

Redes y Sistemas Complejos

Práctica 4

Guillermo Gómez Trenado | 77820354-S
guillermogotre@correo.ugr.es

15 de diciembre de 2018

Índice

1. Introducción	2
2. Datos usados y preguntas	2
2.1. Datos utilizados	2
2.1.1. Entidades comunes	4
2.1.2. Artistas relacionados	4
2.1.3. Fechas similares	4
2.1.4. Filtrado de enlaces	5
2.1.5. Dimensión de los enlaces	5
2.1.6. Análisis de los datos	5
2.2. El problema del tiempo	6
2.3. Cuestiones a analizar sobre la red	6
3. Análisis de la red	6
3.1. Análisis preliminar	6
3.2. Análisis red simplificada	8
3.3. Componente conexa	10
4. Detección de comunidades y análisis cualitativo	12
5. Reflexiones finales	19

1. Introducción

En esta práctica pretendo dibujar una radiografía de los actores implicados en la historia del arte, en concreto en la pintura y materias afines, quiero ver las relaciones y elementos de cohesión y la relevancia de actores concretos, movimientos artísticos y nacionalidades. Para ello vamos a obtener una red desde Wikidata, analizarla y extraer información valiosa para nuestro propósito. Debido al enorme tamaño de la red tendré que ir tomando decisiones críticas a lo largo del camino para asir la complejidad añadida que suma a nuestro problema.

2. Datos usados y preguntas

Wikidata es una base de conocimiento de la Fundación *Wikimedia*, de libre acceso, organizada en tuplas (*Entidad, Relación, Entidad*) que definen tanto sus relaciones con otros agentes como los atributos y propiedades de cada agente. *Wikidata* espone una API de consulta en lenguaje *SPARQL*, con una sintaxis similar a SQL pero orientada a este tipo de datos. El conocimiento acumulado esta base de datos es de una utilidad humana indescriptible y que sea un esfuerzo global y libre no hace más que acentuar su valor, sin embargo, al no ser un esfuerzo estructurado y uniforme organizado en una dirección desde un agente en concreto encontramos que la información es incompleta —no todos los agentes tienen las mismas propiedades— e irregular —es posible que la densidad o la riqueza en la descripción de ciertos actores no refleja su importancia histórica—, aunque vamos a intentar extraer conocimiento a pesar de estas dificultades con la información que sí tenemos.

2.1. Datos utilizados

El proceso seguido fue el de obtener una lista de entidades que fueran instancia de la clase **Persona** y cuya profesión fuera instancia de **Pintor** o una subclase de ésta. Por cada autor obtuve además la siguiente lista de relaciones —nótese que pueden ser relaciones uno a uno o uno a muchas—.

1. Lugar de nacimiento
2. Fecha de nacimiento (Fecha)
3. Nacionalidad
4. Lugar de trabajo
5. Época conocida (Fecha) *Esta relación sólo aparece en los autores en los que no se conozca su fecha de nacimiento o de muerte*
6. Influenciado por
7. Género artístico

8. Espónsor
9. Familiares no directos
10. Estudiantes de los que ha sido maestro
11. Haber sido estudiante de
12. Movimiento artístico
13. Padre
14. Hijo
15. Cónyuge
16. Hermano
17. Lugar de estudios/formación
18. Madre
19. Participante de —Exposiciones y eventos de relevancia—

El número de autores recogidos en Wikidata con el único criterio de ser autores era superior al tipo de consultas que se pueden realizar, éstas están limitadas a tiempos de computación de 5 segundos, y por la forma en la que se almacenan las bases de datos y se optimiza su acceso, la recuperación es no determinística, y el uso de los keywords *LIMIT* y *OFFSET* no tienen validez para recuperar todas las instancias, así que como primera estrategia de filtrado recuperé únicamente las instancias que tuvieran al menos una imagen asociada, pues observé en un análisis preliminar que todos los autores relevantes tenían una o varias fotos, ya fuera de autorretratos, retratos realizados por otros, fotografías, esculturas realizadas de ellos e incluso recreaciones de su apariencia en el caso de autores muy antiguos —tenemos datos de autores desde el 400 antes de Cristo—.

Tras esto, el proceso para crear la red siguió varias etapas, atendiendo al tipo de enlace que describían las relaciones. Debido a la magnitud del problema enfrentado tuve que hacer varios sacrificios que agilizaran el tiempo de desarrollo y ejecución, el primer sacrificio fue asumir que todas las relaciones eran simétricas, lo cual no es cierto en todas, y me daba así un grafo no dirigido, el segundo sacrificio fue asumir que todas las relaciones tenían el mismo peso, así cuando hay una relación entre A y B se suma 1 a su enlace. A continuación describo el proceso de creación de enlaces según el tipo de relación.

2.1.1. Entidades comunes

El primer tipo de relaciones que analicé fue aquellas en las que dos autores, A y B comparten la parte derecha de la tupla relacional $(A, r, C) \wedge (B, r, C)$, si sucedía esto añadía 1 al peso de su enlace. Nótese que esto genera el primer problema, dos personas que compartan el mismo padre es posible que también contengan el atributo hermano donde se referencian recíprocamente, sin embargo, observé que esto no sólo no sucedía siempre, sino que la frecuencia en la que esta información era consistente era bajísima, así que ignoré este tipo de circunstancias. Las entidades están ordenadas por orden alfabético y sólo rellenamos el triangulo superior derecho de la matriz de pesos, para evitar redundancia.

