3.2. Cercanía

En esta sección vamos a estudiar cuántos de los usuarios de la red se encuentran más cerca del centro de la misma. El objetivo es conocer cuáles son los que tendrían que dar un menor número de saltos para alcanzar cualquier tweet. Como podemos observar en la gráfica existe un primer gran grupo de usuarios con cercanía 0, lo que nos indica que son nodos muy alejados del centro de la red. Mientras que por otra parte, existe otro gran grupo de usuarios con cercanía 1, los cuales se encuentran en la parte central. Finalmente, existe un tercer grupo de usuarios minoritario que se encuentra en valores intermedios, de manera que están a una distancia media del centro de la red por lo que no son nodos céntricos pero tampoco están demasiado alejados.

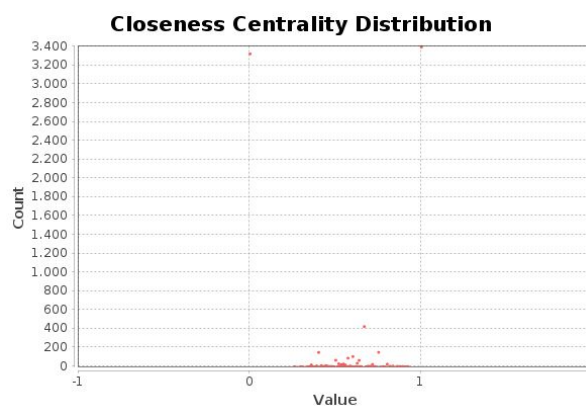


Figura 5. Distribución de la centralidad de cercanía.

### 3.3. Vector propio

En este caso vamos a estudiar la importancia de los usuarios en función de la relevancia de sus vecinos. De este modo podemos visualizar qué usuarios son más influyentes por el hecho de estar conectados a usuarios que se denominan *influencers*. Como podemos apreciar en la gráfica, la mayoría de los nodos tienen una baja importancia debido a que se encuentran conectados con usuarios nada relevantes. No obstante, existen ciertas excepciones de usuarios con un mayor valor. En la sección de visualización podremos observar estos resultados de forma gráfica para comprobar dónde se sitúan este tipo de nodos.

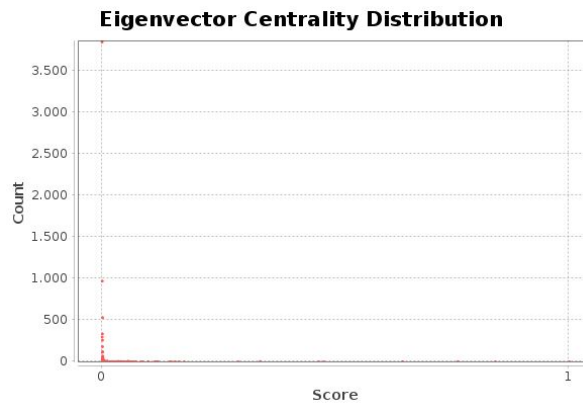


Figura 6. Distribución de la centralidad de vector propio.

## 5. Comunidades en la red

En esta sección vamos a analizar las comunidades existentes dentro de la red obtenida. Sin embargo, por los valores tan bajos tanto en el grado medio, la densidad y el coeficiente medio de clustering podemos intuir que muchas de ellas dispondrán de un conjunto muy reducido de nodos. Del mismo modo, en función de la conectividad general podemos intuir que estas comunidades estarán muy aisladas entre sí y será complicado agruparlas. Y, efectivamente, tras estudiar la **modularidad** de la red con los valores por defecto hemos obtenido un total de **1.410 comunidades**. Como podemos observar en el gráfico de su distribución la gran mayoría de ellas tienen muy pocos miembros. Esto confirma de nuevo la baja conectividad de la red y afirma la teoría de la baja participación en esta temática.

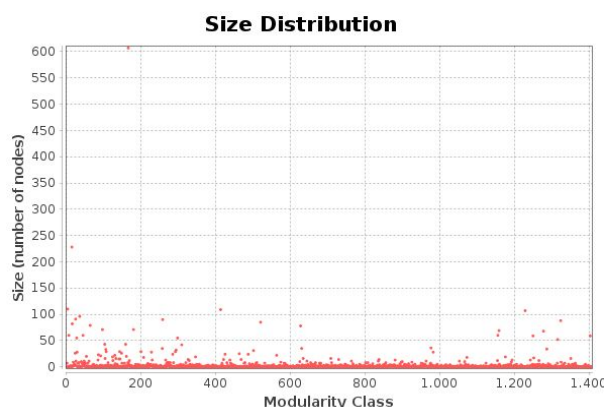


Figura 7. Distribución del tamaño de comunidades.

Tras intentar adaptar el parámetro de la resolución para agrupar aún más las comunidades el valor más pequeño encontrado para la red al completo ha sido de **1.382 comunidades**. Como predijimos, las comunidades disponen de muy pocos miembros y están muy aisladas entre sí, por lo que el algoritmo no es capaz de realizar más agrupaciones. Para intentar paliar esta situación, en lugar de analizar la red al completo, voy a analizar las **comunidades existentes en la componente gigante**, que como hemos podido comprobar en la tabla de medidas reúne a un 26% de nodos. Por lo tanto, al ser una red más reducida y

conectada, el algoritmo podrá obtener mejores resultados. Dejando los parámetros por defecto es capaz de encontrar **solo 45 comunidades**, lo cual es una cifra mucho menor que la mostrada anteriormente con la red completa.

