



UNIVERSIDAD DE GRANADA
MÁSTER DE CIENCIA DE DATOS E INGENIERÍA DE
COMPUTADORES
CURSO ACADÉMICO 2019-2020
MINERÍA DE MEDIOS SOCIALES

Análisis y Visualización Básica de una Red Social con Gephi.

Análisis y Visualización de la red de Facebook de figuras públicas.

Nicolás Cubero

11 de Mayo de 2020

Índice

Índice de figuras	2
1. Introducción	3
2. Análisis preliminar de la red	3
3. Reflexiones sobre Estadísticos básicos	6
4. Análisis de la Centralidad	7
5. Análisis de Comunidades	13

Índice de figuras

1.	Gráfico de la red de páginas figuras públicas de Facebook donde para cada nodo, su tamaño y color se ha determinado en función del número de páginas de otras figuras famosas con los que mantienen algún “ <i>like</i> ” mutuos en <i>Facebook</i> . Aquellos nodos más rojos y de mayor tamaño se corresponden con las páginas de mayor Grado, mientras que los nodos más claros y de menor tamaño constituyen las páginas de menor Grado.	4
2.	Gráfico de la distribución de grados de los nodos de la red, donde se puede apreciar que se ajusta a la regla de la potencia.	5
3.	Gráfico de la distribución de coeficientes de clustering de la red.	6
4.	Gráfico de la distribución de la Intermediación de la red.	8
5.	Gráfico de la distribución de la Cercanía de la red.	9
6.	Gráfico de la distribución de la Cercanía de la red.	9
7.	Gráfico de la red donde el tamaño de cada nodo y su coloración se han establecido en base a sus valores de Intermediación y centralidad de Vector Propio: Los nodos de mayor tamaño se corresponden con los nodos de mayor Intermediación, mientras que los nodos de color más violeta se corresponden con los nodos con mayor centralidad de vector propio.	11
8.	Distribución de los valores de Cercanía de la red frente a los valores de Grado	12
9.	Gráficos de densidad de las comunidades formadas en cada particionamiento obtenido con los diferentes valores de Resolución.	15
10.	Gráfico de la red donde se ha coloreado cada conjunto de nodos y aristas en la comunidad a la que han sido asignados considerando un valor de Resolución de 0.75.	16
11.	Gráfico de la distribución de densidad de las comunidades formadas en el particionamiento con un valor de Resolución de 5.	17
12.	Gráfico de la red donde se ha coloreado cada conjunto de nodos y aristas en la comunidad a la que han sido asignados considerando un valor de Resolución de 5.	18
13.	Gráfico de la distribución de densidad de las comunidades formadas en el particionamiento con un valor de Resolución de 5.	19

1. Introducción

En este proyecto se tratará el Análisis y Visualización de la red de *Facebook* de las páginas de figuras públicas [1], recogida en el año 2017 con la finalidad de poner en práctica todos los conocimientos adquiridos sobre **Análisis de Redes Complejas**.

La red tratada presenta 11565 nodos correspondientes a las páginas de *Facebook* de diferentes figuras públicas mundialmente conocidas, registrando el establecimiento de "like" mutuo, lo cual supone la existencia de interacción entre ambos personajes públicos. En concreto, la red registra más de 67.000 "likes" entre las páginas.

Primeramente, se realizará un **análisis preliminar de la red** apoyándonos en el cálculo de los estadísticos más básicos aplicables a redes complejas. Seguidamente, se entrará más a fondo estudiando la **Centralidad de la red** para tratar de determinar las páginas de figuras famosas más centrales.^{en} esta red. Para acabar, se realizará un estudio de comunidades para tratar de obtener el mejor particionamiento de la red.

2. Análisis preliminar de la red

Primeramente se muestra una imagen de la red donde se han destacado las páginas de mayor Grado, es decir, las que establecen un mayor número de "likes" con otras páginas de figuras famosas.

De este modo, en la red de la figura 1, los nodos más rojos y grandes se corresponden con las páginas de figuras famosas de mayor **Grado**, mientras que los nodos más claros y pequeños, representan el caso contrario, páginas con un menor número de "likes" mutuos a otras páginas:

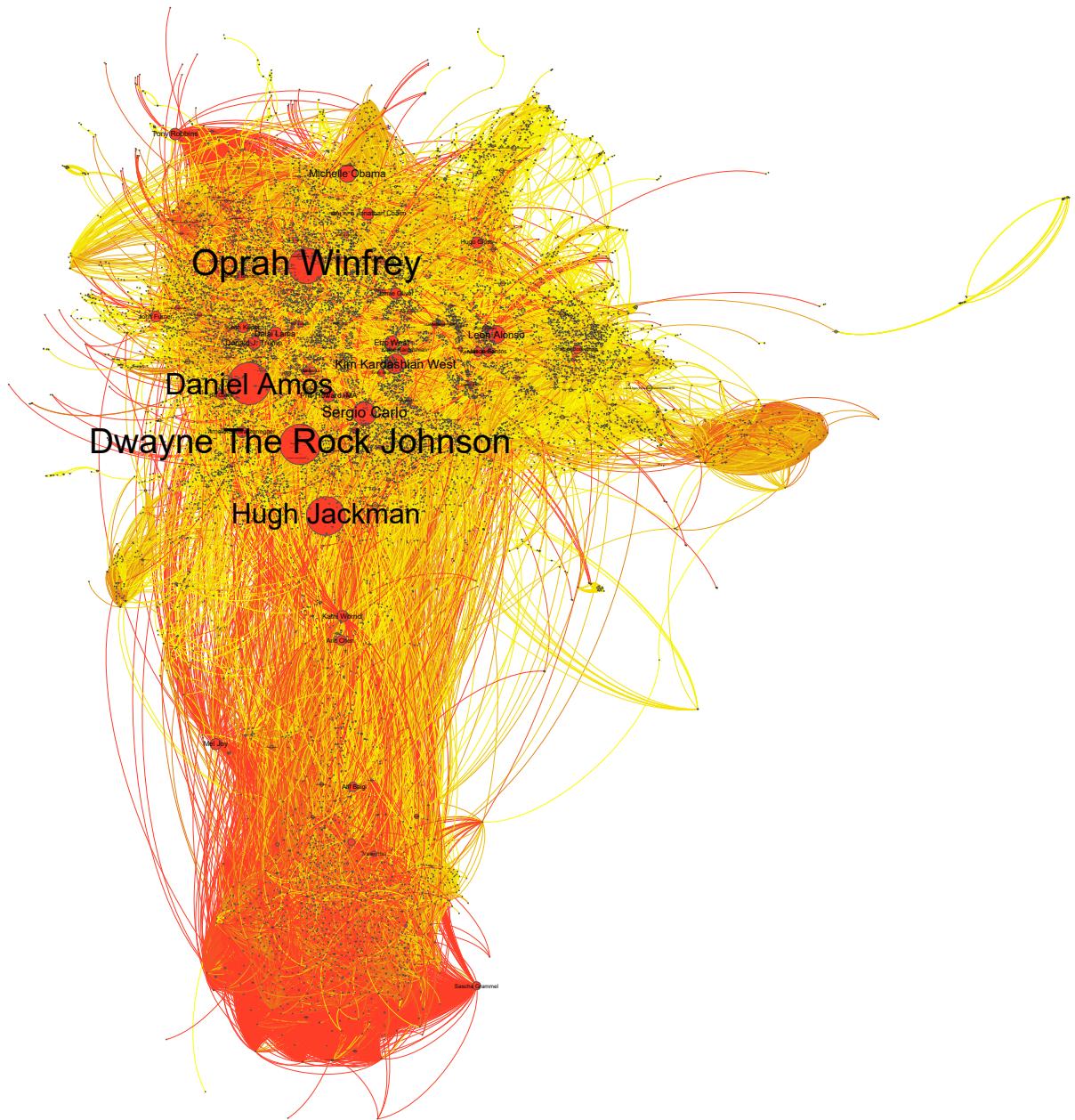


Figura 1: Gráfico de la red de páginas figuras públicas de Facebook donde para cada nodo, su tamaño y color se ha determinado en función del número de páginas de otras figuras famosas con los que mantienen algún “*like*” mutuos en *Facebook*. Aquellos nodos más rojos y de mayor tamaño se corresponden con las páginas de mayor Grado, mientras que los nodos más claros y de menor tamaño constituyen las páginas de menor Grado.

Procedemos a calcular las medidas más básicas de la red:

Estadístico	Valor
Número de nodos (N)	11565
Número de aristas (L)	67038
Número máximo de enlaces(L_{max})	67.038.000
Grado medio (k)	11.593
Densidad del grafo	0.001
Coeficiente medio de clustering C	0.215

Tabla 1: Estadísticos generales de la red

Estudiamos con mayor profundidad la distribución de grados de los nodos en esta red, la cual se representa en la siguiente figura:

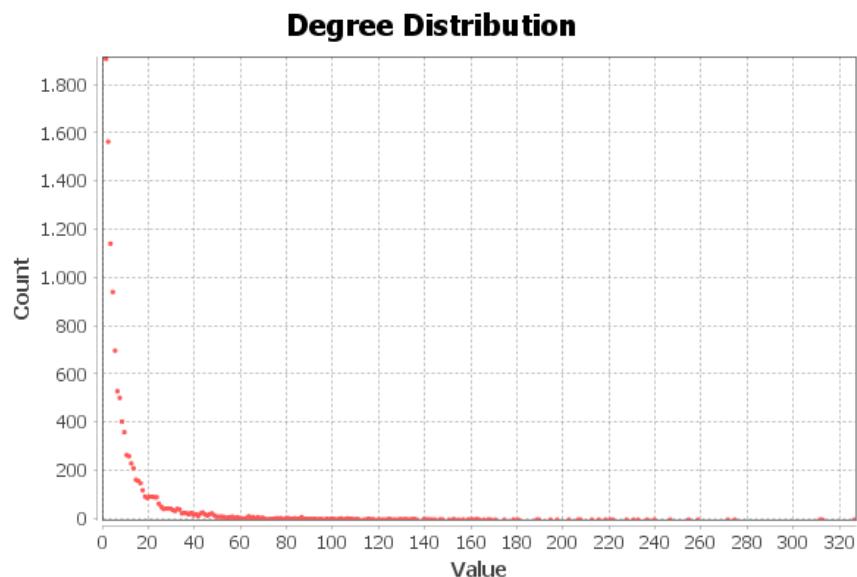


Figura 2: Gráfico de la distribución de grados de los nodos de la red, donde se puede apreciar que se ajusta a la regla de la potencia.

Por su parte, la distribución de los coeficientes de clustering para esta red es la siguiente:

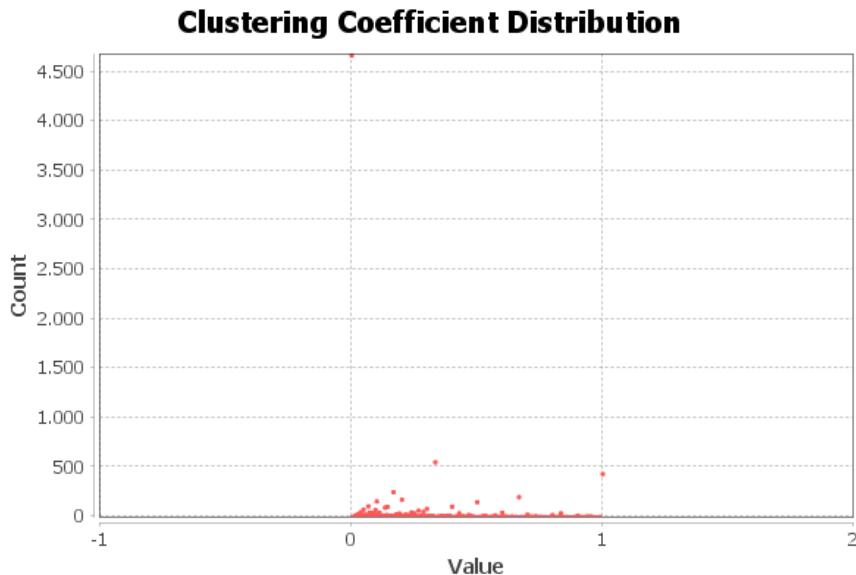


Figura 3: Gráfico de la distribución de coeficientes de clustering de la red.

Nos centraremos ahora en estudiar la conectividad de la red:

$$\text{Número de componentes conexas} \mid 1$$

Se da que **sólo es posible identificar en la red una única componente conexa**, es decir, no se distinguen en la red agrupaciones de nodos con diferente conectividad.

Tratamos de estudiar más detenidamente las medidas de conectividad de la red:

Estadísticos	
Diámetro de la red (d_{max})	15
Distancia media (d)	4.622979301417417

Tabla 2: Estadísticos referente a la centralidad de la red

3. Reflexiones sobre Estadísticos básicos

La **distribución de grados medios**, estudiada en la figura 2 en la anterior sección, se ajusta a la propiedad libre de escalas, es decir, **existen pocas páginas de figuras públicas con una gran cantidad de “likes” mutuos con otras páginas y muchas páginas de figuras públicas con pocos “likes” mutuos con las demás**.