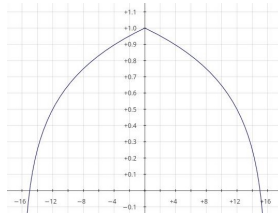
2.1.2. Artistas relacionados

El segundo tipo de relaciones de interés es cuando un autor en alguna de las relaciones tiene a otro autor como parte derecha de la tupla. En este caso, como ya comentábamos tratamos todas las relaciones como no simétricas y finalmente sumamos el valor de los enlaces (A, B) y (B, A) para tratar los enlaces como simétricos, esto introduce cierto ruido semántico en la red pero agilizaba muchísimo el proceso de desarrollo. El objetivo de todo este tipo de estrategias como ya se puede intuir es reducir el tamaño de la red y en especial el número de enlaces, pues como ya analizaremos después, no es pequeño este número. Debido a la falta de consistencia en la red, no son muchos los casos en los que se de que ambos autores se tengan el uno al otro en alguna relación, incluso en aquellas como *Hermano* donde cabría esperar que esto sucediera, y por lo tanto, el impacto de este sacrificio sospecho que no repercute en exceso en la red resultante.

2.1.3. Fechas similares

Para la época lo primero que hice fue unir en una misma tabla la fecha de nacimiento y la época conocida. Para homogeneizar esta relación sumé 30 a la fecha de nacimiento como estimación de cuándo destacaría un autor si no conociéramos su fecha de nacimiento. Posteriormente definí el peso de la relación como

$$dist(x, y) = \frac{\log(N + 1 - |age(x) - age(y)|)}{\log(N)}, \quad |age(x) - age(y)| < N$$



En un principio intenté crear los enlaces con $N = 40$; pero el número de enlaces era tan desproporcionado que se hacía impracticable usar la red, así que fui reduciendo el valor hasta $N = 15$, que da como resultado sólo 36 millones de enlaces.

Tras lo anterior, hice un análisis preliminar de la red sin las distancias, obtuve aquellos nodos con valor de centralidad de vector propio mayor que 0,03 —conservando el 60 % de la red—, me quedé con la componente conexa y a continuación conservé únicamente los enlaces sobre fechas que tuvieran en al menos uno de los extremos a un nodo con *eigenvalue* > 0,03, así conseguí reducir el tamaño en bytes de la lista de enlaces por fecha de 1.5GB a 400MB.

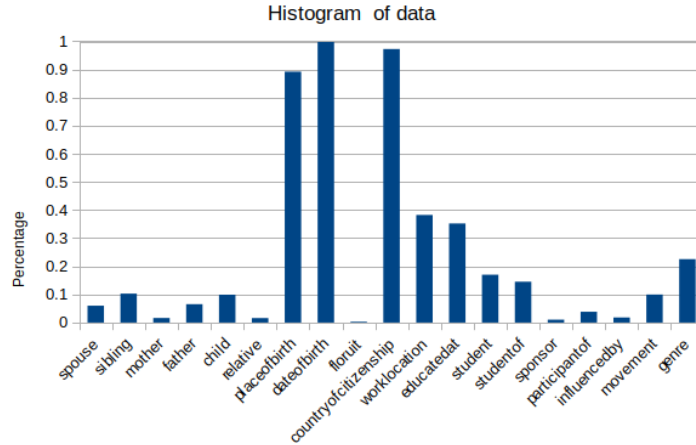
2.1.4. Filtrado de enlaces

Finalmente, junto al paso anterior descarté todos los enlaces con valor estrictamente menor que 1, es decir, es decir tienen que compartir al menos dos relaciones cualitativas o una relación cualitativa y estar próximos en el tiempo.

2.1.5. Dimensión de los enlaces

Todos los enlaces están definidos tal que a mayor es su valor, mayor es la intensidad de la relación, sin embargo, definí otra lista paralela, para su uso posterior en el algoritmo *MST-Pathfinder* donde el peso viene dado por $w' = 1/w$, pudiendo así obtener los caminos de menor coste.

2.1.6. Análisis de los datos



En la gráfica anterior podemos ver la proporción de autores con cada una de las propiedades o relaciones rellenas. Es evidente que la red va a estar gobernada por las relaciones derivadas de la nacionalidad, y la fecha y lugar de nacimiento; y esperamos que los actores centrales sean aquellos que además de estas propiedades tengan más relaciones permitiéndoles crear vínculos más fuertes con los

demás actores.

2.2. El problema del tiempo

Hay una dificultad inherente al diseño de nuestra red y es el papel crucial que juega el tiempo en esta red histórica y que no tenemos forma sencilla de controlar, pues al mezclar relaciones dirigidas —influenciado por— y no dirigidas —misma nacionalidad— es difícil controlar que no se creen caminos que recorran el tiempo en dirección contraria. No me queda más remedio que ignorar este factor.

2.3. Cuestiones a analizar sobre la red

Debido a las características tan particulares de la red es difícil plantear preguntas que se puedan responder con rotundidad, por ejemplo, qué movimiento artístico ha tenido mayor relevancia sobre los siguientes autores, o cuáles son los pintores con mayor impacto en las siguientes generaciones. Sin embargo, al eliminar la restricción temporal las preguntas deberían versar sobre elementos estáticos, como puede ser la nacionalidad o los autores centrales en la historia del arte. Por otro lado, las propiedades que constan en todos los autores con mayor frecuencia están limitadas a características espaciales y temporales. Por todo lo anterior, si bien mi objetivo principal es dibujar la radiografía de la historia del arte y la importancia relativa de sus agentes, me pregunto cuáles son los **países con mayor impacto** en esta red de artistas, así como las **épocas claves** que definen la centralidad de los autores. Y me temo que una pregunta de interés para mí como era la relevancia del movimiento pictórico al que se adscribían los autores va a quedar desplazada por el peso de las propiedades dominantes en la red.