Si visualizamos el gráfico de la izquierda podemos observar como, a diferencia del anterior, no existe una agrupación tan masificada de comunidades con valores cercanos a 0. Sin embargo, la mayoría de ellas siguen teniendo menos de 50 miembros, exceptuando un par de ellas que superan la centena. Aunque hayamos reducido bastante el número de comunidades, sigue siendo un valor demasiado elevado como para intentar graficarlo por lo que experimentando aún más con el valor de resolución he llegado a obtener un valor de **9 comunidades**. En este caso, como se puede observar en la gráfica de la derecha, la mayoría supera la centena de miembros, lo que da lugar a comunidades más grandes y más fácilmente representables de forma gráfica. Esta se visualizará en la siguiente sección donde se estudiará más profundamente el significado de esta característica.

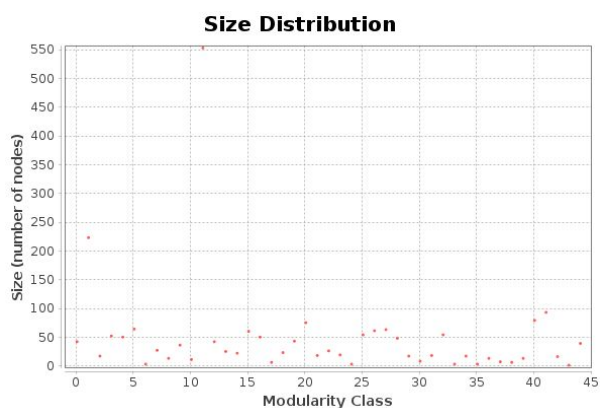


Figura 8. Distribución del tamaño de comunidades en la componente gigante con resolución = 1.

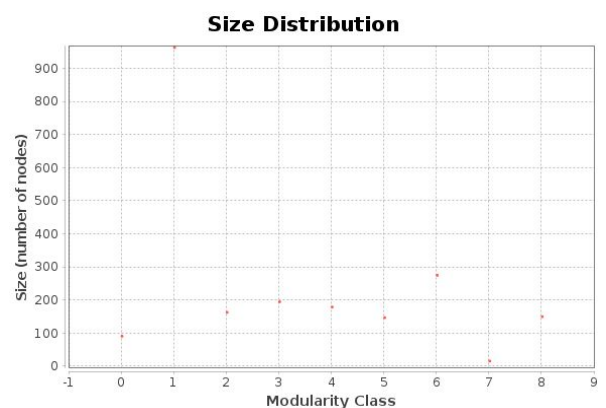


Figura 9. Distribución del tamaño de comunidades en la componente gigante con resolución = 20.

Un segundo enfoque consiste en seleccionar aquellos nodos con un mayor número de conexiones de modo que obtengamos una subred con los usuarios más participativos. Para ello vamos a aplicar el algoritmo **K-core**. Tras realizar diversos experimentos modificando tanto el grado mínimo en el algoritmo como la resolución, he podido observar que al establecer  $k=3$  he llegado a obtener un mínimo de 272 comunidades con 2.069 nodos y 5.395 aristas, mientras que con  $k=4$  existían 84 comunidades con 605 nodos y 2.015 conexiones. Como podemos observar, pese a que el grado mínimo establecido es bastante elevado comparado con el grado medio, el número de comunidades sigue siendo demasiado alto. Continuando con  $k=5$  con un mínimo de nodos obtenemos 27 comunidades. Como podemos observar en el gráfico siguiente la gran mayoría se encuentra entorno a cinco miembros por lo que, si además son dispersas, el algoritmo apenas puede hacer algo para agruparlas.

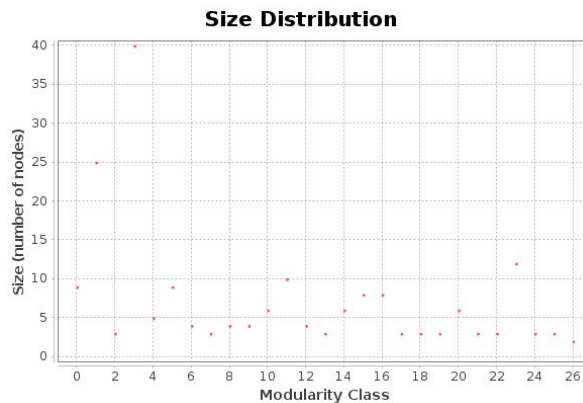


Figura 10. Distribución del tamaño de comunidades para la red podada con  $K$ -core ( $k=5$ ).

Con valores sucesivos los resultados son peores puesto que apenas quedan algunos nodos y si bien el número de comunidades es más bajo, estas se componen de apenas unos cuantos nodos por lo que creo que no merece la pena continuar este estudio. Las conclusiones que podemos extraer es que, debido a la escasa interacción entre los usuarios y al ser extremadamente local, el algoritmo detección de comunidades de Gephi no es capaz de agruparlos para conseguir un menor número de comunidades sin reducir exageradamente el número de nodos.

## 6. Visualización

En esta penúltima sección vamos a estudiar gráficamente algunas de las propiedades explicadas anteriormente de modo que nos ayude a obtener las conclusiones finales.

### 6.1. Centralidad intermedia

En esta primera gráfica podemos observar la red al completo en la que el **color** de los nodos está definido por el valor de la **centralidad intermedia**. Aquellos que son más azules son los que disponen de un menor valor, por lo tanto son nodos que apenas intervienen en algunos caminos. Estos usuarios no tienen ninguna capacidad de control de la información que se difunde.