Estos datos dan idea una red donde **cada página establece “likes” con una media de 12 figuras aproximadamente.**

Deteniéndonos en el **análisis del coeficiente de clustering en la red**, el coeficiente de clustering medio nos indicaría que, en media, un 21.5 % de las páginas tienden a establecer “likes” entre sí.

Por su parte, el **análisis de la distribución de los coeficientes de clustering**, permite concluir los siguientes hechos relevantes:

- Existen alrededor de 4500 páginas de figuras públicas cuyos vecinas (páginas de otras figuras públicas a las cuales se ha hecho ”like”) no establecen “likes” mutuos entre sí.
- Por el contrario, existen alrededor de 500 páginas de figuras públicas cuyas vecinas también establecerían “likes” entre sí.
- La distribución de los coeficientes de clustering tiende a la uniformidad a lo largo de su dominio, lo que indica que en la red, existen regiones de páginas poco conectadas entre sí y otras regiones de páginas muy conectadas entre sí sin que exista predominancia de un tipo de región sobre otra.

Respecto a las **medidas de conectividad estudiadas**, se determinó que la red únicamente presenta una única componente conexa, la red en su totalidad, por lo que las medidas de distancia máxima (d_{max}) y distancia media (d) se calcularon para toda la red.

Los valores de distancia máxima y distancia media nos permiten observar lo siguiente:

La distancia media entre cada par de páginas cualesquiera de la red es, aproximadamente de 5 enlaces, es decir, desde una página cualquiera, se necesitarían pasar por otras 4 páginas intermedias entre las que se halla realizado ”like” para llegar a la página destino siendo 15 el número máximo de enlaces existentes entre dos páginas cualesquiera de la red, es decir, 14 páginas intermedias.

Comparando estos valores con el número de enlaces de la red, nos da una idea de una **red densamente conectada y concentrada en la que es posible contactar con cualquier figura pública considerando pocas figuras intermedias.**

4. Análisis de la Centralidad

Se estudia la **Centralidad** de los nodos a partir de las medidas de **Grado**, **Intermediación** y **Cercanía**, para tratar de determinar las figuras públicas centrales en esta red de páginas de *Facebook*.

El estudio que se llevará a cabo en esta sección, tratará de obtener las 5 figuras públicas más centrales del grafo según cada una de estas medidas:

La **distribución de grados** que se estudió en la figura 1, mostraba una distribución de grados que se ajustaba a la **ley libre de escalas**, con pocas páginas con un grado superior a 1.800 y menos de 10 figuras con un grado igual a superior a 500.

Considerando ahora la medida de **Intermediación**, se calcula la intermediación de cada nodo de la red y se representa su distribución en la siguiente figura:

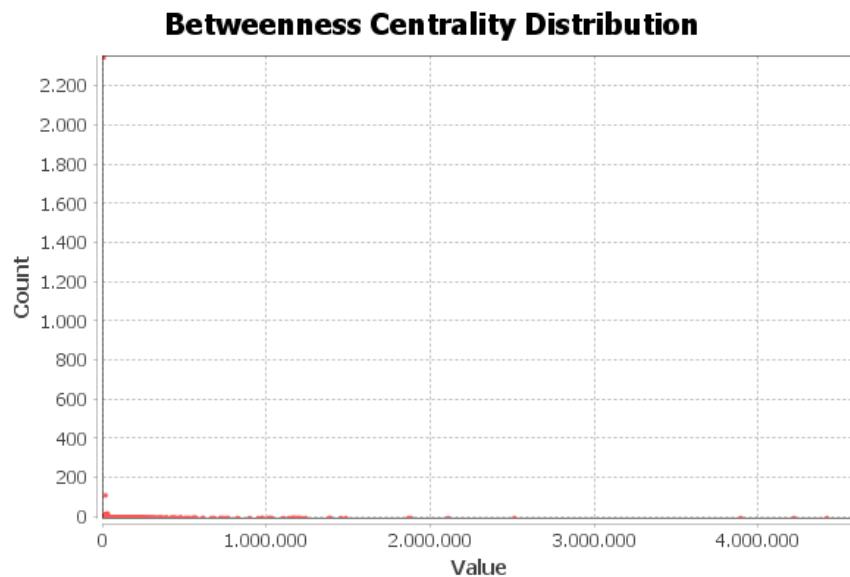


Figura 4: Gráfico de la distribución de la Intermediación de la red.

La distribución de esta medida se ajusta de forma más exacta a la ley libre de escalas: En esta distribución apreciamos muy pocas figuras públicas con un valor de intermediación cercano a 100 y alguna que tiene un valor superior a 2.200.

Respecto a la **Cercanía**, su distribución queda reflejada en la siguiente gráfica:

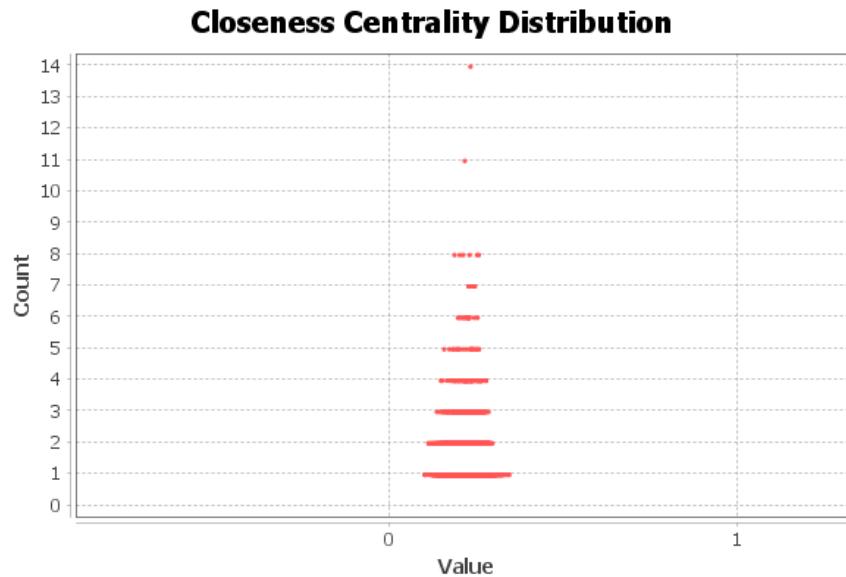


Figura 5: Gráfico de la distribución de la Cercanía de la red.

La distribución de esta métrica revela la **existencia de un número de nodos igual a 1 para los valores más grandes de Intermediación**.