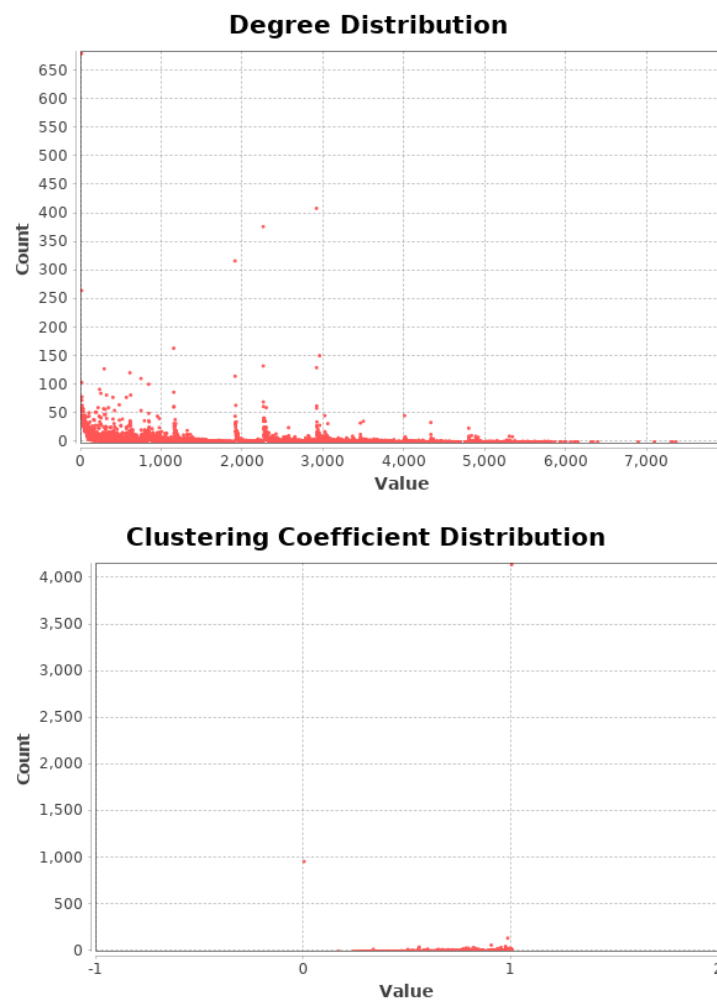
3. Análisis de la red

Vamos a dividir el análisis en dos partes, en primer lugar un análisis preliminar de la red obtenida con los pasos anteriormente descritos, y después voy a podar y analizar esta nueva red simplificada. Debido al gran tamaño de la red no podemos hacer análisis que involucren las distancias mínimas más que sobre un conjunto aleatorio de nodos y sus enlaces que realizo sobre la red simplificada. Asimismo, el análisis preliminar será muy superficial y analizaré con mayor detenimiento la red tras la primera poda en enlaces.

3.1. Análisis preliminar

1. **Número de nodos** 25197
2. **Número de enlaces** 16745049
3. **Grado medio** 1329.13

4. **Grado medio con pesos** 1451.252
5. **Densidad de la red** 0.0263
6. **Componentes conectadas** 689
7. **Tamaño de la componente gigante** 21431 (85.05 %)
8. **Coefficiente de clústering** 0.824
9. **Coefficiente de clústering red aleatoria** 0.05



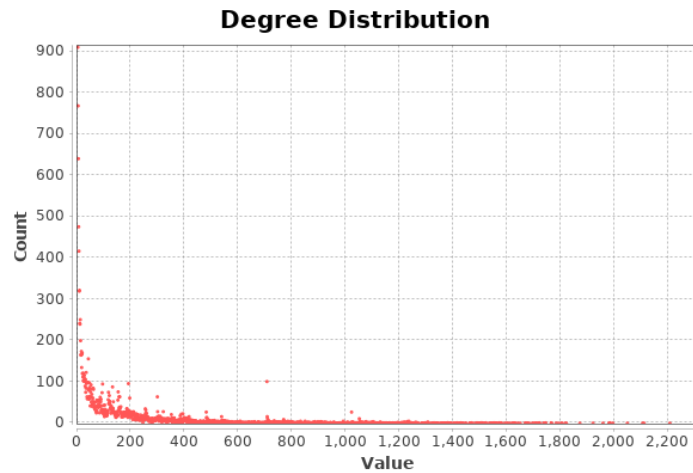
Como podemos observar en el gráfico de la distribución de grados tenemos una red libre de escala, con una densidad relativamente alta para este tipo de redes, con coeficiente de clústering alto, pero antes de proseguir con el análisis

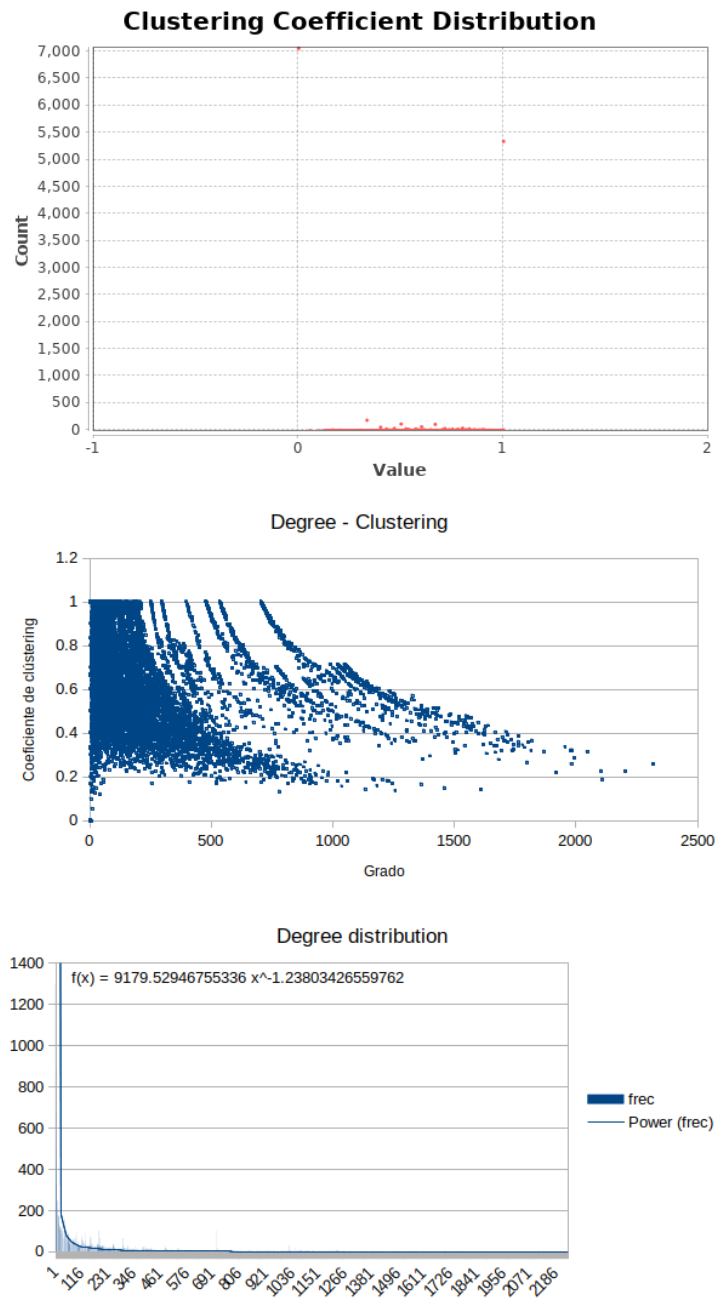
de esta red, ya que vemos la alta conectividad de la red, como ya adelantábamos en el apartado de *Filtrado de enlaces* vamos a eliminar aquellos enlaces menos significativos.