Por otro lado, los nodos con tonos más claros, como el que podemos visualizar de color blanco en el lateral inferior izquierdo, disponen de una mayor influencia. Sin embargo, el que mayor capacidad de decisión tiene, en cuanto al flujo de los caminos mínimos en los que interviene, es el único nodo de color rojo. Su valor de centralidad intermedia es muy alto, por lo que interviene en una gran cantidad de caminos mínimos. Se podría decir que es uno de los nodos *hubs* más influyentes de la red. Por lo tanto, en el caso de mi red este nodo representa a un usuario que ha recibido y/o ha participado en muchos tweets. Si se elimina, los caminos en los que se encuentre como intermediario quedarían desconectados.

Además, si observamos el **tamaño** de los nodos dispuesto en función del **grado de los mismos**, podemos comprobar que el nodo de color rojo tiene un tamaño superior a la mayoría. Esto indica su alta participación, en comparación con el resto. Por tanto, el hecho

de estar enlazado a muchos usuarios le brinda la posibilidad de ejercer su influencia en los caminos mínimos en los que participa.

En la figura de la derecha podemos apreciar los usuarios más influyentes según esta medida. De nuevo, el código de colores y tamaños es el mismo que en la representación general. Aquellos con colores más cálidos son los más influyentes mientras que los de mayor tamaño son los más participativos. Por lo tanto, el usuario más influyente y participativo es [@ramblarosae](#) que se trata de una cuenta que convierte los tweets en *rosas algorítmicas*. El resto de usuarios con un alto valor de centralidad intermedia son periodistas, divulgadores tanto de España como de países hispanoamericanos, espacios de lectura que tuitean sus propuestas así como los propios colectivos y organizaciones fans de escritores.

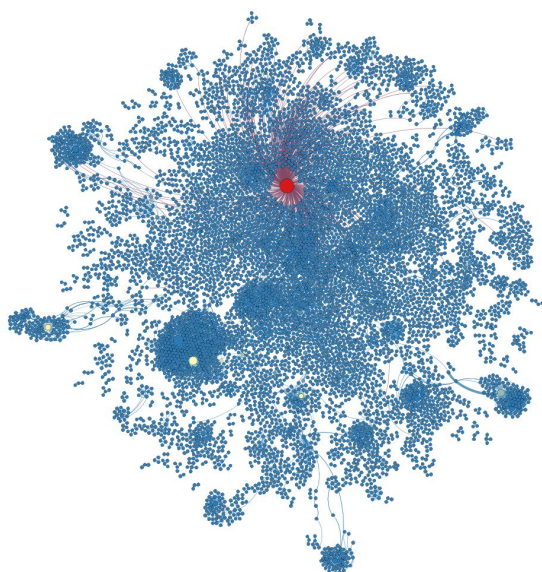


Figura 11. Gráfica de la centralidad de intermediación.

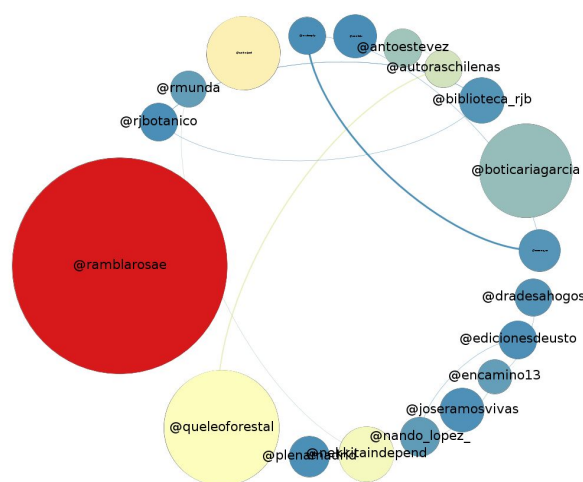


Figura 12. Usuarios más influyente según la centralidad de intermediación.

## 6.2. Centralidad de cercanía

Con el estudio de esta segunda medida el objetivo era conocer qué nodos se encuentran más cercanos al centro de la red. Generalmente, aquellos que sean más céntricos dispondrán de caminos más cortos para alcanzar cualquier nodo. Como podemos observar en el gráfico de la izquierda, el **tamaño** se encuentra, de nuevo, definido en función del **grado** mientras que el **color** representa el valor de la **centralidad de cercanía**. Por ende, los nodos azulados son los que tienen que recorrer una mayor distancia para alcanzar el centro de la red. Mientras que los nodos rojizos son los que están más cerca de esta parte central. Si apreciamos las distintas tonalidades de los nodos podemos visualizar nodos alejados que están situados en el centro de la red. Esto nos indica que la posición del nodo no está relacionada con su centricidad, puesto que el algoritmo *Force Atlas 2* no tiene en

cuenta esta cualidad. Lo mismo ocurre con algunos nodos rojizos situados en la periferia, los cuales están conectados con otros nodos centrales que les facilitan el acceso a la zona central de la red, donde se encuentra el meollo del asunto. Si observamos de forma general el gráfico, parece que existen un mayor número de nodos cercanos al centro que alejados. Esto acompaña al hecho de que la distancia media sea tan pequeña para alcanzar cualquier nodo así como el diámetro sea de tan solo seis nodos para alcanzar el más lejano.