Por último, analizamos la centralidad de Vector Propio. En la siguiente figura 6, se representa la distribución de los valores de centralidad de vector propio calculados para cada nodo de la red:

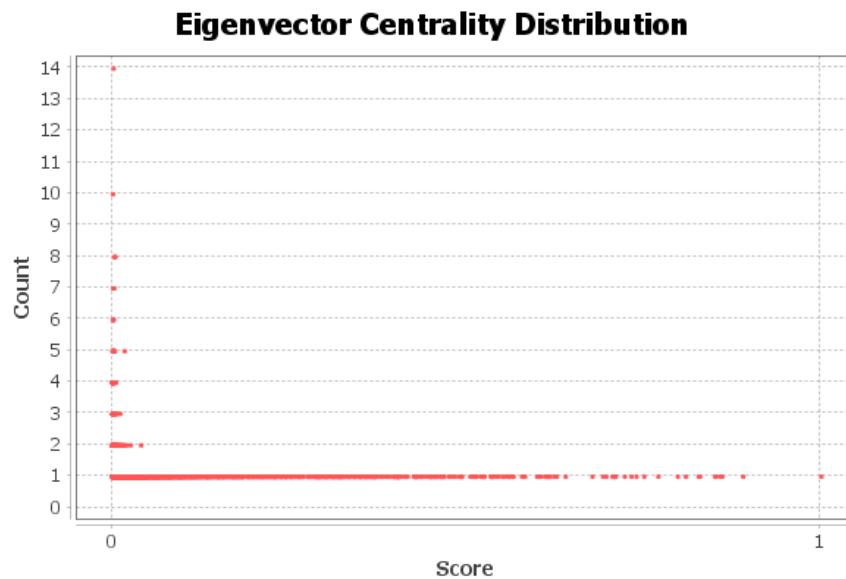


Figura 6: Gráfico de la distribución de la Cercanía de la red.

La distribución de valores de centralidad de vector propio, nos muestra la existencia de una única figura con un valor de centralidad 1, es decir, constituye el nodo más central considerando esta métrica.

Recogemos en la siguiente tabla las 5 figuras públicas más centrales según cada medida de centralidad:

Ranking	Grado	Intermediación	Cercanía	Centr. Vector Propio
1 ^a	Oprah Winfrey: 326	Daniel Amos: 4592547.561942	Daniel Amos: 0.340739	徐若瑄 Vivian Hsu: 1
2 ^a	徐若瑄 Vivian Hsu: 312	Dwayne The Rock Johnson: 4410508.138512	Dwayne The Rock Johnson: 0.340028	黃沐妍-小豬: 0.889804
3 ^a	Dwayne The Rock Johnson: 311	Hugh Jackman: 4209620.160734	Hugh Jackman: 0.337882	Sascha Grammel: 0.860307
4 ^a	煎熬弟 鍾明軒: 274	Oprah Winfrey: 3884964.431251	Oprah Winfrey: 0.333583	王彩樺-台灣濱 崎步: 0.856889
5 ^a	Hugh Jackman: 271	Sergio Carlo: 2504022.951233	Sergio Carlo: 0.331603	Remembering S R Nathan: 0.850986

Tabla 3: Ranking de actores/figuras públicas más centrales según las métricas de Grado, Intermediación, Cercanía y Centralidad de Vector Propio

Se observa que las figuras públicas que aparecen en la mayor cantidad de rankings son *Dwayne The Rock Johnson*, *Hugh Jackman* y *Oprah Winfrey*. *Daniel Amos* constituye la figura con más intermediación y con mayor Cercanía, es decir, constituye la figura de la que más se depende para comunicar comunidades diferentes y también, más cercana a todas las figuras.

Por último, 徐若瑄 *Vivian Hsu* junto con *Dwayne The Rock Johnson* obtienen posiciones altas en los rankings de Grado y Centralidad de vector Propio, lo cual indica que constituyen dos de las figuras públicas con mayor número de relaciones con otras figuras que también tienen un gran número de relaciones con otras figuras públicas, es decir, relaciones de importancia.

A continuación, se vuelve a representar el grafo destacando los nodos más intermedios y con mayor centralidad de vector propio:

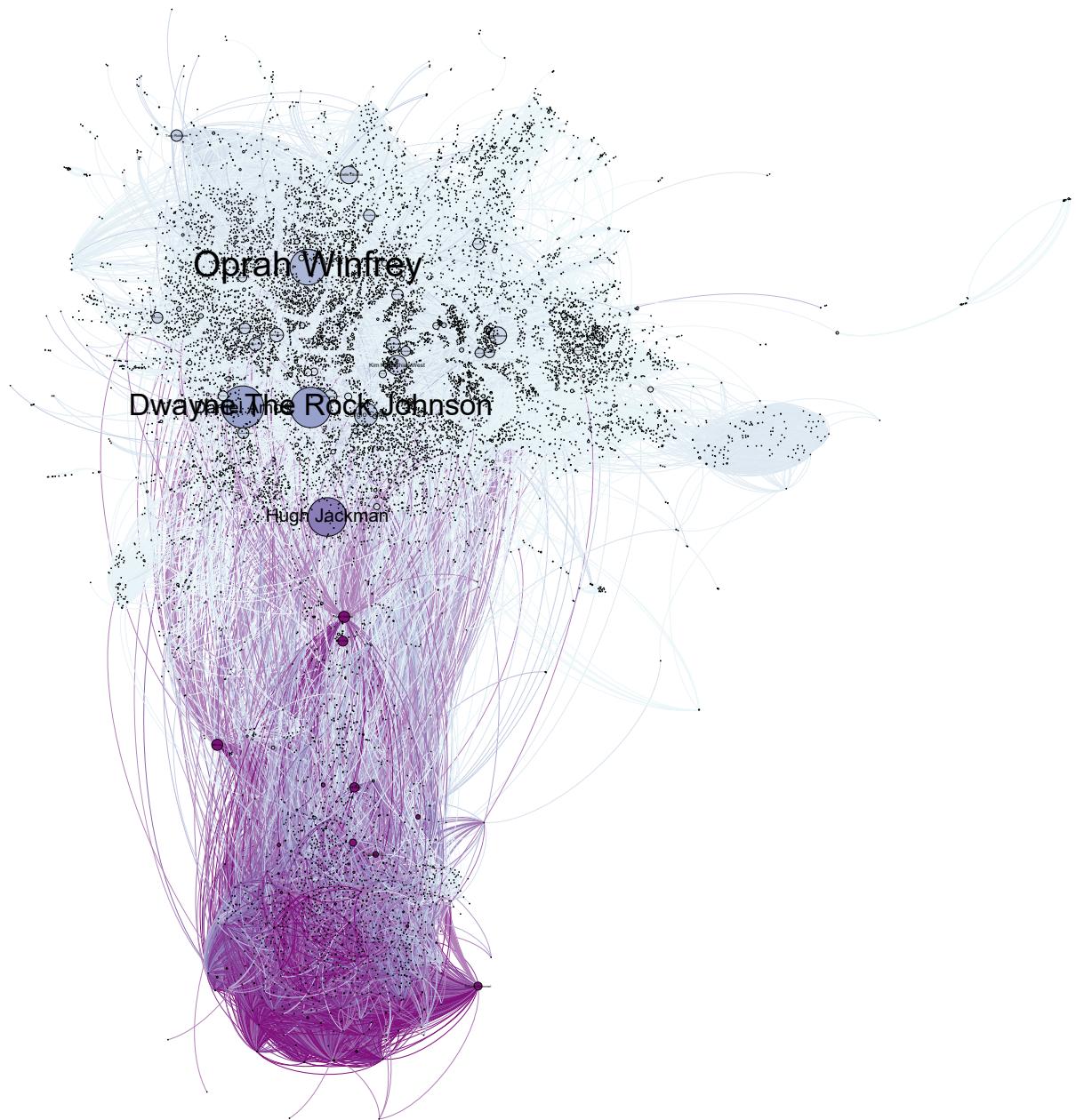


Figura 7: Gráfico de la red donde el tamaño de cada nodo y su coloración se han establecido en base a sus valores de Intermediación y centralidad de Vector Propio: Los nodos de mayor tamaño se corresponden con los nodos de mayor Intermediación, mientras que los nodos de color más violeta se corresponden con los nodos con mayor centralidad de vector propio.

Por último, se desea profundizar con mayor detalle en la relación entre

las medidas de Cercanía y grado. En el gráfico de la figura 8 se muestra la distribución de los valores de Cercanía de los nodos frente a los valores de Grado:

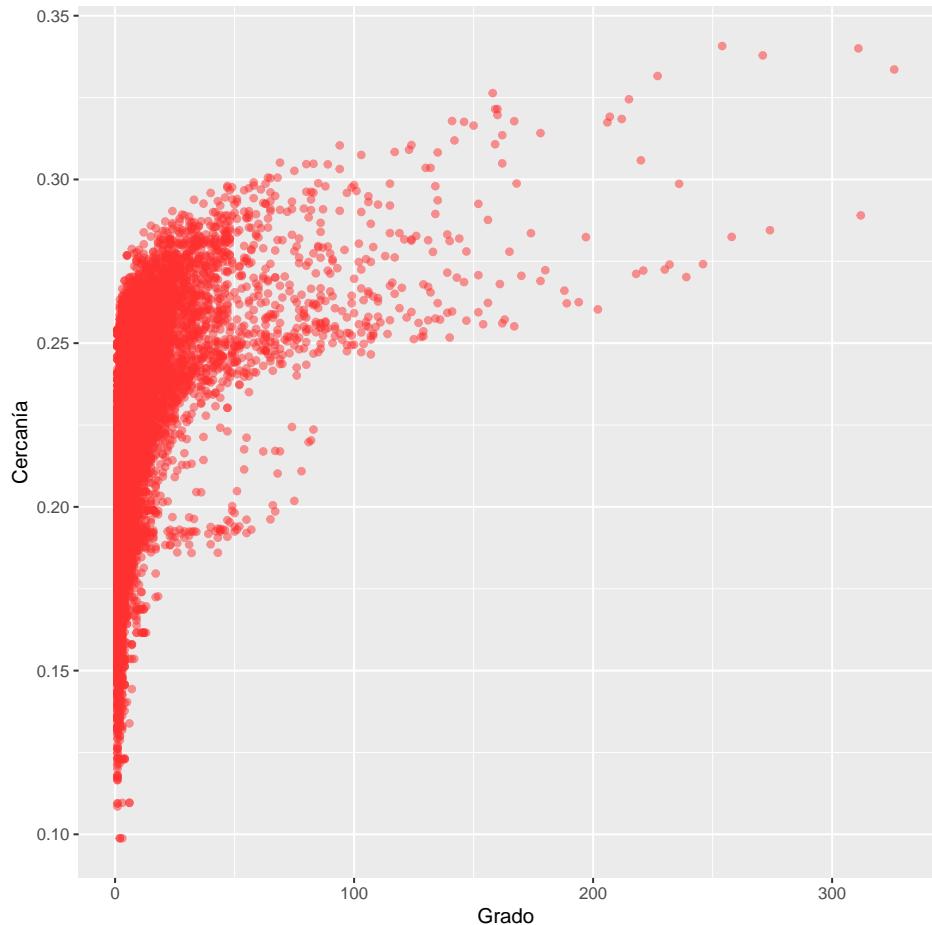


Figura 8: Distribución de los valores de Cercanía de la red frente a los valores de Grado

En el anterior gráfico se refleja un hecho relevante: existe una relación creciente entre la Cercanía de los nodos y su Grado, de modo que cuanto mayor es el grado de un nodo, su Cercanía tiende también a ser mayor. La relación entre ambas métricas muestra una relación que se asemeja a una función logarítmica.

Por su parte, el gráfico indica que la mayor parte de las páginas de la red tienen un grado inferior o igual a 50 al tiempo que su Cercanía es inferior o igual a 0.3.

5. Análisis de Comunidades

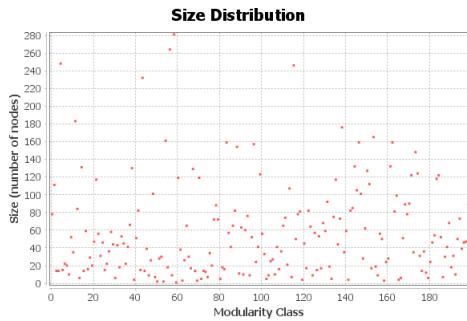
Finalmente, se tratará la división de la red en comunidades, estudiando diferentes particionamientos y se evaluarán los particionamientos más oportunis desde el punto de vista de su Modularidad (Q) y del número de Comunidades.

