

# **Minería de medios sociales: Práctica Bloque I.1- Análisis y Visualización Básica de una Red Social con Gephi**

Máster en ciencia de datos e ingeniería de computadores - UGR

*11-5-2018*

**M<sup>a</sup> Cristina Heredia Gómez**

crstnheredia@correo.ugr.es

## Índice

<b>Descripción de la red</b>	<b>3</b>
Componentes y estructura . . . . .	3
Cálculo de medidas . . . . .	4
Gráficos . . . . .	5
<b>Análisis de la red (en función de sección 2.1)</b>	<b>6</b>
Grados internos y externos . . . . .	6
Diámetro de la red . . . . .	7
Conectividad . . . . .	7
Coeficiente medio de Clustering . . . . .	8
<b>Análisis de la centralidad de los actores en la red (sección 2.2)</b>	<b>8</b>
<b>Estudio de comunidades(sección 2.3)</b>	<b>10</b>
<b>Visualizaciones y gráficos adicionales</b>	<b>10</b>

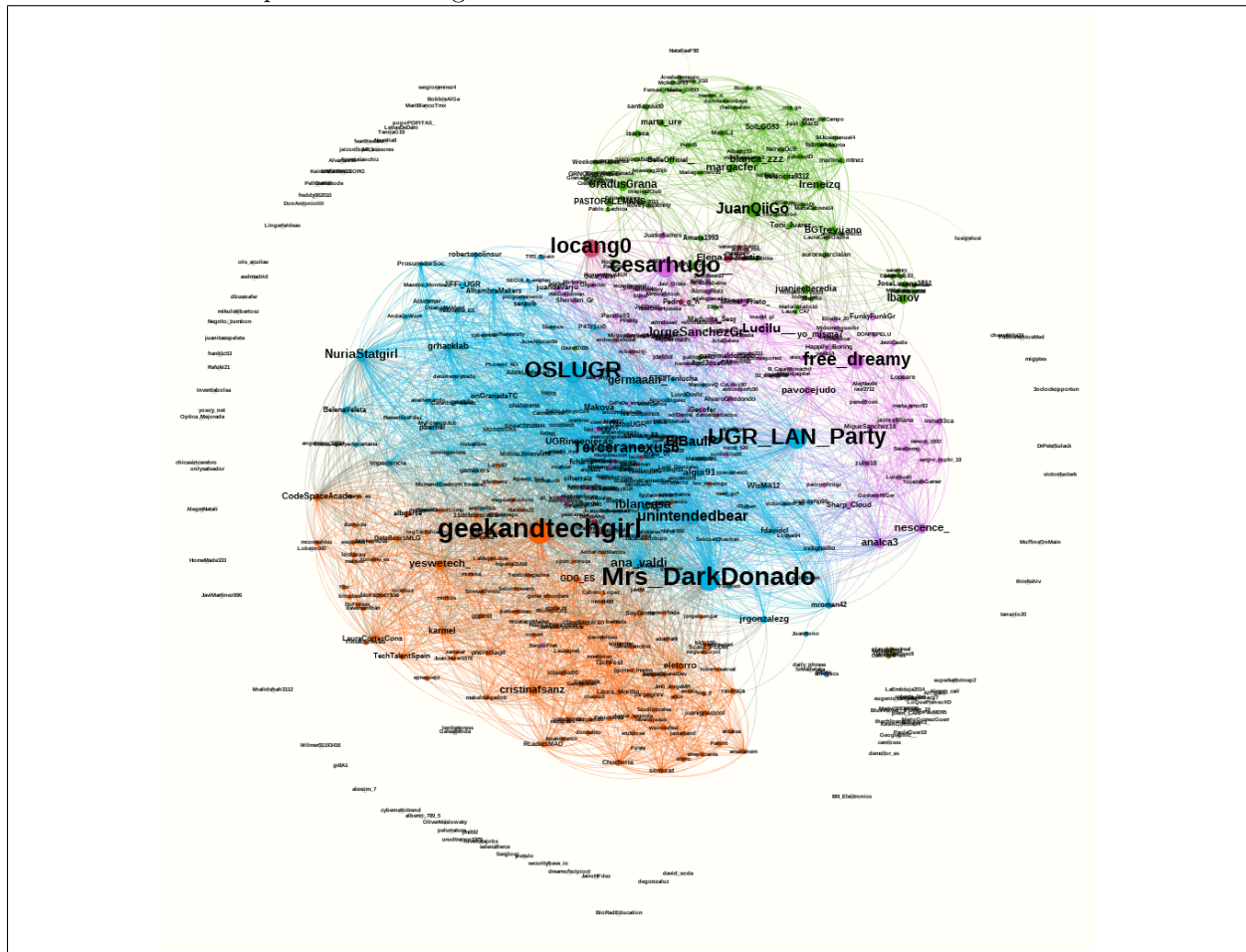
## Descripción de la red

La red escogida para el análisis es mi red personal de twitter, que contempla tanto a mis seguidores como a los amigos de mis seguidores. Las propiedades de la red se reflejan en la siguiente tabla:

Medida	Valor
Número de nodos N	574
Número de enlaces L	5781
Número máximo de enlaces $L_{max}$	328902
Densidad del grafo $L/L_{max}$	0.018
Grado medio $\langle k \rangle$	10.071
Diámetro $d_{max}$	10
Distancia media d	3.319137834962768
Coefficiente medio de clustering $\langle C \rangle$	0.309
Número de componentes conexas	92 weak, 155 strong
Número de nodos componente gigante (y %)	0.8293
Número de aristas componente gigante (y %)	0.9981

## Componentes y estructura

La red es una red dirigida con 574 nodos y 5781 enlaces obtenida con la herramienta en python [twecol](#). La red con todas sus componentes es la siguiente:



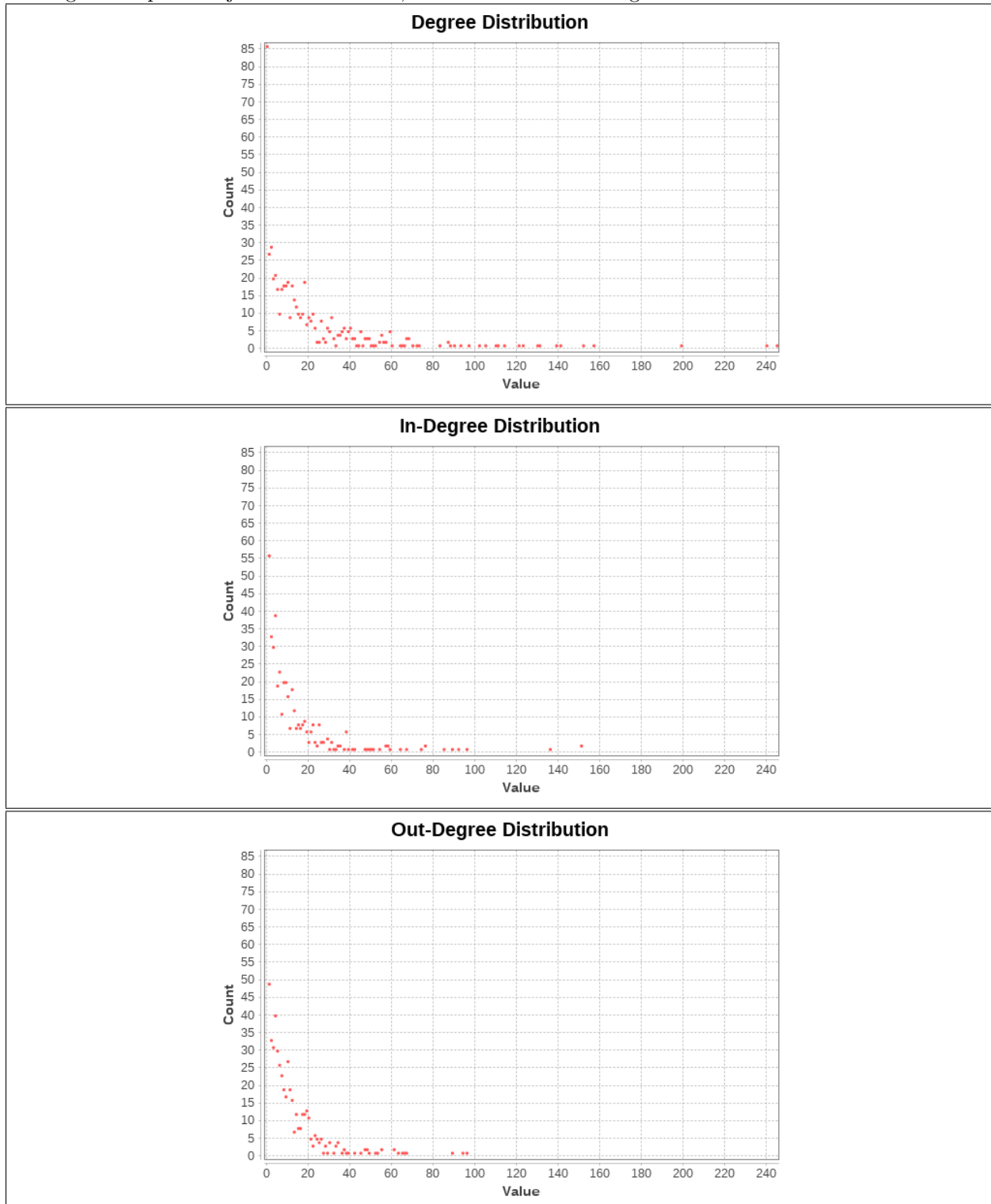
Además de las medidas mencionadas al comienzo de la sección, también se emplearon otras medidas como la **centralidad de grado**, la **intermediación**, la **cercanía** y la medida de **vector propio**, obteniendo los siguientes resultados:

Centralidad de Grado	Intermediación	Cercanía	Vector propio
Mrs_DarkDonado: 242	Geekandtechgirl: 0.09433	RoCioo_4: 1.0	Geekandtechgirl: 1.0
Geekandtechgirl: 238	Mrs_DarkDonado: 0.09009	CaerAtusPies: 1.0	Mrs_DarkDonado: 0.97515
OSLUGR: 198	OSLUGR: 0.07871	DaniiMartinez8: 1.0	OSLUGR: 0.89142
Terceranexus6: 157	UGR_LAN_Party: 0.07845	PanyPm: 1.0	Terceranexus6: 0.78381
Unintendedbear: 152	_cesarhugo.: 0.07232	Gafastienda: 1.0	ana_valdi: 0.77785

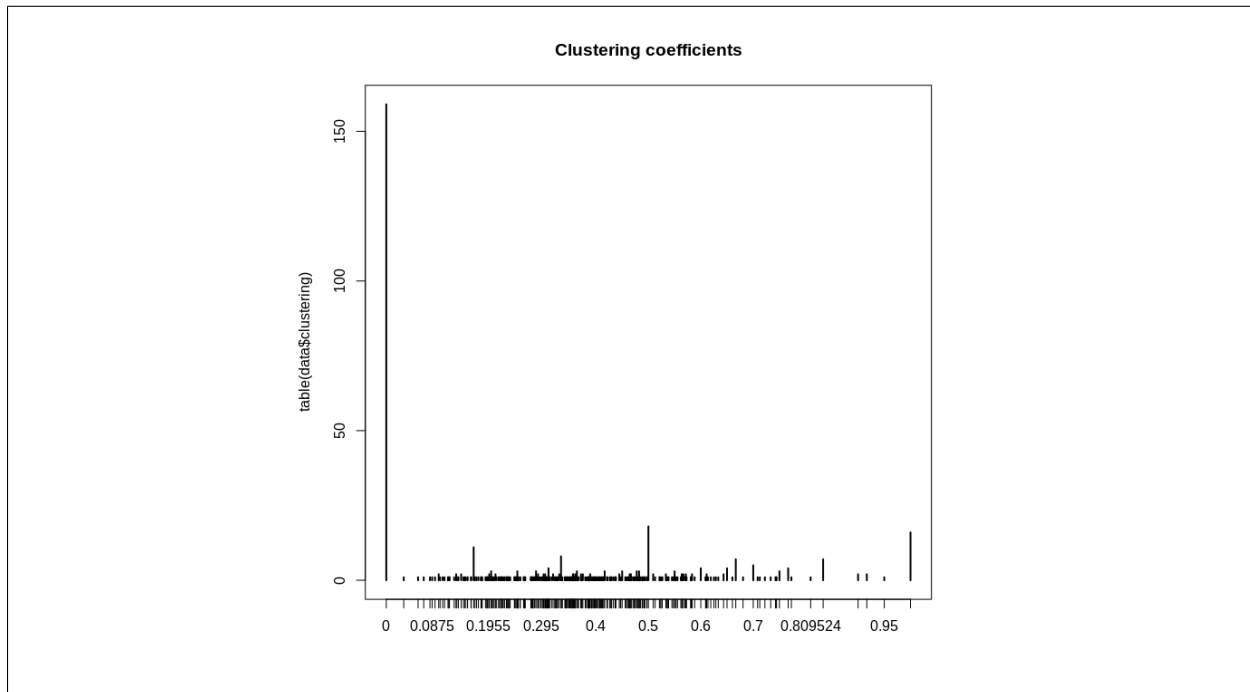
Page 4 of 11

## Gráficos

Junto con el cálculo de las medidas mencionadas en la sección **Descripción de la red** se han generado varios gráficos que se adjuntan en el Excell, sobre distribuciones de grado:



y sobre coeficiente de clustering:

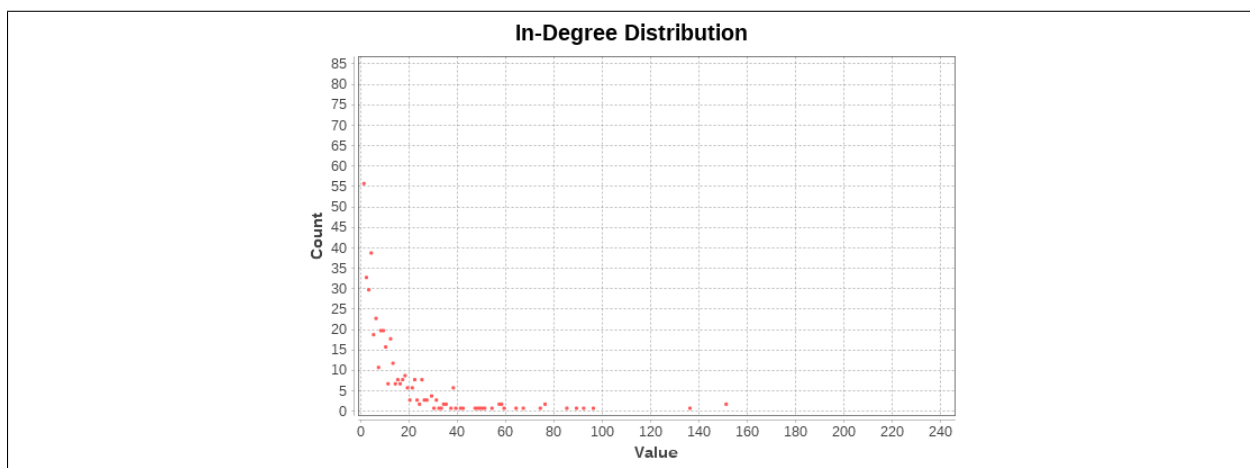


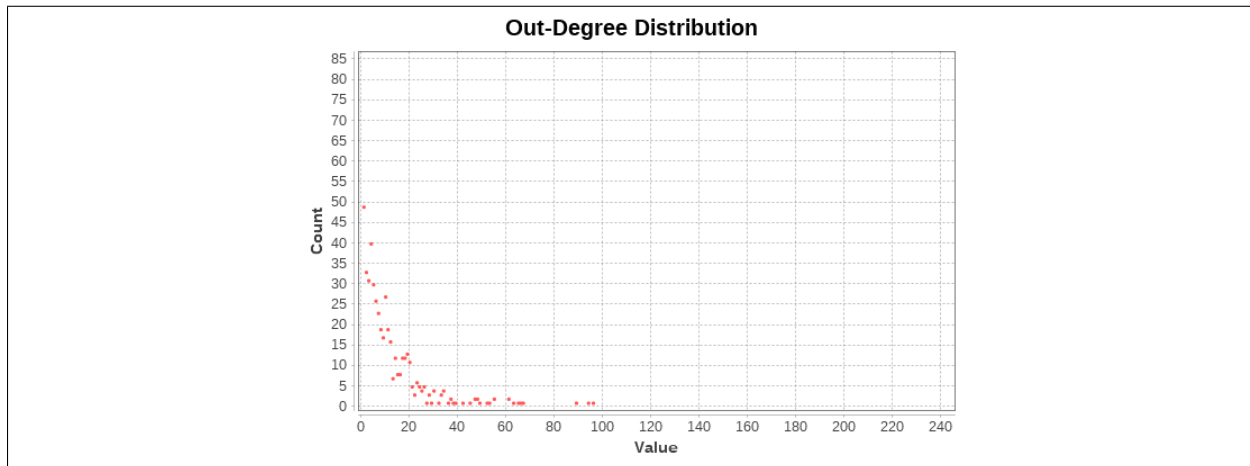
que serán analizadas más ampliamente en la sección siguiente del **Análisis de la red**.