3.2. Análisis red simplificada

En definitiva nos estamos dirigiendo a un modelo simplificado de nuestra red, reduciendo la dimensionalidad de ésta, donde se conserven los actores más relevantes, al aplicar la transformación antes mencionada obtenemos una red con los siguientes valores

1. **Número de nodos** 25197
2. **Número de enlaces** 1580734
3. **Grado medio** 166.34
4. **Grado medio con pesos** 261.746
5. **Densidad de la red** 0.009
6. **Componentes conectadas** 5879
7. **Tamaño de la componente gigante** 19237 (75.35 %)
8. **Coefficiente de clústering** 0.741
9. **Coefficiente de clústering red aleatoria** 0.007





Vemos un resultado consistente con lo que esperábamos, hemos reducido el número de enlaces casi al 10 %, resultando en una red mucho menos densa, más parecido a lo que estamos acostumbrados a ver en otras redes sociales $\sim 0,01$. El mayor impacto lo hemos tenido en el número de componentes, que a aumentado

en 5 el número, debido a que aquellos actores que estaban únicamente relacionados por el país de origen o la fecha han sido desconectados, resultando en un número mayor de pequeños grupos o pintores aislados; sin embargo el tamaño de la componente gigante no ha cambiado significativamente, lo cual nos invita a pensar que vamos por buen camino en nuestra poda.

Nos encontramos, como ya adelantábamos con una red libre de escala donde a pesar de eliminar los enlaces más débiles seguimos teniendo una alta conectividad local, con el coeficiente de clústering igual a 0.741, donde tres de cada cuatro vecinos de un nodo están conectados entre sí, lo cual refleja un poco el comportamiento esperado en el mundo del arte, donde los artistas se relacionan entre ellos con alta frecuencia en su comunidad. Por otro lado, la relación entre grado y coeficiente de clústering nos muestra una gráfica donde como es esperable los agentes con grado menor tienen un coeficiente de clústering alto por relacionarse en comunidades densas, aunque a la izquierda de la gráfica encontramos una masa ruidosa que no sigue una tendencia clara, mi principal sospecha es que se debe a que todos los puntos tienen el mismo diámetro, y esos puntos que no siguen la tendencia —al ser 25000 nodos— ocupan mucho área de la gráfica pero no representan realmente la tendencia mayoritaria, cuyo protagonismo se ve reducido por solaparse unos sobre otros.

Por otro lado, vemos cómo el exponente de grado, obtenido aproximando una recta logarítmica a la distribución de grados, es de 1,24, algo bajo para lo que estamos acostumbrados, aunque quizás se deba a la poda de enlaces que hemos realizado.

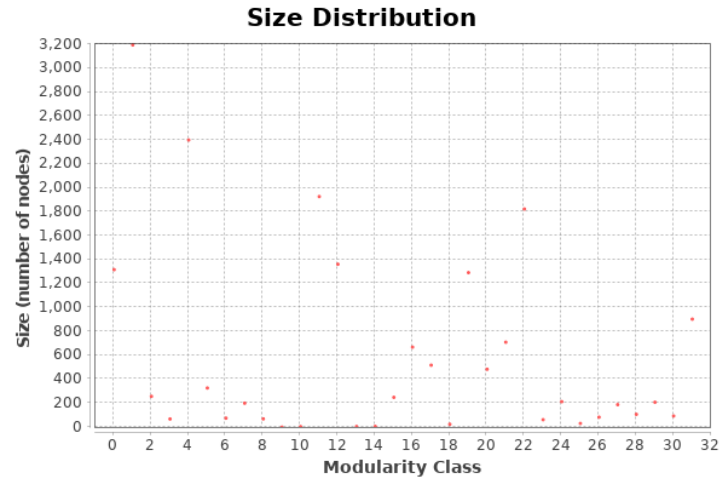
Finalmente, debido al tamaño de la red no he podido obtener el diámetro, radio y distancia media, porque su computación era inasumible, aunque haciendo experimentos sucesivos con subconjuntos aleatorios de la red observo que este valor suele encontrarse entre 10 y 11, lo cual es consistente con el modelo de mundos pequeños, siendo $\ln(25197) = 10,13$.

3.3. Componente conexas

Para la siguiente etapa de nuestro análisis exporté la componente conexas y apliqué el algoritmo *MST-Pathfinder* ($r = \infty, q = n - 1$) para obtener una PF-NET más asumible en el dibujado. Para este paso, utilizo el valor de centralidad de vector propio calculado sobre la red anterior antes de podar, como estimador de la centralidad de un actor utilizo el *eigenvalue* por dos motivos, en primer lugar porque es buen reflejo comprobado de la centralidad de un actor en la red y en segundo lugar, porque otros valores de interés, como la intermediación o la cercanía dependen del cálculo de caminos mínimos y era inasumible en una red de este tamaño.