Probando con diferentes valores para el parámetro **Resolución**, obtenemos diferentes particionamientos para los cuales se anota el número de comunidades y la Modularidad (Q) del particionamiento:

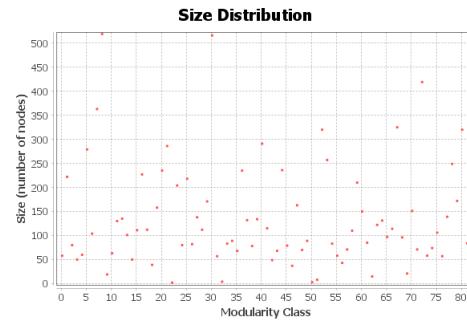
Resolución	Q	Nº de Comunidades
0.1	0.476	199
0.25	0.554	83
0.5	0.667	52
0.75	0.679	38
1	0.677	29
2	0.651	18
3	0.587	17
5	0.564	11
7	0.430	5
10	0.420	2

Tabla 4: Número de comunidades y Modularidad (Q) obtenida para cada particionamiento con cada valor de Resolución

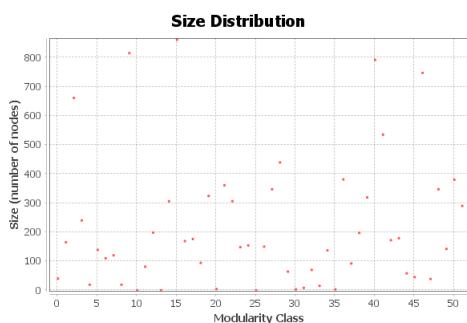
Analizamos también la distribución de los nodos en las comunidades para cada particionamiento realizado.



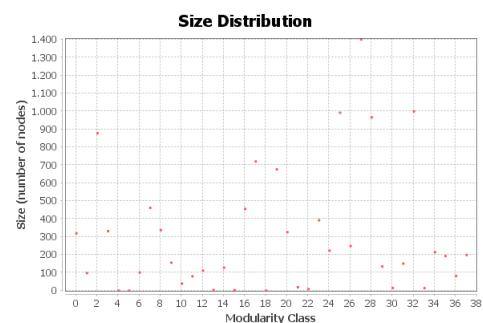
Resolución: 0.1



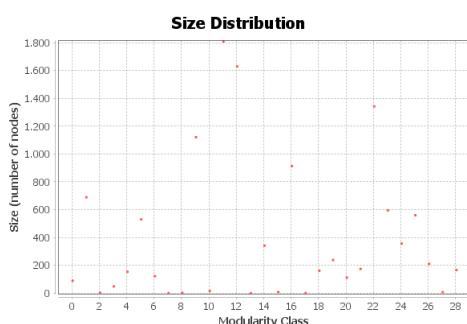
Resolución: 0.25



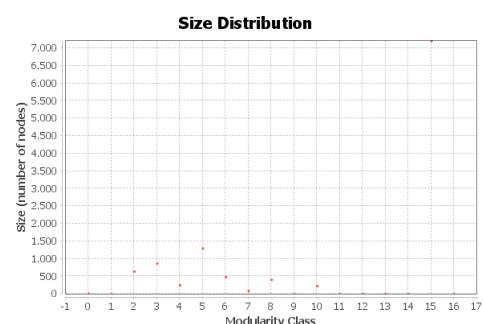
Resolución: 0.5



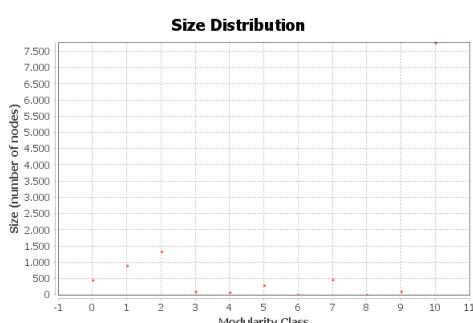
Resolución: 0.75



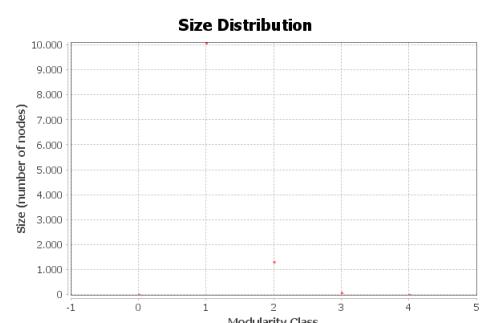
Resolución: 3



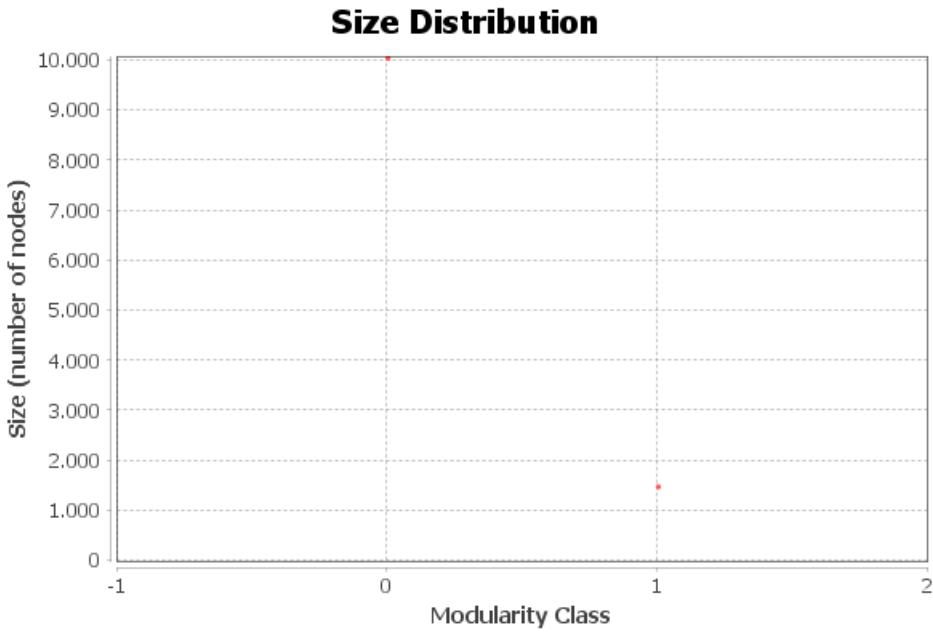
Resolución: 3



Resolución: 5



Resolución: 7



Resolución: 10

Figura 9: Gráficos de densidad de las comunidades formadas en cada particionamiento obtenido con los diferentes valores de Resolución.

El valor de modularidad de la red crece al tiempo que el número de comunidades hasta que se alcanza un número de comunidades entorno a 38. Una vez alcanzada esta reducción, conforme el número de comunidades sigue decrementándose, la modularidad también decrece.

Se observa por tanto, que los mejores valores de modularidad se alcanzan entorno a un número de comunidades entre 20 hasta 50 aproximadamente, alcanzándose el valor óptimo de modularidad para un número de comunidades de 38.