## Análisis de la red (en función de sección 2.1)

En esta sección se realiza un análisis básico de la red, que contempla el cálculo de medidas como número máximo de enlaces, diámetro de la red, coeficiente de clustering... etc (en definitiva, la tabla de la página 3). La red tiene  $N = 574$  nodos y  $L = 5781$  enlaces, y por lo tanto un número máximo de enlaces ( $L_{max}$ ) =  $N \cdot (N - 1) = 328902$ . La densidad del grafo se calcula como  $\frac{L}{L_{max}}$  y es igual a 0.018. El grado medio de la red figura 1 pág 5 es de 10.071, lo que denota que, en media un actor está relacionado con unos 10 actores de la red, por lo que se puede decir que la interacción en media es alta. En la gráfica se observa que hay algunos hubs con más de 200 conexiones, en concreto hay uno con 250 conexiones.

## Grados internos y externos





Como el grafo es dirigido, tenemos una gráfica de cómo se distribuyen los grados internos y grados externos. En la primera se observa que los enlaces que conectan a un nodo con otros nodos de su subgrafo oscila también en torno a 10, y que hay unos pocos hubs con más de 100 enlaces internos. En el segundo grafo, observamos que la media de enlaces del nodo con el exterior de su subgrafo (el resto de la red) está también en torno al 10, pero esta vez hay menos hubs y todos tienen menos de 100 conexiones con el exterior. Como casi siempre en redes sociales reales, observamos que hay pocos nodos fuertemente conectados, lo que indica que se cumple la propiedad libre de escala.

## Diámetro de la red

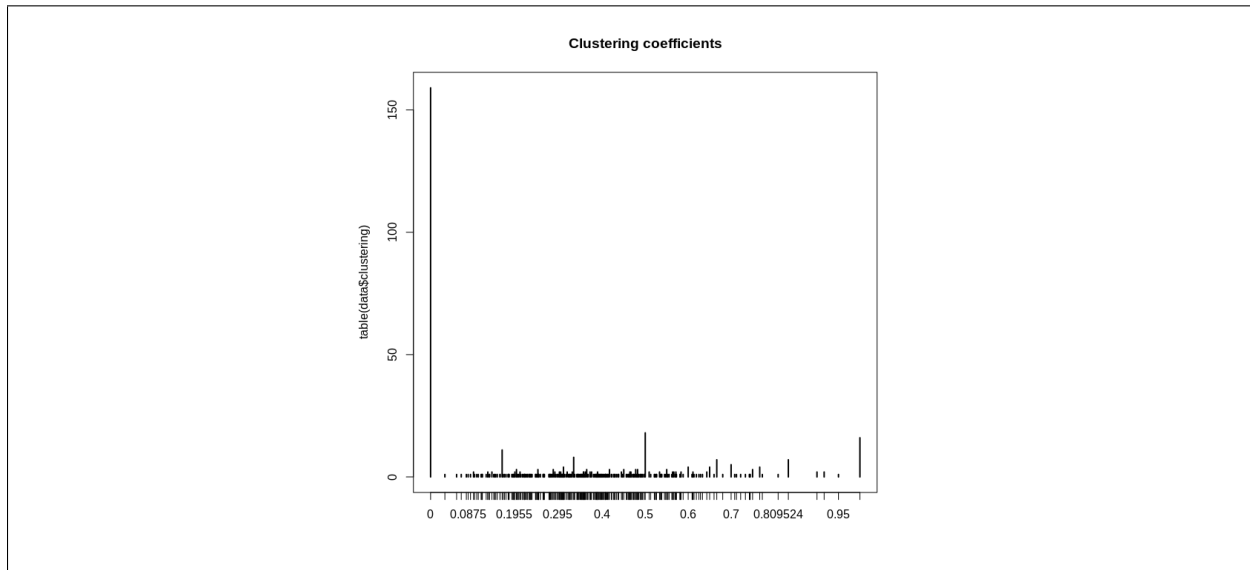
El diámetro mide la distancia máxima entre los dos puntos más lejanos de la red. En este caso, obtenemos que el diámetro máximo de la red original es 10, y que la distancia media es de 3.319137834962768, que se puede decir que es poca, si tenemos en cuenta que es una red con 574 nodos y 5781 enlaces. Por lo tanto se cumple la propiedad de mundos pequeños, dado que la distancia entre muchos nodos es aproximadamente 3.31913...