En el gráfico de la derecha podemos visualizar los usuarios con un mayor valor de centralidad de cercanía, es decir, aquellos que se encuentran más cercanos al centro de la red. El código de colores y tamaños es el mismo, por lo que aquellos con tonos más rojizos son los nodos más céntricos, mientras que los de mayor tamaño disponen de un mayor número de conexiones por su alto grado.

Investigando acerca de las cuentas que aparecen en este segundo gráfico, he descubierto que entre algunas se encuentran periodistas, asociaciones y/o clubes de lectura, es decir, entidades y/o usuarios relacionados con este tema. Sin embargo, en su mayoría son usuarios habituales de Twitter sin conexión aparente además de otras asociaciones, por ejemplo, de animales. A diferencia de la medida anterior, en esta aparecen nodos que a simple vista no están relacionados con la temática, pero que sin embargo han interactuado de alguna forma con alguno de los tres *hashtags*. Es por ello por lo que parece que esta medida reúne a un grupo más diverso de usuarios, que en principio, no tienen por qué aportar información interesante para el estudio de esta red.

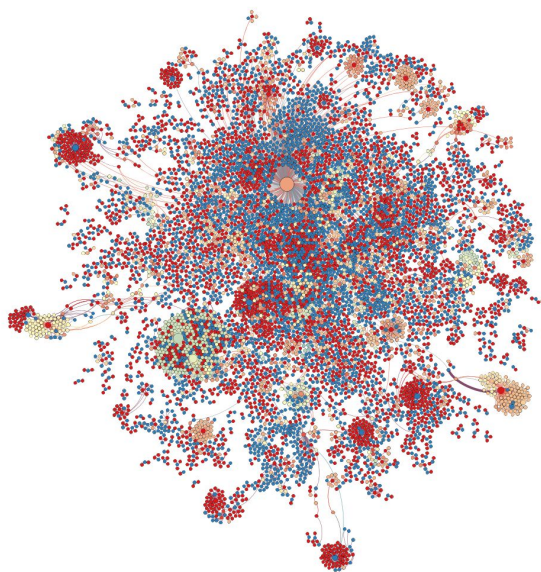


Figura 13. Gráfica de la centralidad de cercanía.

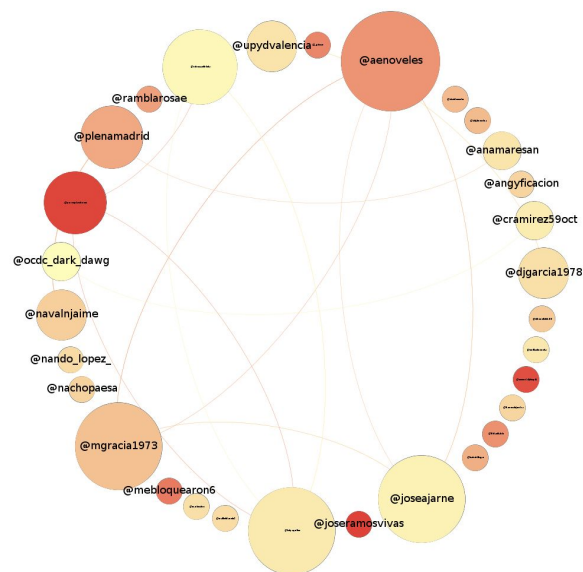


Figura 14. Usuarios más influyentes según la centralidad de cercanía.



### 6.3. Vector propio

Como en los dos casos anteriores, el **tamaño** de los nodos viene dado en función del **grado** de los mismos, solo que para esta sección el **color** se encuentra definido por el valor del **vector propio**. Este valor indica cuán importante es un nodo en función tanto de su influencia como la de los vecinos a los que se encuentra conectado. Los colores azules representan nodos nada relevantes, mientras que los más cercanos al color rojizo son los más influyentes.

En un principio podemos pensar que estos últimos pueden ser los *hubs*, puesto que ciertamente tienen mucha influencia sobre la red de sus seguidores. Sin embargo, como podemos apreciar, el *hub* de mayor tamaño que se encuentra casi en el centro del gráfico tiene un color azul. Esto indica que a pesar de su influencia, según esta medida el nodo pierde relevancia puesto que sus vecinos no son nada importantes. Por el contrario, los nodos más influyentes tienen un menor tamaño lo que indica que han recibido o participado en menor medida pero su poder de influencia es mayor, seguramente, por estar conectados a algunos de los *influencers* de esta red.

En el gráfico de la derecha se muestran estos usuarios más influyentes. El código de colores y tamaños se mantiene de modo que los más rojizos son los que más control pueden ejercer, mientras que los de mayor tamaño son los que disponen de un número mayor de conexiones con otros nodos de la red. De nuevo, investigando acerca de estas cuentas he podido descubrir que se tratan de asociaciones de cultura, como la *Unidad de Cultura Científica e Innovación (UCC+i) de la Universidad de Castilla-La Mancha*, divulgadores como [@boticariagarcia](#) que ya ha aparecido en anteriores gráficos, además de diversos usuarios habituales de Twitter, la mayoría de ellos de países hispanoamericanos.

