

Universidad Nacional de Educación a Distancia  
(UNED)

E.T.S. de Ingeniería Informática — Máster Universitario en Tecnologías del  
Lenguaje

# Técnicas Basadas en Grafos Aplicadas al Procesamiento del Lenguaje Natural

## Tarea 1: Análisis de grafos con Gephi

**Autor:** Franly Iris Urbina Franco

**DNI:** 71.997.418-N

**Correo:** furbina7@alumno.uned.es

**Fecha:** 30 de octubre de 2025

# Índice

<b>1. EuroSisGenerale.gexf</b>	<b>4</b>
1.1. Descripción . . . . .	4
1.2. Datos generales . . . . .	4
1.3. Principales características . . . . .	5
1.4. Distribución del grado . . . . .	7
1.4.1. Análisis de la heterogeneidad . . . . .	7
1.5. Clasificación del grafo . . . . .	8
<b>2. RetGraph.gephi</b>	<b>11</b>
2.1. Descripción . . . . .	11
2.2. Datos generales . . . . .	12
2.3. Principales características . . . . .	13
2.4. Distribución del grado . . . . .	13
2.4.1. Análisis de la heterogeneidad . . . . .	14
2.5. Clasificación del grafo . . . . .	15
<b>3. LesMiserables.gexf</b>	<b>17</b>
3.1. Descripción . . . . .	17
3.2. Datos generales . . . . .	17
3.3. Principales características . . . . .	19
3.4. Distribución del grado . . . . .	19
3.4.1. Análisis de la heterogeneidad . . . . .	20
3.5. Clasificación del grafo . . . . .	20
<b>4. BoundaryCountries.gephi</b>	<b>23</b>
4.1. Descripción . . . . .	23
4.2. Datos generales . . . . .	23

4.3. Principales características . . . . .	24
4.4. Distribución del grado . . . . .	25
4.4.1. Análisis de la heterogeneidad . . . . .	27
4.5. Clasificación del grafo . . . . .	27

## Introducción

El presente informe tiene como objetivo analizar distintas redes representadas mediante grafos con el propósito de identificar sus propiedades estructurales y determinar su posible clasificación dentro de los principales modelos de red: *mundo pequeño* (*small-world*), *red aleatoria*, *libre de escala* (*scale-free*) y su *heterogeneidad*. Para cada grafo se ha seguido una metodología sistemática que permite caracterizar su comportamiento y comprender la forma en que se organizan las conexiones entre sus nodos.

En primer lugar, se realiza una descripción general del grafo y se analizan sus **datos generales** obtenidos a partir de la vista general y de la pestaña de estadísticas de *Gephi*, donde se muestran métricas clave como el número de nodos y aristas, el tipo de grafo (dirigido o no dirigido), la ponderación de las aristas, el diámetro de la red, la longitud media de los caminos, el grado medio, el coeficiente medio de clustering y el número de componentes conexas.

En la **distribución del grado**  $P(k)$  de cada red, analizamos la frecuencia con la que los nodos presentan los diferentes niveles de conectividad con una representación gráfica de la distribución de grados generada por *Gephi*. Estos se conseguirá exportando los datos de las aristas y generando una gráfica **logarítmica doble (log–log)** con el fin de comprobar si los datos se ajustan a una ley de potencias del tipo  $P(k) \sim k^{-\gamma}$  para evaluar la tendencia a una red libre de escala.

# 1. EuroSisGenerale.gexf

## 1.1. Descripción

El grafo **EuroSisGenerale** representa la red de relaciones entre asociaciones científicas europeas generada por el proyecto *European Science Information System (EuroSIS)*.