Respecto a la evolución de la aglomeración de las figuras en las distintas comunidades, observamos en la figura 9, que, al considerar un número bajo de Resolución existe una dispersión más uniforme de las figuras respecto de las comunidades. Al incrementar la resolución, disminuye esta dispersión uniforme dando lugar a la existencia de comunidades que tienden a aglomerar una gran parte de las figuras públicas y, para las cuales, el número de figuras que cae en las mismas tiende a estabilizarse conforme aumenta la resolución.

Por último, se proponen como particionamientos más interesantes, tan-

to desde el punto de vista del número de comunidades obtenido como de la modularidad obtenida, los particionamientos obtenidos para un nivel de Resolución de 0.75 y 5:

El primero ofrece el particionamiento más modular mientras que el segundo ofrece un particionamiento compacto respecto al número de comunidades conservando un valor alto valor de modularidad.

Para la partición obtenida con un nivel de Resolución de 0.75, la red queda dividida de la siguiente forma:

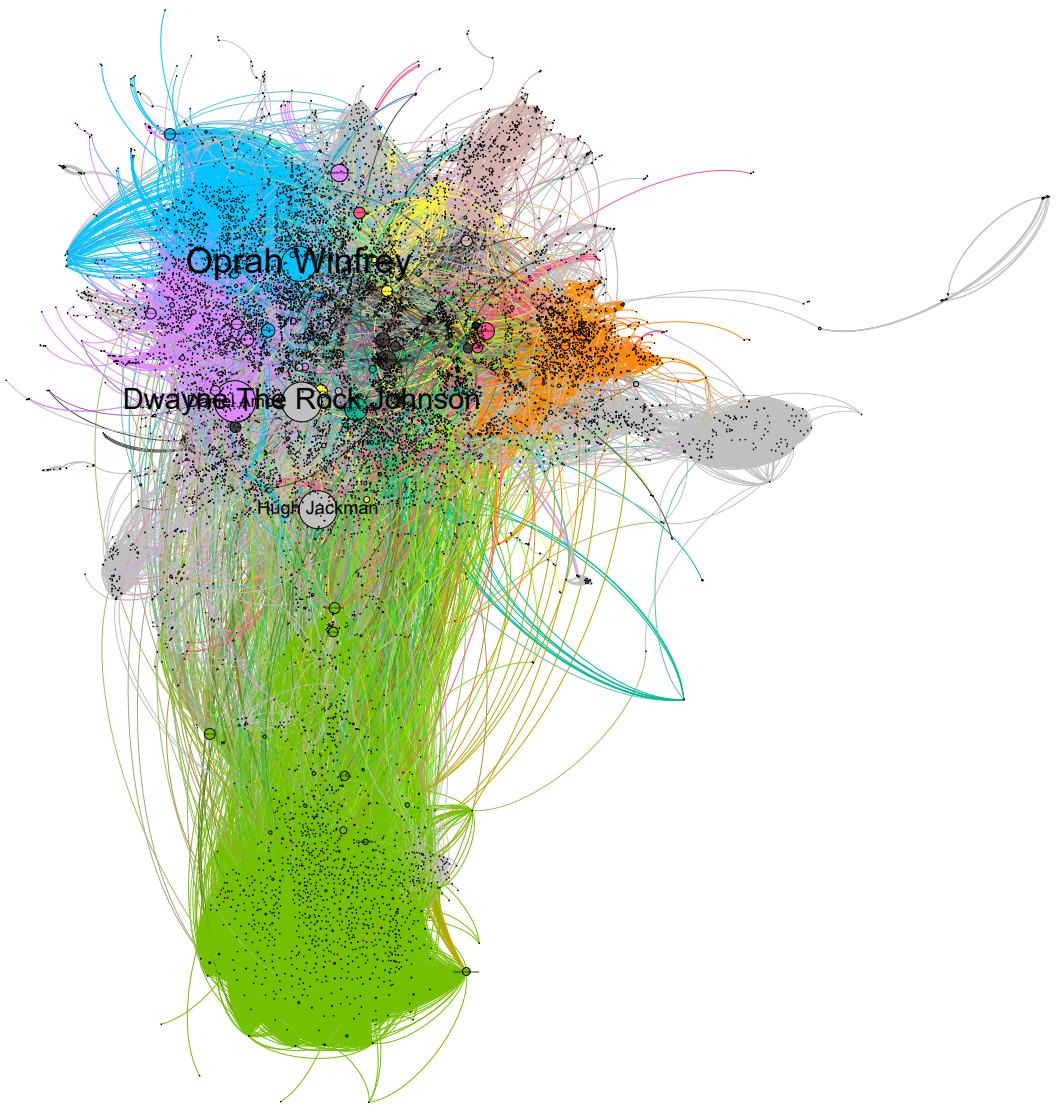


Figura 10: Gráfico de la red donde se ha coloreado cada conjunto de nodos y aristas en la comunidad a la que han sido asignados considerando un valor de Resolución de 0.75.

Cada comunidad queda formada con el siguiente conjunto de nodos:

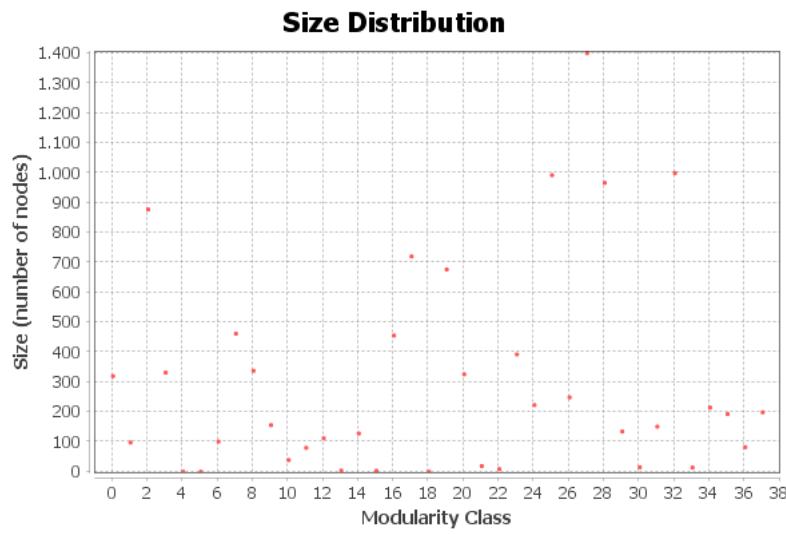


Figura 11: Gráfico de la distribución de densidad de las comunidades formadas en el particionamiento con un valor de Resolución de 5.