Tras filtrar nodos con menos de dos enlaces, de nuevo obtenemos un diámetro de 10 y una distancia media de 3.2808523562887655, muy parecida a la obtenida antes de filtrar. Además, con el cálculo del diámetro Gephi nos proporciona gráficas de intermediación, cercanía, y excentricidad, que discutiremos en el apartado siguiente de **Análisis de centralidad de actores en la red**.

## Conectividad

La red tiene 247 componentes conexas, de las cuales 92 son débiles (el grado interno total del grafo es mayor que el grado externo total) y 155 son fuertes (cada nodo tiene más enlaces dentro de la componente que fuera de ésta). Hay muchos nodos aislados o conectados a otro nodo aislado por un único enlace, por lo que los nodos de la componente gigante son el 82.93 % del total de nodos sin embargo el número de enlaces de la componente gigante es el 99.81 % (los enlaces de mantienen casi todos).

## Coeficiente medio de Clustering



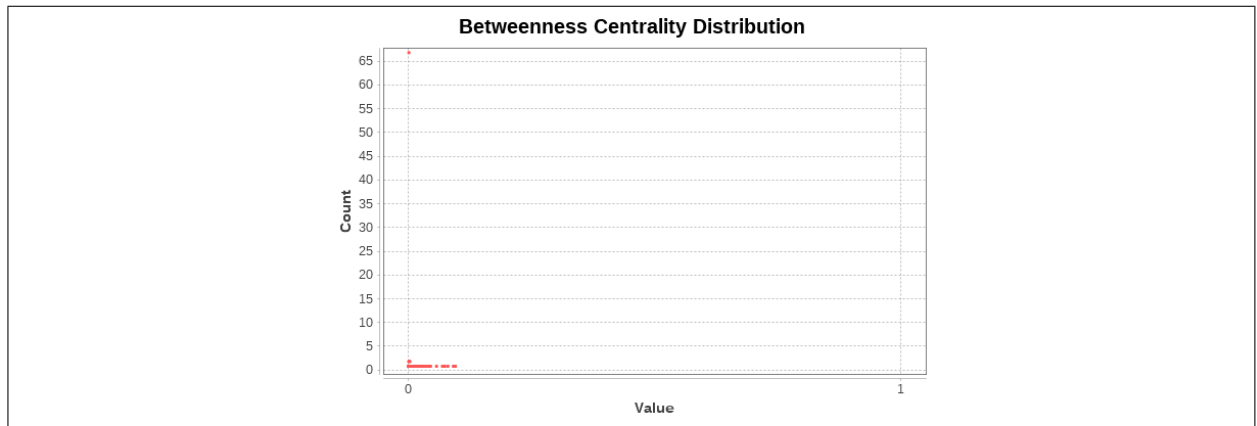
El coeficiente medio de clustering indica si el grado de clustering local que se aprecia en la red es o no significativo. En este caso, el coeficiente medio obtenido es de 0.309, que es muy alto, lo que indica que sí se produce mucho clustering local. En el gráfico se observa un despunte en la primera barra que indica como el clustering local es muchísimo más alto en nodos con pocas conexiones, ya que por la jerarquía de redes, estos nodos se sitúan en vecindarios localmente densos.

## Análisis de la centralidad de los actores en la red (sección 2.2)

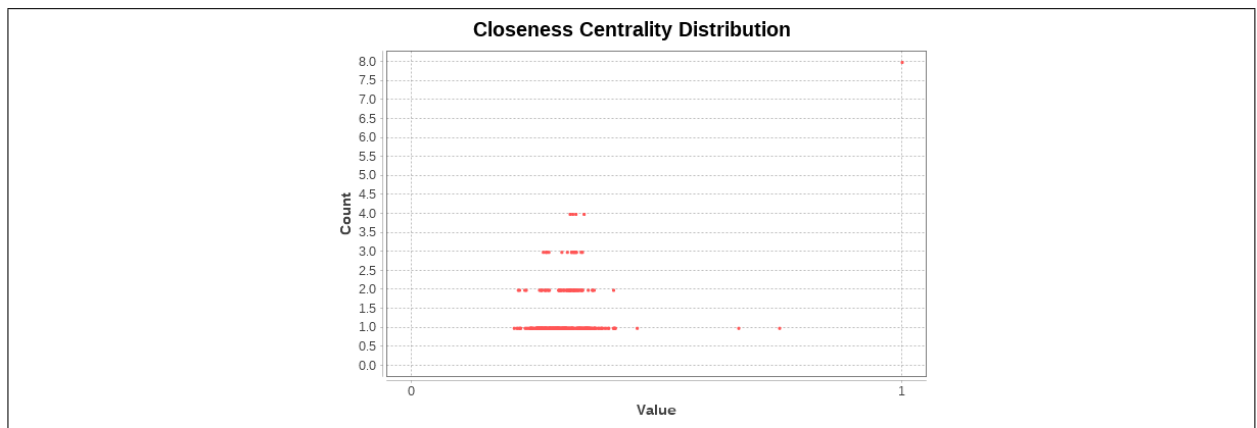
Como se comentaba en el apartado anterior, junto con el cálculo del diámetro de la red Gephi genera tres gráficos más: el de intermediación, el de cercanía y el de excentricidad. Para medir la centralidad de los actores en la red se han utilizado 5 métricas: la centralidad de grado, la intermediación la cercanía y el vector propio, extrayendo los 5 actores principales para las 5 medidas como se muestra en la tabla de la página 4.