1. **Número de nodos** 19001
2. **Número de enlaces** 972198
3. **Grado medio** 102.331

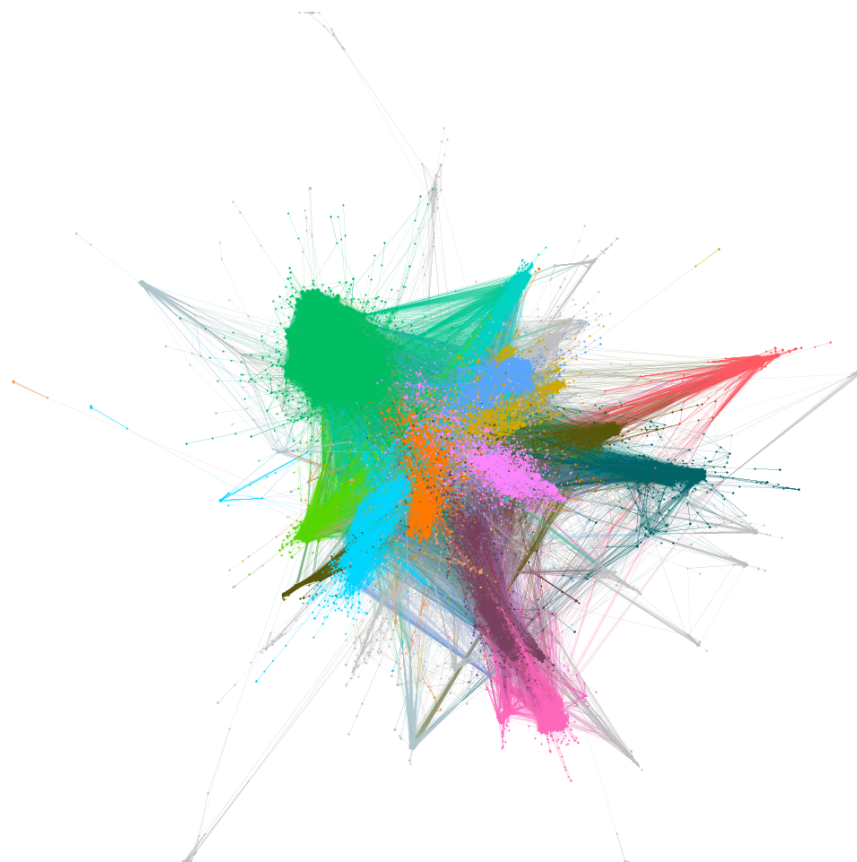
4. Modularidad 0.696
5. Número de comunidades 32



No me voy a detener en describir el resultado del valor de vector propio para actores individuales porque no tiene excesivo sentido este análisis en una red de este tamaño, aunque sí que analizaremos posteriormente esta tendencia en relación a las comunidades.

La primera visualización que vamos a ver de la comunidad, utilizando *ForceAtlas 2* para su *layout* es poco descriptiva, pero es francamente bonita, podemos detenernos un segundo en apreciarla y seguir con el análisis.

Cada nodo está coloreado según el color único del clúster para los 12 módulos de mayor tamaño, y el resto están coloreados en gris —son éstos los nodos que vemos empujados a la periferia—. Cabe destacar el grupo de color verde neutro, el clúster 1, la masa superior izquierda, porque agrupa a los actores con mayor valor de centralidad de vector propio con diferencia así como es el clúster de mayor tamaño.



4. Detección de comunidades y análisis cualitativo

He utilizado para la detección de comunidades el método de Lovaina implementado en Giphi, descargué otra versión en C++¹ que me permitía obtener los clústers en distintos niveles de recursividad, pero la información que añadían era tan poca —sólo los dos primeros niveles introducían cambios cuantitativos suficientemente interesantes— que decidí dejarlo al margen del análisis. Evidentemente tuve que descartar el algoritmo de Girvan-Newman debido al tamaño de la red.

Utilizando las comunidades detectadas por el método de Lovaina he analizado los tres valores más frecuentes respecto al tamaño total de la clase y respecto al número de objetos etiquetados con esa relación, la tabla es demasiado grande

¹<https://sourceforge.net/projects/lovain/>

para mostrarla aquí y la adjunto en una hoja de cálculo en este mismo zip, este es un ejemplo de resultado para el clúster 1.