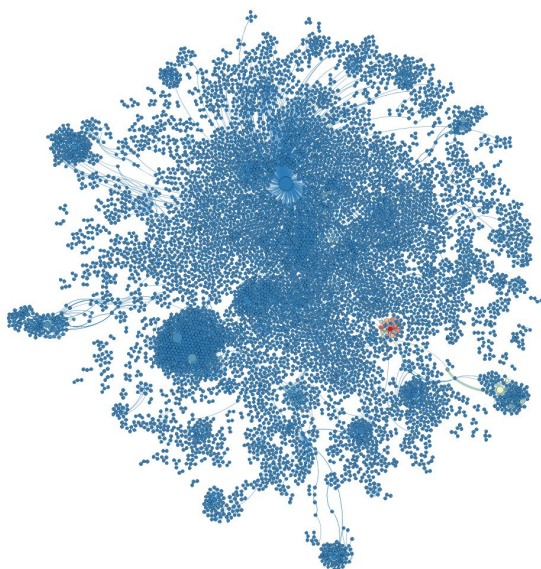


Figura 15. Gráfica del vector propio.

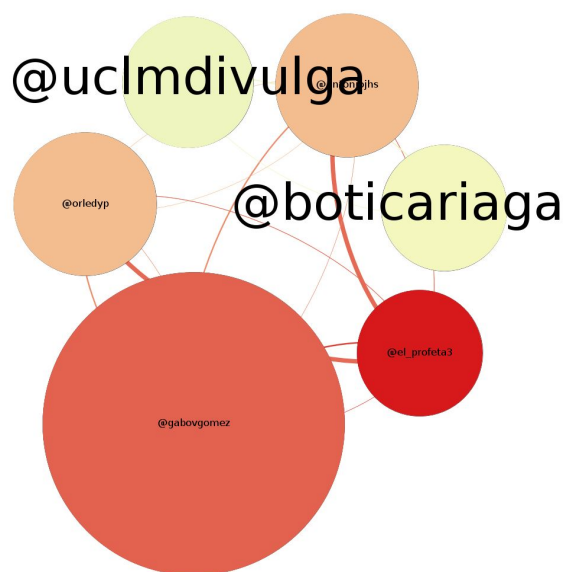


Figura 16. Usuarios más influyentes según el V.P.

#### 6.4. Comunidades de la red y subredes

En este último apartado de la visualización vamos a analizar el estudio de comunidades realizado anteriormente para conocer qué ha influido en los resultados. En primer lugar, destacamos la imposibilidad de representar todas las comunidades en la red al completo puesto que se encuentran demasiado dispersas y no ha sido posible obtener un número menor de 1.382 comunidades. Sin embargo, en el grafo de la izquierda se muestra la **comunidad más grande** que contiene un 26,41% (2.206 nodos) de los nodos y un 30,36% (4.184) de las aristas. El **tamaño** de los nodos viene representado por el **grado medio** de estos mientras que su **color** refleja el **coeficiente medio de clustering**. De este modo podemos analizar el número de conexiones que tienen y la conectividad local con sus vecinos.

Como podemos apreciar los nodos más oscuros son los que tienen una mayor conectividad local, como los dos grupos situados en los laterales del gráfico. Esto significa que son grupos de usuarios, que probablemente se conocen, y por ende están todos conectados entre sí y que interaccionan con sus tweets, como cualquier persona con sus amigos. Del mismo modo también podemos observar pequeños grupos también de color oscuro muy conectados entre sí. No obstante, se encuentran muy dispersos y la mayoría tiene muy pocos nodos por lo que el número de usuarios es bastante reducido. Por otro lado, existen una gran cantidad de nodos claros con una conectividad local bastante baja, entre ellos el nodo *hub* que se puede observar casi en el centro del gráfico. Estos tratan de conectar los distintos grupos que hemos detectado para reunirlos en una misma comunidad.



Si en esta primera comunidad los grupos se encuentran bastante dispersos, en la red general lo están aún más puesto que la **segunda comunidad más grande** solo llega a disponer de un 1,41% de nodos y 1,78% de aristas, la cual se puede ver en el gráfico de la derecha. En este caso, la comunidad está más agrupada y algunos de los nodos tienen una conectividad local bastante alta, los cuales serán usuarios que conocen a toda comunidad. Mientras que los nodos más claros son aquellos que no tienen muchas conexiones con el resto.

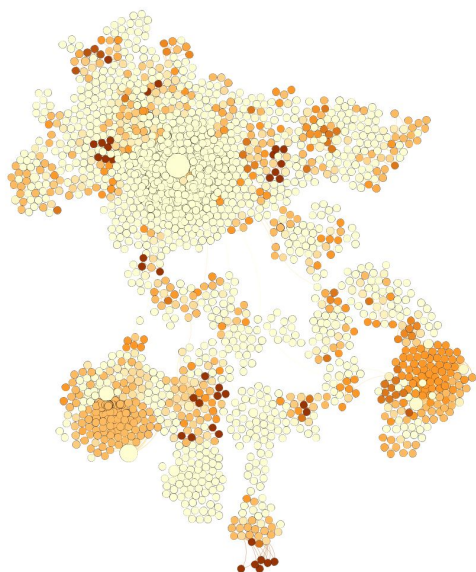


Figura 17. Primera comunidad más grande de la red al completo.

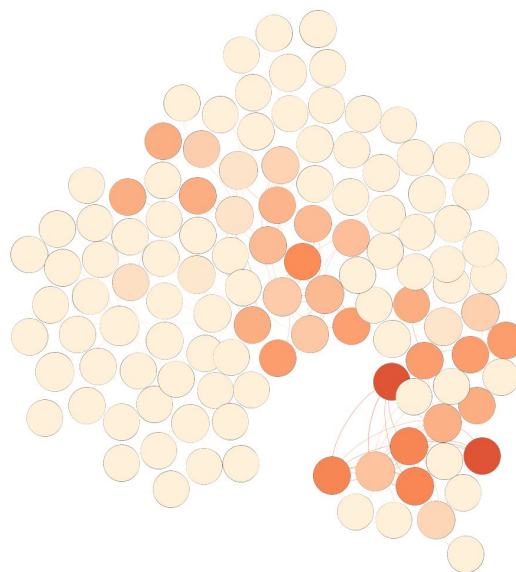


Figura 18. Segunda comunidad más grande de la red total.