- **Centralidad de Grado:** Mide el grado ( $n^\circ$  de enlaces) de un actor en la red. Según esta medida, el actor con más enlaces es Marta (Mrs\_DarkDonado) con 242 enlaces, seguida por la comunidad Geekandtechgirl con 238, la OSL con 198, Terceranexus6 con 157 y Unintendedbear con 152
- **Intermediación:** Esta medida mide como de frecuentemente aparece un nodo en el camino más corto para ir a otros nodos. En este caso, los 5 actores con mayor intermediación en la red son Geekandtechgirl, Mrs\_DarkDonado, la OSL, UGR\_LAN\_Party y \_cesarhugo\_. En el siguiente gráfico podemos ver como la mayoría de los nodos no tienen intermediación, mientras que hay unos pocos que tienen valores más altos:

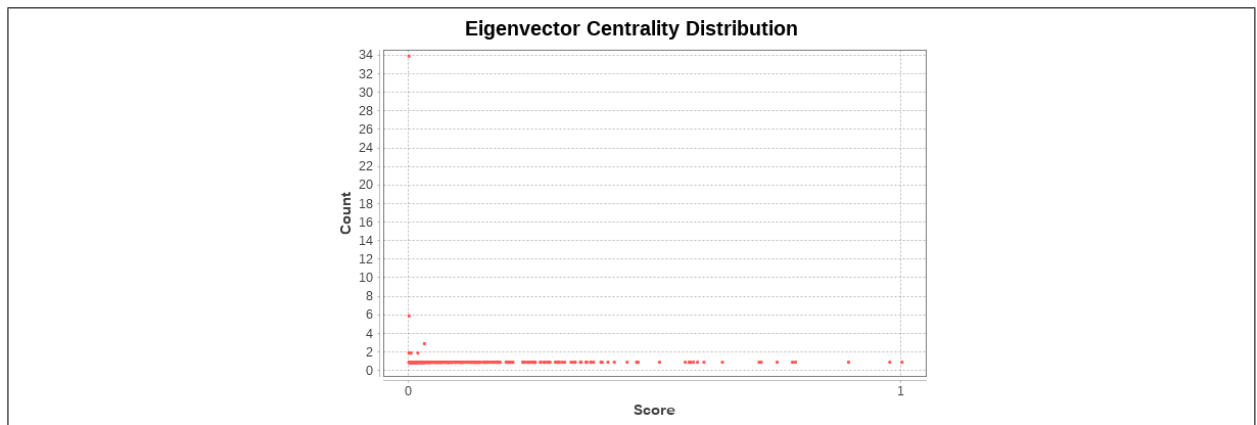




- **Cercanía:** Mide la distancia media de un nodo a todos los demás nodos de la red. En este caso tenemos que los nodos con mayor cercanía en la red son: RoCio\_4, CaerAtusPies, DaniMartinez8, PanyPm y Gafastienda, todos con un valor de 1. En la siguiente gráfica podemos observar como en general las cercanías varían para los distintos actores, y que, sin embargo, hay 8 actores con el máximo valor de cercanía.