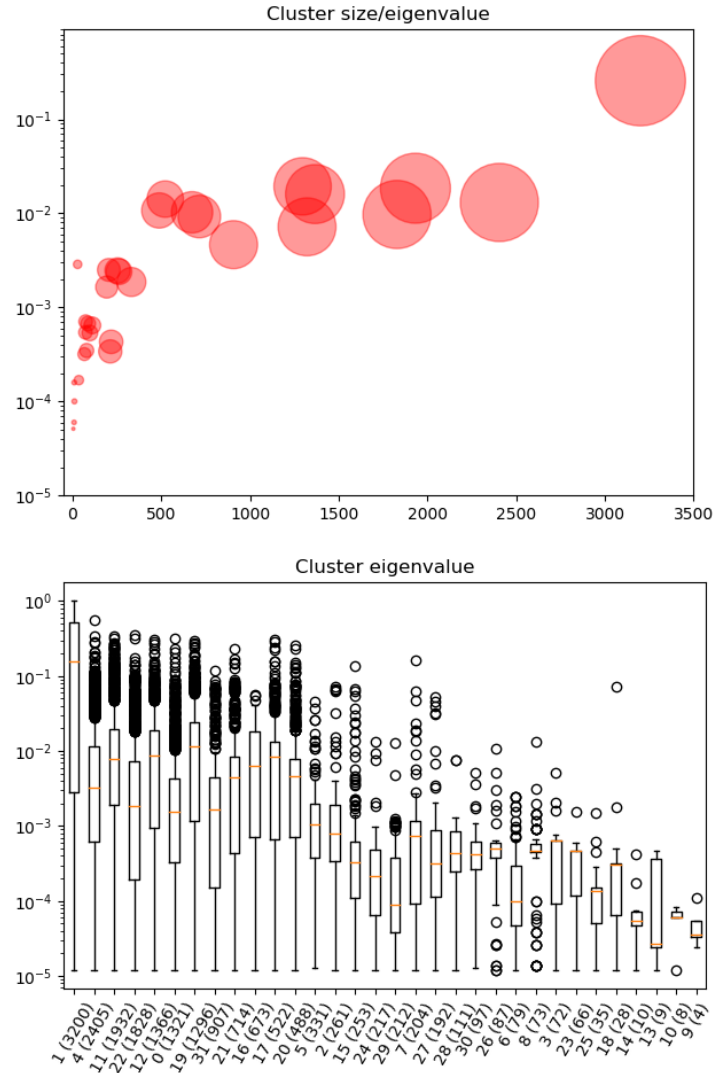
Cluster 1 (Eigenvalue: 0.25724971999999935) (Size: 3200):				Prop total	Prop etiq		Prop total	Prop etiq	
placeofbirth	Paris (Q90)			0.226875	0.241756242	Lyon (Q456)	0.0278125	0.02863768	Marseille (Q29482)
date	18			0.48711875	0.488740082	19	0.2669375	0.26654072	17
countryofcitizenship	France (Q142)			0.8021875	0.81673566	United Kingdom of Great Britain and Ireland (Q174193)	0.015625	0.015006368	Switzerland (Q39)
worklocation	Paris (Q90)			0.1809375	0.430483271	France (Q142)	0.026875	0.00394052	Rome (Q220)
educatedat	École nationale supérieure des Beaux-Arts (Q273593)			0.1653125	0.447625487	Académie Julian (Q337480)	0.0340625	0.002394666	École nationale supérieure des Beaux-Arts (Q273593)
student	Jean-Jacques Henner (Q655818)			0.0040625	0.0040625	Paul Émile Chabes (Q294791)	0.000625	0.00058606	Maurice Chabas (Q2391228)
studentof	Jean-Léon Gérôme (Q212499)			0.0140625	0.04	Alexandre Cabanel (Q310437)	0.0125	0.035555556	William-Adolphe Bouguereau (Q2391228)
sponsor	Louis Philippe I (Q7771)			0.00125	0.047058824	Philip the Good (Q239337)	0.0009375	0.035294118	Charles the Bold (Q1511205)
participantof	documenta III (Q26553)			0.01125	0.222222222	Documenta II (Q265448)	0.0090625	0.179012346	documenta 1 (Q265376)
influencedby	Paul Cézanne (Q35549)			0.00340625	0.00340625	Master of Spaccio (Q294982)	0.000625	0.040618825	Gustave Courbet (Q194128)
movement	academic art (Q189458)			0.019375	0.100813008	Impressionism (Q40415)	0.011125	0.068202683	neoclassicism (Q14376)
genre	portrait (Q134307)			0.1434375	0.496216216	history painting (Q742333)	0.046875	0.162162162	landscape art (Q191165)

Lo que observo es que el mayor elemento de cohesión dentro del grupo, con mucha diferencia es en primer lugar la nacionalidad, y en segundo lugar valores como el centro donde se formó y en casos excepcionales el movimiento artístico; el valor para la fecha —expresado como el siglo menos uno, posteriormente en las gráficas el siglo sí está corregido— aunque la frecuencia es muy alta en todos los clústers, más que un elemento de cohesión es una representación de la frecuencia general, pues los autores del siglo XIX y XX son mucho más numerosos que el resto. También encontramos descrito en la tabla, para cada clúster, el valor medio de centralidad de vector propio y el tamaño de la misma.

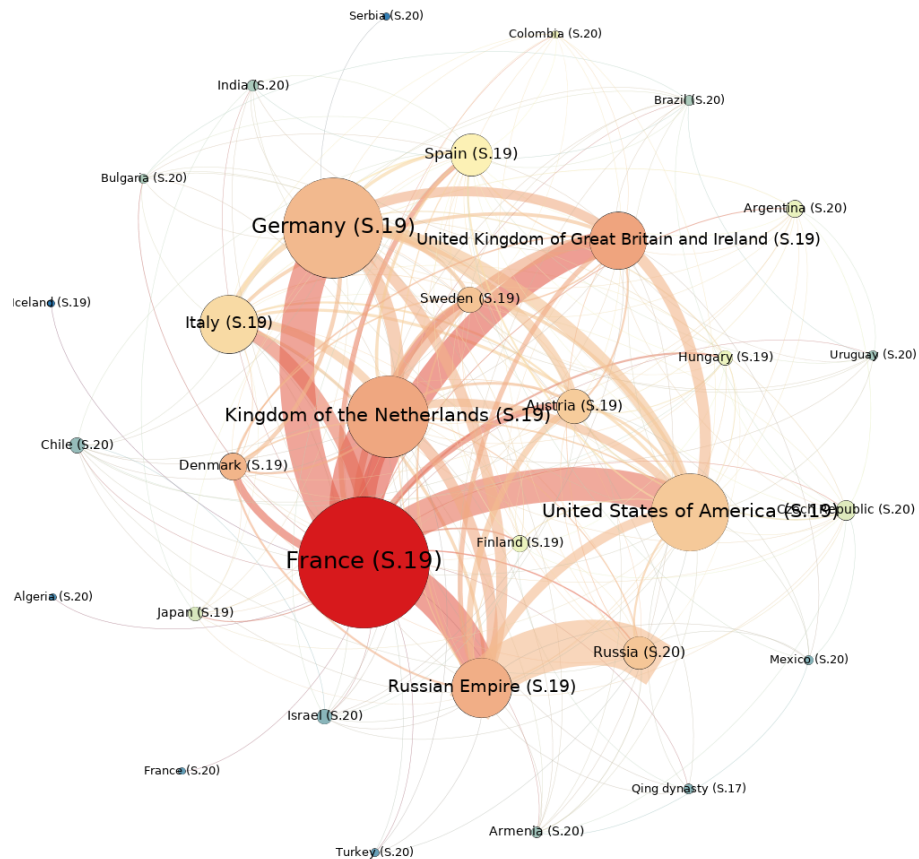
Existen una serie de países, que aglutinan en ellos al grueso de artistas de Wikidata y absorben a aquellos de otros países con menor influencia en el conjunto.