Tras analizar las comunidades de la red total y no obtener buenos resultados, recordemos que realizamos el mismo procedimiento pero con la **componente gigante**. Como se explicó anteriormente, el objetivo de utilizar esta subred reside en que existe una mayor probabilidad de que los nodos se encuentren más conectados entre sí de modo que le facilitemos al algoritmo la obtención de un número representable de comunidades. Esta subred se puede visualizar en la siguiente gráfica, en la que el tamaño de nuevo está asociado al grado medio mientras que el color lo está al coeficiente medio de clustering.

Si bien está representada con otra paleta de colores, podemos apreciar grandes similitudes entre ella y la primera comunidad más grande de la red total visualizada en la figura 17. De hecho, **es exactamente la misma subred** si recordamos el porcentaje de nodos y aristas de la componente gigante (26,41% y 30,36%, respectivamente). Es por ello por lo que el análisis de comunidades sobre la componente gigante se convierte en un sub-análisis de comunidades sobre la comunidad más amplia de la red total.

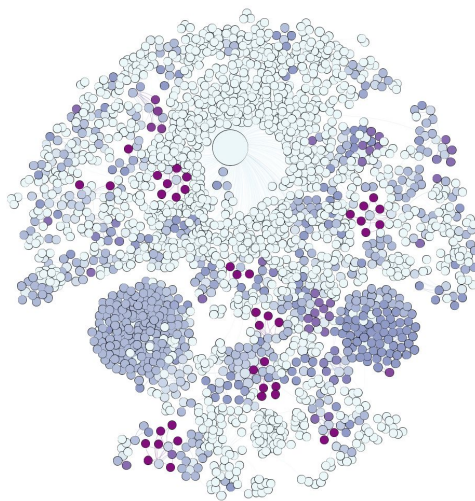


Figura 19. Gráfica de la componente gigante.

En la figura de la izquierda se representan las **9 comunidades** existentes dentro de la componente gigante. Como podemos observar, existe una componente mayor de color naranja que engloba a una gran parte de nodos. Este grupo de usuarios es el mayoritario con respecto a los otros. Asimismo, esta comunidad se cruza con otras tres (azul, gris y dorada) de modo que sus usuarios están conectados entre sí. Con un valor más alto de resolución, el algoritmo es capaz de agruparlas a las cuatro en una gran comunidad, como se muestra en la figura de la derecha. En este caso, además, ha sido capaz de agrupar también las comunidades de color cian y gris oscuro en una sola. Sin embargo, parte de esta última ha sido ubicada dentro de la comunidad fucsia junto con la amarilla que también se encontraba al lado de esta. Por lo tanto podemos concluir que varias de las comunidades originales están conectadas entre sí y esto permite agrupar a sus usuarios en una sola.

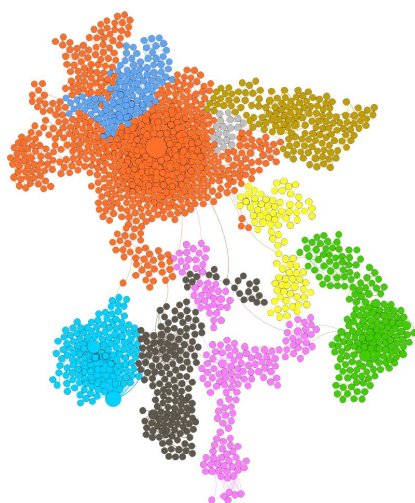


Figura 20. Nueve comunidades en la componente gigante.



Figura 21. Cuatro comunidades en la componente gigante.

## 6.5. Grado bidireccional, de entrada y salida

Como hemos podido observar a lo largo del análisis de esta red, una de sus principales características es su baja conectividad. Por ello, en esta última subsección vamos a estudiar en profundidad el número de nodos aislados y conexiones con pocos miembros, tanto de entrada, de salida como en ambas direcciones, por ser un grafo dirigido.

Comenzamos con la **distribución de grados bidireccional**, en la que hemos obtenido la siguiente subred compuesta únicamente por usuarios que han participado en el día del libro, tanto recibiendo como aportando interacción a la red. Para quedarnos solo con los nodos que disponen de al menos una conexión, hemos establecido un rango bidireccional con límite inferior en 1 para eliminar todos aquellos nodos que estén aislados. En total son **368 nodos y 739 aristas** (4,41% y 5,29%, respectivamente). Como podemos comprobar la cifra de usuarios y conexiones es altamente inferior comparada con la de la red total, por lo que podemos deducir que hay muy pocos usuarios participativos en ambas direcciones.

A continuación se muestra una gráfica que representa esta subred. Al utilizar el algoritmo *Radial Axis Layout* hemos conseguido, además, que los **agrupe por ejes** desde un grupo central en función del **grado medio** calculado anteriormente. Aquellos nodos con **color** verde más oscuro son los que disponen de una **centralidad de cercanía** mayor, por lo tanto se encuentran más cercanos al centro de la red. Mientras que los más claros y/o rosáceos, son los más alejados. Con esto podemos situar dónde se encuentra cada nodo dentro de la red global. Como podemos observar, la mayoría de los usuarios son bastante céntricos por lo que esta razón puede ser la que explique el por qué tienen tanto conexiones de entrada como de salida, ya que tanto son más alcanzables para otros usuarios además de disponer de una mayor facilidad para interaccionar con otros.