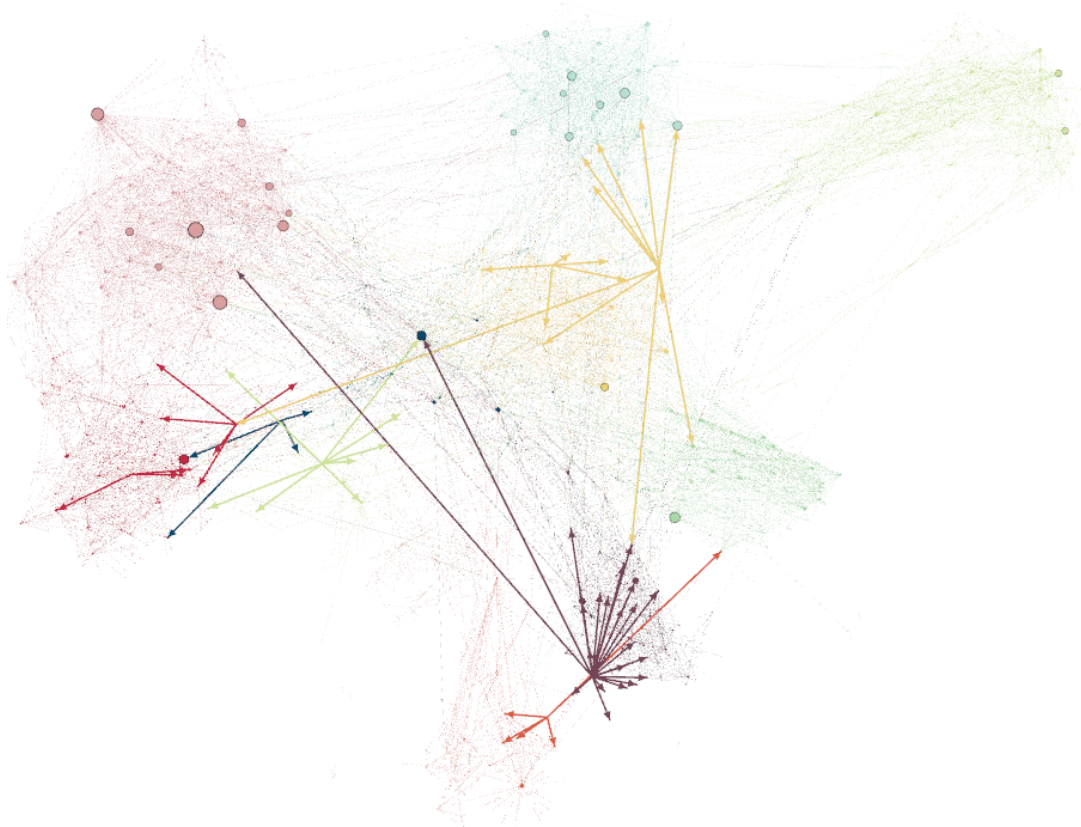


Figura 1: Visualización del grafo EuroSisGenerale en Gephi.

## 1.2. Datos generales

Al realizar un primer acercamiento, podemos ver un grafo dirigido y ponderado que conecta diferentes instituciones universitarias como La Universidad de Bielskok, La Red Europea de Centros de Ciencia y Museos, el Instituto de Chemii Przemyslowej junto con otras instituciones científicas europeas. Las aristas reflejan posibles vínculos de colaboración o dependencia entre ellas. Una cualidad destacable de esta red es su estructura jerárquica, la presencia de unos pocos nodos con un número muy elevado de conexiones sugiere la existencia de instituciones que actúan como puntos intermedios de conexión. Esta configuración favorece una difusión eficiente de información en toda la red, aunque implica dependencia estructural hacia estos nodos principales.

### 1.3. Principales características

Para obtener los datos del grafo que conforman sus principales características, accedemos a dos secciones distintas. La primera la encontraremos en la Vista General, en una pestaña lateral derecha llamada Estadísticas con la que conformaremos la tabla de datos del enunciado, donde podremos hacer click en los botones de 'ejecutar' para generar el resultado.