ID	Label	Size	Mean eigen *100	Highest artist	Highest eigen
1	France (S.19)	3200	25.72	Edouard Louis Dubufe	1.000
4	Germany (S.19)	2405	1.31	Léon Hornecker	0.560
11	Kingdom of the Netherlands (S.19)	1932	1.85	Huib Luns	0.346
22	United States of America (S.19)	1828	0.97	Charles Christian Nahl	0.348
12	Russian Empire (S.19)	1366	1.60	Piotr Michałowski	0.301
0	Italy (S.19)	1321	0.72	Luigi Calamatta	0.320
19	UK and Ireland (S.19)	1296	1.94	Rudolf Lehmann	0.290
31	Spain (S.19)	907	0.46	Diego Velázquez	0.116
21	Austria (S.19)	714	0.93	Ferdinand Jagemann	0.234
16	Russia (S.20)	673	1.03	Simon Kozhin	0.057
17	Denmark (S.19)	522	1.42	Carl Balsegaard	0.311
20	Sweden (S.19)	488	1.07	Amalia Lindegren	0.253
5	Czech Republic (S.20)	331	0.19	František Tkadlík	0.046
2	Argentina (S.20)	261	0.24	Prilidiano Pueyrredón	0.073
15	Finland (S.19)	253	0.25	Aleksander Lauréus	0.136
24	Chile (S.20)	217	0.04	Louis Buvelot	0.013
29	Israel (S.20)	212	0.03	Dani Karavan	0.013
7	Hungary (S.19)	204	0.25	Bertalan Pór	0.164
27	Japan (S.19)	192	0.16	Utagawa Kuniyoshi	0.053
28	India (S.20)	111	0.06	Bichitr	0.008
30	Armenia (S.20)	97	0.05	Karen Smbatyan	0.005
26	Brazil (S.20)	87	0.07	Agostinho José da Mota	0.011
6	Qing dynasty (S.17)	79	0.03	Qian Du	0.002
8	Bulgaria (S.20)	73	0.07	Daniel Buren	0.013
3	Uruguay (S.20)	72	0.05	Joaquín Torres-García	0.005
23	Mexico (S.20)	66	0.03	David Alfaro Siqueiros	0.002
25	Turkey (S.20)	35	0.02	Burhan Doğançay	0.002
18	Colombia (S.20)	28	0.29	Ramón Torres Méndez	0.073
14	Algeria (S.20)	10	0.01	Choukri Mesli	0.000
13	France (S.20)	9	0.02	Georges Moustaki	0.000
10	Serbia (S.20)	8	0.01	Đorđe Krstić	0.000
9	Iceland (S.19)	4	0.01	Thorstein Elias Hjaltelin	0.000

Aunque no se observa aquí es curioso un fenómeno que sucede en clústers como el 0, donde la zona geográfica ha cambiado de denominación y ordenación política a lo largo del tiempo, y se ve en la tabla referida anteriormente cómo el agrupamiento es inmune a este fenómeno, dándose en el caso mencionado que el 36 % de los autores son de Italia, el 23 % del Reino de Italia y el 10 % de la República de Venecia. Sucede igual con la República Checa, los Países Bajos, Reino Unido, China, Rusia, Japón, Turquía y la India. Con todo lo anterior, todos los clúster representan una frecuencia uniforme en la variable nacionalidad de entre el 80 % y el 95 %.



Vemos cómo el valor medio de eigenvalue y el tamaño del clúster están linealmente correlacionados, excepto en la horquilla entre 500 y 2500 nodos donde se producen bolsas de pintores con demasiada varianza y el valor medio describe la tendencia global.



En el gráfico anterior cada nodo representa a un clúster con la etiqueta de nacionalidad y siglo mayoritario, el tamaño hace referencia al número de nodos que engloba y el color al logaritmo del eigenvalue normalizado entre 0 y 1. Se observa cómo las potencias históricas occidentales dominan esta red social de artistas, con la excepción de Estados Unidos que es posiblemente la nación más reciente y sin embargo compite en número de artistas y centralidad con Alemania, Países Bajos y supera incluso a Reino Unido. Esta instantánea está claramente dominada por Francia, desafiando mi intuición que esperaba que fueran España e Italia —posiblemente por la formación sesgadamente mediterránea que obtenemos en el instituto—. Se aprecia el fuerte intercambio de influencias entre Francia y Estados Unidos en el siglo 19, fenómeno recogido históricamente como inspiradores de la voluntad secesionista norteamericana. También vemos cómo tenemos dos Rusias, la anterior a la revolución comunista, seducida por la fuerte influencia francesa, como el resto de países europeos en el siglo XIX —pues era tomada como referente de modernidad e ilustración cultural—; y la Rusia comunista del siglo 20, totalmente aislada del resto del mundo excepto por su herencia histórica. España por otro lado, se mantiene en un discretísimo