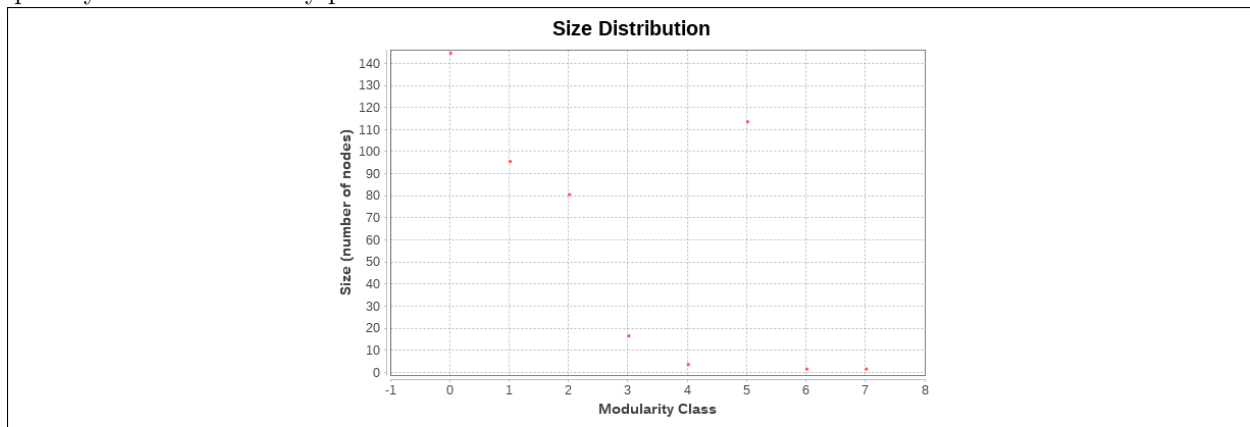
- **Vector propio:** mide la importancia de un nodo en base a la prominencia (cómo de centrales son sus vecinos), por lo que, a diferencia de la centralidad de grado, no todas las conexiones importan igual. En esta red los 5 actores con mayor vector propio son por orden: Geekandtechgirl, Mrs.DarkDonado, OSLUGR, Terceranexus6 y ana\_valdi. En el gráfico podemos observar como la mayoría de los actores tienen un autovalor inferior a 0.5, incluso hay 34 actores con un autovalor=0, mientras que hay unos pocos actores con autovalores próximos a 1.



## Estudio de comunidades(sección 2.3)

Para el análisis de comunidades se ha utilizado el algoritmo de Lovaina disponible en Gephi. El algoritmo de Lovaina se basa en la modularidad de la red, que es una medida que comprueba si la cantidad de enlaces entre comunidades es menor que el de una red aleatoria. La modularidad toma un valor alto cuando la red tiene más enlaces hacia el interior de la comunidad que hacia el exterior.

En la imagen siguiente podemos observar las 8 comunidades detectadas, así como el número de nodos de cada comunidad, pudiendo observar que hay una comunidad muy grande con más de 140 nodos, mientras que hay otras dos con muy pocas conexiones.



En concreto, las comunidades detectadas son las siguientes, como se observa en la figura de la página 4:

- **Comunidad azul:** Es la comunidad más grande detectada, que engloba a entidades y gente que participan en comunidades dentro de la UGR como la OSL, Interferencias, GDG, etc.
- **Comunidad naranja:** La segunda comunidad de mayor tamaño, que representa una comunidad de mujeres tech con aprox. 115 nodos, formada por mujeres tech donde nuestra comunidad granadina Geek&TechGirls cobra importancia porque a través de ella conectan mujeres tech de Málaga, Madrid y otras partes de España con otras comunidades de la UGR.
- **Comunidad rosa:** Compuesta principalmente por gente del grado de informática y del doble grado.
- **Comunidad verde:** Formada por amigos y conocidos del instituto, el colegio y el pueblo.
- **Comunidad roja:** Comunidad pequeña compuesta por unos 16 nodos que representan egresados de la ETSIT de promociones mucho más anteriores a la mía que tienen relaciones principalmente con comunidades de la UGR y gente de allí, pero que algunos como *locang0* conectan su comunidad con geek&tech girls y la comunidad verde además de las de la URG.
- **Comunidad blanca:** Comunidad externa a la UGR y a Granada, de pequeño tamaño, formada por profesionales del mundo de la programación.
- **Comunidad azul marino:** Comunidad pequeña formada por cuentas de congresos sobre IA que se siguen entre sí.
- **Comunidad marrón:** Comunidad pequeña sin vínculos con actores de la componente grande, compuesta por gente que me sigue que yo no sigo pero que se siguen entre ellos.

## Visualizaciones y gráficos adicionales

En la primera gráfica se representa el tamaño de los nodos según su valor de intermediación (a más grandes mayor valor), y el color del nodo según su autovalor (a más oscuro más elevado):