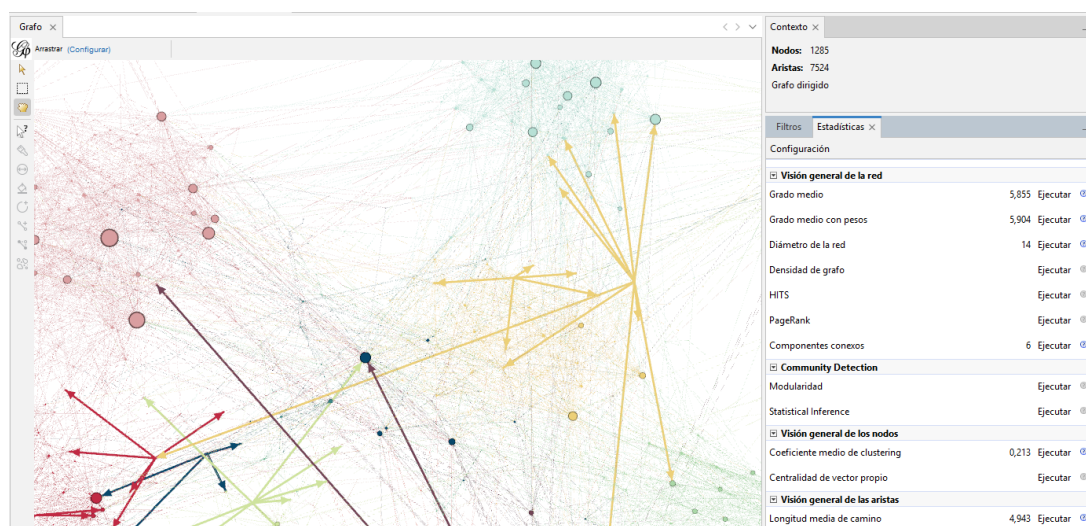


Figura 2: Imagen de datos estadísticos de EuroSisGenerale.gexf.

Para confirmar que se trata de un grafo ponderado, accedemos a la sección 'Laboratorio de datos' y veremos una tabla exportable con los datos principales con los que queramos trabajar. Dentro de estos datos se encuentra la columna 'id' que corresponde a la identificación de la conexión entre los nodos 'Origen' y 'Destino', el tipo de conexión si es dirigida o no dirigida y la columna 'weight', que es el peso que posee las diferentes aristas. Se considera que es un grafo ponderado si contiene valores distintos a 1. El peso predeterminado para las aristas en Gephi es 1, esto significa que si un grafo posee todas sus aristas con dicho valor, se considerará no ponderado y ponderado cuando al menos uno de sus valores posee un valor distinto de 1. Como podemos ver en la siguiente imagen, existen valores que nos permiten confirmar que efectivamente, se trata de un grafo ponderado.

Id	Origen	Destino	Tipo	Weight
8992	1581 - University of Bielsko-Biala, Institute of Textile Engin...	1862 - Beskidzki Festiwal Nauki i Sztuki	Dirigida	2.0
8995	1581 - University of Bielsko-Biala, Institute of Textile Engin...	2097 - Lahti University of Applied Sciences	Dirigida	2.0
8996	1581 - University of Bielsko-Biala, Institute of Textile Engin...	2120 - Lappeenranta University of Technology	Dirigida	2.0
8997	1581 - University of Bielsko-Biala, Institute of Textile Engin...	2267 - University of Oulu	Dirigida	2.0
9001	1581 - University of Bielsko-Biala, Institute of Textile Engin...	2272 - Polish Academy of Sciences	Dirigida	2.0
8999	1581 - University of Bielsko-Biala, Institute of Textile Engin...	2510 - Turku University of Applied Sciences	Dirigida	2.0
8998	1581 - University of Bielsko-Biala, Institute of Textile Engin...	2511 - Tampere University of Technology	Dirigida	2.0
8994	1581 - University of Bielsko-Biala, Institute of Textile Engin...	2530 - Université de nancy	Dirigida	2.0
9000	1581 - University of Bielsko-Biala, Institute of Textile Engin...	2548 - Sofia University 'St. Kliment Ohridski'	Dirigida	2.0
8993	1581 - University of Bielsko-Biala, Institute of Textile Engin...	2556 - University of Debrecen	Dirigida	2.0
10108	1768 - European Network of Science Centres and Museum	1578 - Association of Science and Technology centers	Dirigida	2.0
10109	1768 - European Network of Science Centres and Museum	1697 - Cité de l'espace- Toulouse	Dirigida	2.0
10110	1768 - European Network of Science Centres and Museum	1698 - Cité des Sciences et de l'Industrie	Dirigida	2.0
10107	1768 - European Network of Science Centres and Museum	2193 - Museo Nazionale della scienza e della tecnologia	Dirigida	2.0
11183	1976 - Instytut Chemii Przemysłowej	1518 - Akademia Górniczo-Hutnicza im. Stanisława Sta...	Dirigida	2.0
11180	1976 - Instytut Chemii Przemysłowej	1969 - Polish Hydrogen and Fuel Cell Association Web...	Dirigida	2.0

Figura 3: Tabla de datos de EuroSisGenerale.gexf.