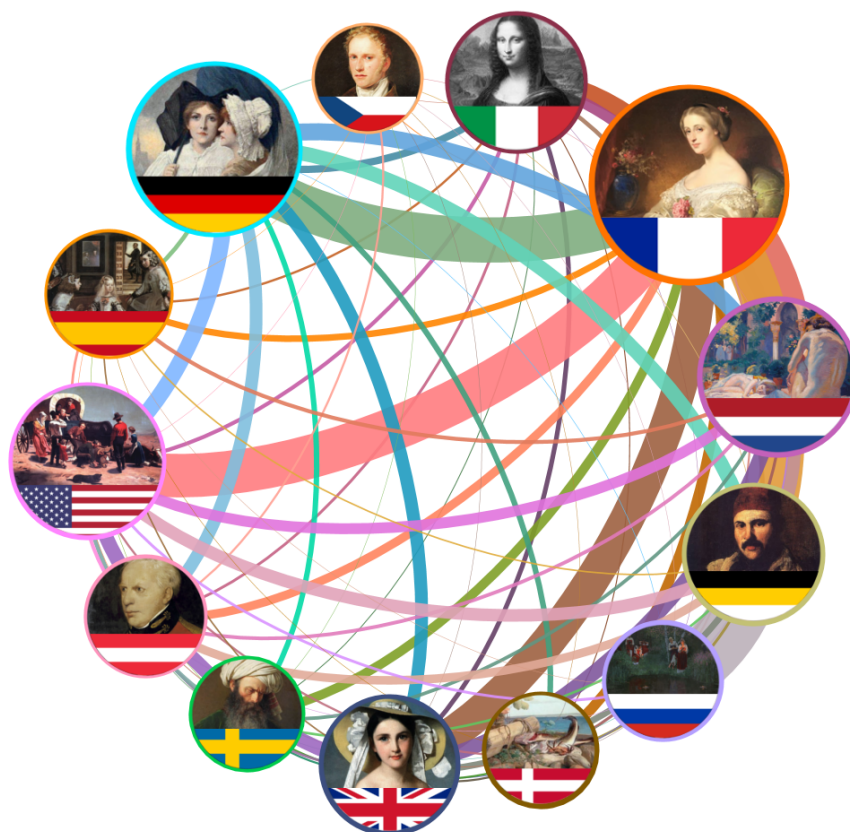
segundo plano, y casi que el único elemento destacable es que para sorpresa de nadie, el autor de mayor centralidad español es Diego Velázquez (eigenvalue de 0.12), aunque me sorprendió ver cómo el valor de centralidad de vector propio de Francisco de Goya era tan sólo 0.013, estoy convencido de que esto es debido a la primacía temporal del primero, pues la influencia del segundo —así como de Velázquez— en el arte del siglo XX es de sobra reconocida.

La red social está fuertemente dominada —como ya adelantaba— por el arte del siglo XIX y XX y no se aprecian las influencias artísticas que recorren el tiempo desde el pasado hacia el futuro, esencialmente por dos motivos: en primer lugar, el movimiento artístico es sólo uno de todos los atributos analizados, responsable únicamente de 1/18 del máximo peso posible de un enlace, mientras que los demás factores, especialmente geográficos y espaciales son mucho más numerosos; y en segundo lugar, la frecuencia de pintores con el movimiento incluido entre sus atributos está entre el 5 % y el 10 %.



Por último, quién demonios es Édouard Dubufe y por qué es la figura clave de esta red de artistas, con mayor *eigenvalue* y de los mayores respecto a centralidad de grado. Édouard-Louis Dubufe no es ni de lejos la figura más relevante de la historia del arte francés ni mucho menos de la historia del arte global, por lo que he podido analizar, vivió durante la época de mayor densidad de pintores en nuestra red (1819-1883), padre e hijo de artistas de relevancia similar a la suya, trabajó en París, Londres e Italia, conectando así con autores de diversísimas nacionalidades. Estudió en la Escuela Nacional Superior de Bellas Artes de París, donde estudiaron el 45 % de los autores franceses con esta etiqueta y fue alumno de otro autor de relevancia que compartió con otros muchos autores formados en Francia así como tutor. En definitiva, si bien su relevancia artística no ha trascendido en exceso, en esta red social que analizamos supo *ser el perezjil de todas las salsas*.

La siguiente gráfica resume lo analizado hasta ahora



En esta gráfica, quizás la más intuitiva para la gente sin formación en nuestra materia observamos los clústers, hemos conservado únicamente los 13 con mayor valor de centralidad de vector propio. En éstos el tamaño del círculo hace referencia al número de artistas que engloba —al igual que en la gráfica anterior—, el color es sencillamente un color distintivo para cada grupo y en cada nodo vemos la bandera del país dominante en la parte inferior y un cuadro del autor con mayor eigenvalue en la parte superior, he decidido utilizar una representación circular por dos motivos, un número tan reducido de nodos luce bien en este *layout*, y en segundo lugar, se da mucho protagonismo al peso de las conexiones, teniendo todo el espacio central para expandirse, que es al fin y al cabo el elemento con carga semántica de mayor interés de este análisis que nos hemos propuesto.

5. Reflexiones finales

Estoy muy satisfecho con los resultados, hemos podido extraer conclusiones que de alguna manera reflejan los eventos históricos que conocemos a pesar de las dificultades que presentaban los datos. Sin embargo hemos ignorado ciertos elementos que han podido introducir un sesgo en nuestro análisis y que sólo mencionamos superficialmente en el apartado relativo a la toma de los datos, en especial los referentes a la diferencia tan radical de centralidad de grado y de vector propio en los autores franceses, especialmente en aquellos que estudiaron en la Escuela Nacional Superior de Bellas Artes de París; y no he podido encontrar ninguna referencia la respecto pero me pregunto si los resultados obtenidos no se deberán en parte a alguna iniciativa particular francesa, que haya automatizado el proceso de rellenar de información las fichas en Wikidata de los autores que pasaron por este centro, pues en un ojeado superficial me da la impresión de que está más ricamente documentado que otros autores de relevancia similar de nacionalidades distintas. En definitiva, sea o no cierto lo anterior, nuestra red social es vulnerable no sólo a la red real subyacente que describimos, sino a la riqueza con la que se han documentado los autores en *Wikidata*, y es que una documentación más rica generará mayor número de conexiones y mayor valor de éstas, y no tenemos ninguna herramienta para controlar este sesgo.

Por último, dejamos en el tintero introducir análisis que contemplen la componente temporal y hagan una lectura distinta y quizás más interesante de la red que tenemos entre manos, así como reevaluar estas relaciones en una red donde no todos los enlaces sean no dirigidos, definiendo quizás enlaces paralelos que reflejen con mayor fidelidad la naturaleza de las relaciones que describen.