Se aprecia que la mayor parte de las comunidades formadas tienen un número de nodos inferior a 900, mientras que las comunidades 25, 27, 28 y 32 poseen un número de figuras públicas superior a 1000 (colores azul claro, rosa, negro y verde claro no olivo respectivamente).

Por su parte, para el particionamiento obtenido con un valor de Resolución de 5, la red se divide de la siguiente forma:

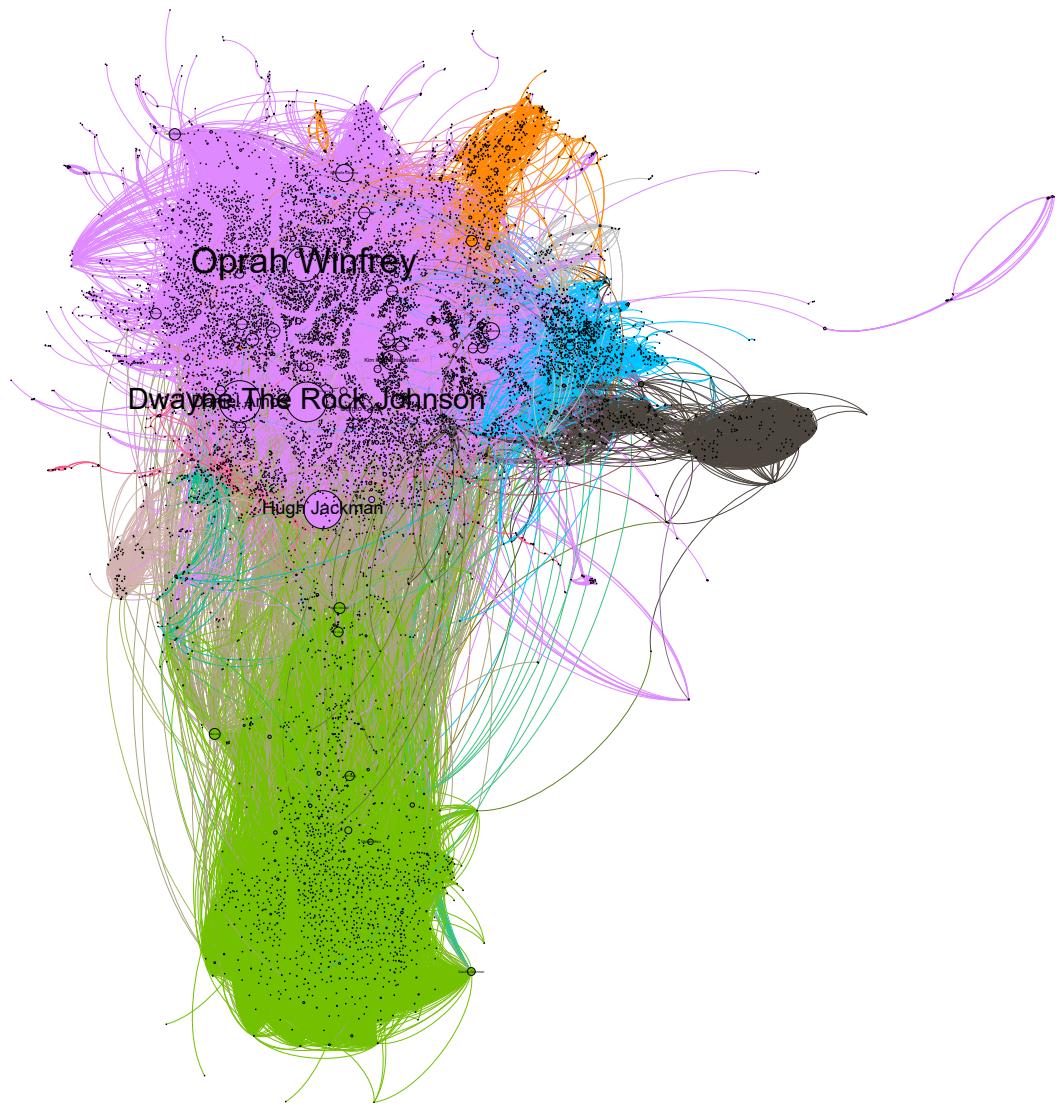


Figura 12: Gráfico de la red donde se ha coloreado cada conjunto de nodos y aristas en la comunidad a la que han sido asignados considerando un valor de Resolución de 5.

En esta partición, cada comunidad queda formada con el siguiente conjunto de nodos:

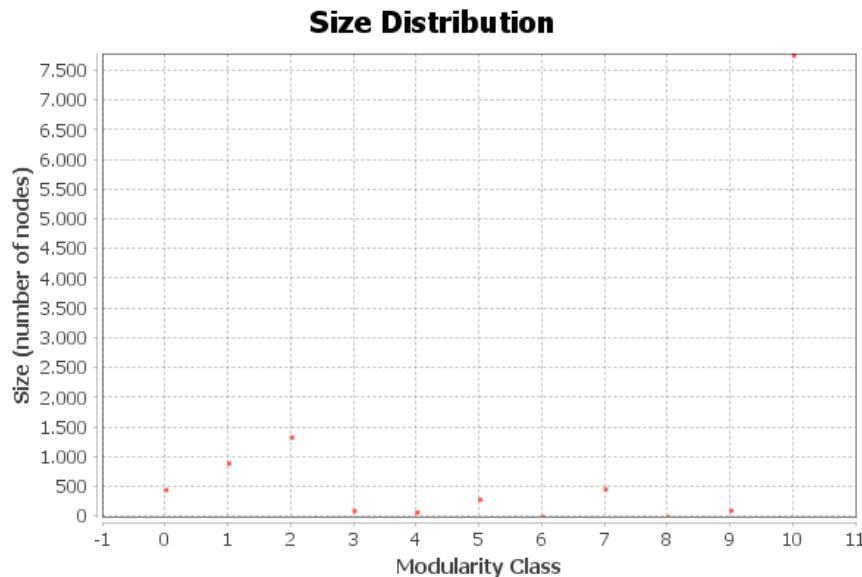


Figura 13: Gráfico de la distribución de densidad de las comunidades formadas en el particionamiento con un valor de Resolución de 5.

En esta partición, la comunidad 10 (color rosa) aglomera la mayor cantidad de figuras públicas (más de 7.500 nodos), mientras que el resto de comunidades aglomeran un número de figuras inferior a 1.500.

Referencias

- [1] Ryan A. Rossi and Nesreen K. Ahmed. The network data repository with interactive graph analytics and visualization. In *AAAI*, 2015.