Tras el análisis de ambas secciones podemos concluir con la siguiente tabla las principales características del grafo EuroSiSGenerale.

Dato	Valor
Nodos	1285
Aristas	7524
Grafo dirigido	Sí
Grafo con pesos	Sí
Longitud media de los caminos	4.94
Coeficiente de clustering medio	0.213
Diámetro de la red	14
Grado medio	5.855
Grado medio (ponderado)	5.904
Componentes conexas fuertes (SCC <sup>1</sup> )	511
Componentes conexas débiles (WCC <sup>2</sup> )	6

<sup>1</sup>**SCC**: Strongly Connected Component.

<sup>2</sup>**WCC**: Weakly Connected Component.

Cuadro 1: Métricas de la red EuroSisGenerale.

Una cualidad destacable de esta red es su estructura jerárquica, la presencia de unos pocos nodos con un número muy elevado de conexiones sugiere la existencia de instituciones que actúan como puntos intermedios de conexión. Esta configuración favorece una difusión eficiente de información en toda la red, aunque implica dependencia estructural de estos nodos principales.

## 1.4. Distribución del grado

La probabilidad de que un nodo tomado al azar tenga grado  $k$  se define como:

$$P(k) = \frac{N_k}{N}$$

donde  $N_k$  es el número de nodos con grado  $k$ .

En EuroSiSGenerale, el grado medio es 5.855, lo que indica que la mayoría de los nodos están conectados con alrededor de seis enlaces. La red tiene un número total de 1285 nodos y 7524 aristas, lo que sugiere cierta heterogeneidad.

### 1.4.1. Análisis de la heterogeneidad

La heterogeneidad de una red se refiere a la variación en el número de conexiones (*grado*) que posee cada nodo. En una red homogénea o regular, todos los nodos presentan el mismo grado; mientras que en una red heterogénea, existen nodos con muchos enlaces y otros con muy pocos.

Para cuantificar la heterogeneidad se pueden emplear diferentes métricas, entre las cuales destacan:

- **Varianza o desviación estándar del grado.**

Permite estimar la dispersión de los grados respecto a la media:

$$\sigma_k^2 = \langle k^2 \rangle - \langle k \rangle^2$$

donde  $\langle k \rangle$  es el grado medio y  $\langle k^2 \rangle$  el promedio del cuadrado de los grados. Cuanto mayor sea el valor de  $\sigma_k$ , mayor será la heterogeneidad del grafo.

- **Coefficiente de heterogeneidad de Newman (2003).**

Una medida alternativa basada en la relación entre el segundo y el primer momento del grado:

$$H = \frac{\langle k^2 \rangle}{\langle k \rangle^2}$$

Si  $H = 1$ , la red es completamente homogénea. Valores  $H > 1$  indican la presencia de nodos con grados significativamente distintos, es decir, una red heterogénea.

En el caso de la red **EuroSiSGenerale**, el valor medio de grado de 5,855 y la presencia de nodos con un número muy elevado de conexiones respecto a otros sugieren una estructura heterogénea, coherente con las redes complejas observadas en entornos científicos y de colaboración.



El coeficiente de clustering, la longitud media de camino y el diámetro indica que hay caminos cortos que acortan la red y los clustering locales muestran como grupos vecinos también se conectan entre sí. El grafo posee una gran cantidad de relaciones fuertemente conexas entre entidades pequeñas y solo unas pocas sugieren tener relaciones fuertemente conexas con algunas grandes, lo que hace ver una asimetría entre instituciones centrales que reciben comunicación con entidades pequeñas, pero no siempre hay reciprocidad, tal como se puede esperar de entidades institucionales grandes.

## 1.5. Clasificación del grafo

Las métricas calculadas indican que el grafo presenta un **coeficiente de clustering medio relativamente alto** y una **longitud media de los caminos corta**, lo que sugiere una estructura de tipo **small-world**. Asimismo, la distribución del grado muestra un comportamiento similar a una ley de potencias con exponente entre 2 y 3, característico de una **red libre de escala (scale-free)**.