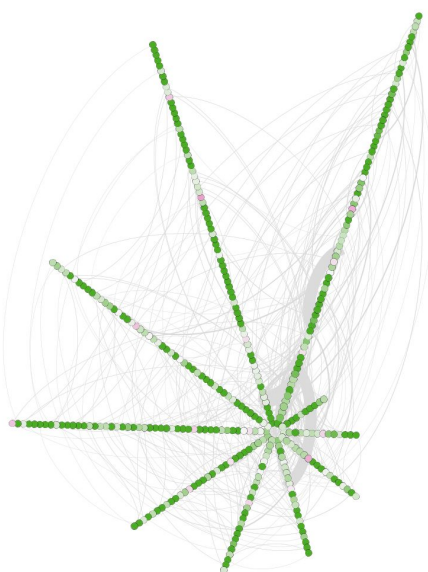


Figura 22. Gráfico de grados bidireccionales.

Para finalizar vamos a estudiar los gráficos asociados a los **grados de entrada y salida**, con el objetivo de comprobar cuáles son los actores con mayor número de interacciones recibidas y realizadas, así como la importancia de cada uno de ellos dentro de los caminos en los que intervienen. Comenzamos estableciendo un rango de grados de entrada con un valor de **tres o más conexiones de entrada**. El gráfico de la izquierda representa la situación general de estos nodos. Aquellos de **color azul oscuro** son los más influyentes en sus caminos mínimos según la medida de **centralidad de intermediación**, mientras que los más claros son los que menos control ejercen. En cuanto a los **tamaños**, lo más grandes representan un **grado de entrada** mayor, mientras que los más pequeños son los que entran dentro del intervalo pero tienen un menor número de conexiones. Como podemos observar existen unos cuantos nodos azules con un tamaño más o menos mayor al del resto de nodos, que nos indica que son usuarios influyentes y que han recibido un número de interacciones mayor al del resto. Sin embargo, también podemos visualizar un nodo en particular de un azul más oscuro pero de menor tamaño, situado aproximadamente en el centro de la red. Este es más *influencer* que los mencionados anteriormente pero no ha recibido tanta interacción por parte de otros usuarios.

En el gráfico de la derecha podemos visualizar los nodos más influyentes según la medida de cercanía de intermediación. El nodo azul oscuro que anteriormente he mencionado es [@ramblarosae](#), el cual también fue un usuario destacado al analizar la misma medida pero para la red al completo. Con este análisis complementario conocemos que si bien es el más influyente, no es el que ha recibido la mayor cantidad de interacciones por parte de los demás usuarios. El resto de *influencers* coloreados en tonos azules más claritos también los hemos visualizado anteriormente, por lo que el rango de personas influyentes no cambia. Lo que sí conocemos es que la interacción que han recibido en sus tweets por el resto de los usuarios es mayor a la media.

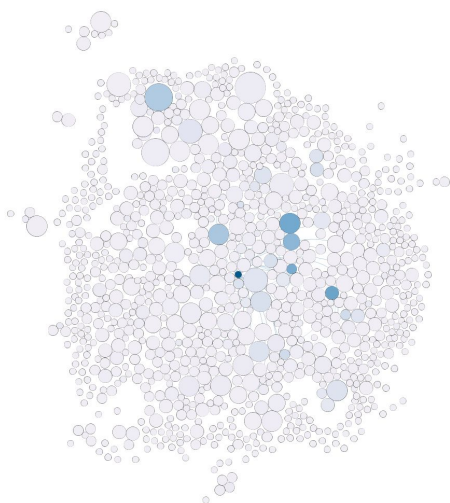


Figura 23. Gráfico de grados de entrada.

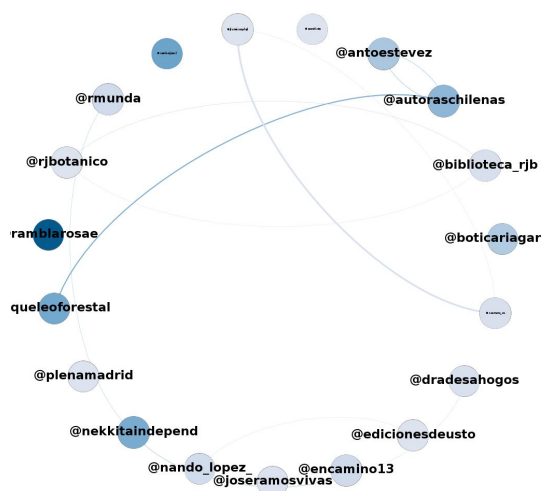


Figura 24. Usuarios relevantes según la centralidad de intermediación y el grado de entrada.



Por otro lado, en el siguiente gráfico de la izquierda se muestra el rango de nodos con un número de **conexiones salientes iguales o mayores a tres**. De igual modo, el código de colores y tamaños no varía solo que en este caso este último atributo se encuentra en función del grado de salida. Como podemos apreciar, en este caso el número de usuarios más participativos desciende considerablemente si lo comparamos con la interacción que reciben. Exceptuando un nodo de color azul oscuro claramente más diferenciado que el resto. Este usuario, además de interactuar bastante con los tweets de otros, es muy influyente según, de nuevo, la centralidad de intermediación. Este indicio significa que puede conocer su posición y está ejerciendo un control de información acerca de qué tweets retuitea, menciona y/o cita para que sus seguidores los vean.

Mediante el gráfico de la derecha podemos conocer quién es el usuario destacado anteriormente. Si lo observamos, podemos reconocer que son los mismos usuarios obtenidos con los grados de entrada, y de nuevo, el más participativo es [@ramblarosae](#). Por lo tanto, esta asociación es la más influyente de la red y la que más participación y participativa ha estado en este tópico de Twitter.

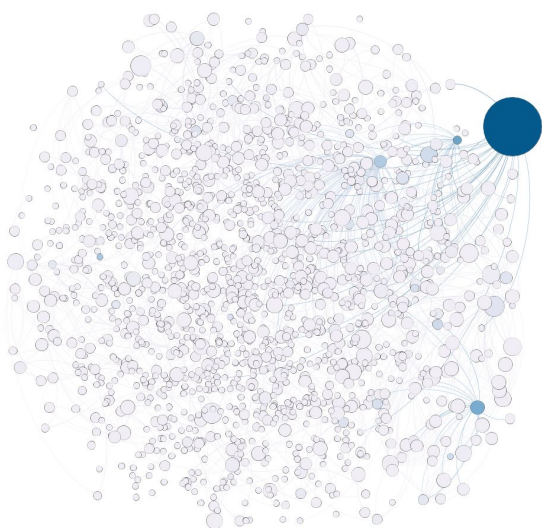


Figura 25. Gráfico de grados de salida.

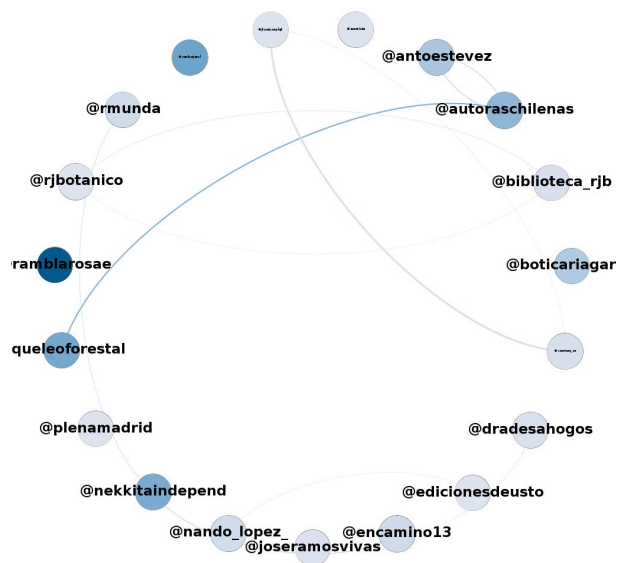


Figura 26. Usuarios relevantes según la centralidad de intermediación y grado de salida

## Conclusiones

Como hemos podido comprobar a lo largo del documento, la red obtenida con los tres *hashtags* definidos y asociados al día del libro dispone de unas características particulares.

- En primer lugar destacamos la escasa participación de la mayoría de usuarios en esta temática tanto a la hora de interactuar con otros tweets como al recibirla en

los suyos propios. Parece ser que estos *hashtags* han sido utilizados para celebrar de forma puntual el día del libro pero no existen demasiadas discusiones, con lo que la participación es bastante escasa.

- La conectividad es bastante local, lo que indica que los grupos de usuarios están compuestos por pocos miembros. Esto nos indica que en esta temática la mayoría de interacciones realizadas se encuentra entre los conocidos y amigos de la red.
- La cualidad anterior da lugar a que las comunidades sean muy dispersas de modo que es necesario reducir la red para encontrar un número fácilmente representable. Otro hecho a destacar es que la mayoría de ellas se encuentran mezcladas entre sí, lo que da lugar a una agrupación más fuerte en la que se consiguen un menor número de comunidades más pobladas.
- Pese a la baja densidad y número de conexiones, tanto el camino mínimo como el que conecta con el nodo más alejado no son nada amplios, lo que nos indica que existen usuarios *hubs* determinantes a la hora de conectar varios grupos de nodos. Si se eliminasen, surgirían muchos nodos aislados y los valores de las dos propiedades anteriores disminuirían aún más.
- Uno de los usuarios más influyentes según la centralidad de intermediación y que más participación ha recibido y realizado en la red completa es [@ramblarosae](#), una asociación que traduce los tweets dedicados al día de Sant Jordi, que está directamente relacionado con la temática del día del libro, para convertirlos en rosas algorítmicas. Este usuario es capaz de controlar el flujo de información en todos los caminos en los que participa de modo que puede manipular los datos que le llegan a una parte y a otra.

## Bibliografía

1. Stackoverflow, *What is the maximum number of edges in a directed graph with  $n$  nodes?*, <https://stackoverflow.com/questions/5058406/what-is-the-maximum-number-of-edges-in-a-directed-graph-with-n-nodes>
2. Stackoverflow, *What is the definition of the density of a graph?*, <https://math.stackexchange.com/questions/1526372/what-is-the-definition-of-the-density-of-a-graph>
3. Documentación sobre la función *FRECUENCIA*, <https://support.google.com/docs/answer/3094286?hl=es>
4. Universidad de Waterloo, Sarah McTavish, *Network Graphing Archived Websistes With Gephi*, <https://cloud.archivesunleashed.org/derivatives/gephi>