### *Red Small-World*

La red **EuroSiSGenerale** puede considerarse de tipo *mundo pequeño* porque combina un alto valor medio de *clustering* local de 0.213 con caminos cortos entre nodos distantes de 4.93. En este tipo de redes, aun cuando cada nodo se conecta con un número pequeño de vecinos, se puede alcanzar a otro nodo en pocos pasos a través de intermediarios.

La existencia de nodos altamente conectados, como grandes agencias científicas y universidades, permite que asociaciones o centros más pequeños queden indirectamente conectados.

Este patrón cumple la propiedad característica del *pequeño mundo* descrita por **Easley y Kleinberg**:

$$L \approx L_{rand} \quad \text{y} \quad C \gg C_{rand}$$

EuroSiSGenerale es un ejemplo donde se produce el mismo efecto observado en la red de colaboraciones científicas que inspiró el concepto de los *seis grados de separación*; ya que combina una longitud media de los caminos pequeña con un coeficiente de *clustering* alto en comparación con una red aleatoria.

### Red libre de escala (*scale-free*)

Se define una **red libre de escala** como aquella donde la distribución del grado sigue una ley de potencias:

$$P(k) = k^{-c}$$

con  $c$  entre aproximadamente 2 y 3.

Se puede observar que la red no muestra una distribución uniforme o aleatoria de conexiones, sino una disminución rápida y continua del número de nodos al aumentar el grado, con una cola larga hacia la derecha. Este patrón es coherente con una distribución de tipo potencia, característica de las redes complejas de tipo *scale-free*.

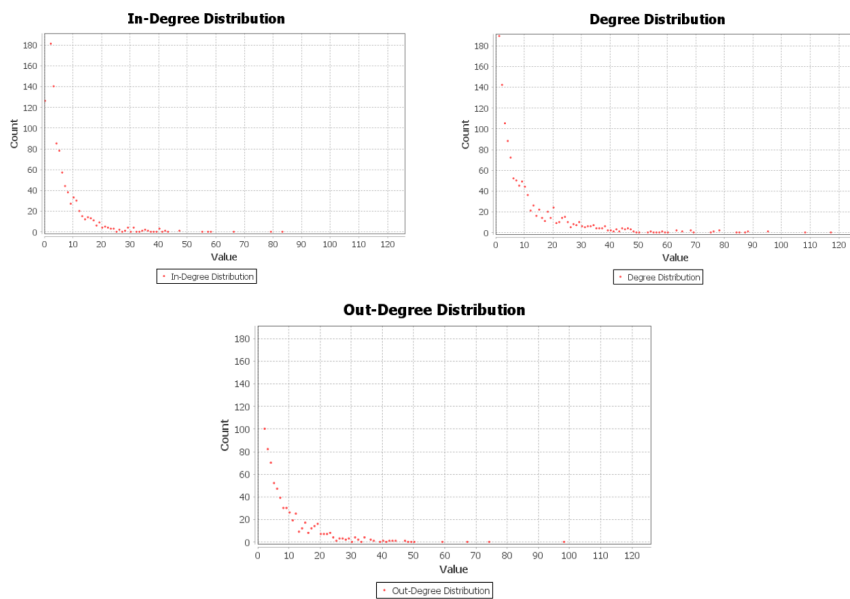


Figura 4: Representación de la distribución de los grados de EuroSisGenerale obtenida por Gephi.

Para examinar la posible naturaleza libre de escala (*scale-free*) de la red EuroSisGenerale, se exportaron los datos del grafo desde *Gephi* en formato `.csv`, obteniendo una tabla con las columnas **Source**, **Target** y **Weight**. A partir de dicha tabla, se procedió a combinar los nodos de origen y destino para obtener una lista única de nodos, sobre la cual se calculó el grado  $k$  de cada nodo, es decir, el número de aristas incidentes en él.

Posteriormente, se determinó la frecuencia  $N(k)$  de cada valor de  $k$ , y la probabilidad asociada  $P(k)$  mediante la expresión:

$$P(k) = \frac{N(k)}{\sum_k N(k)}$$

Con estos valores, se representó la distribución de grados en escala logarítmica doble

(*log-log*) mediante las transformaciones:

$$x = \log_{10}(k) \quad y = \log_{10}(P(k))$$

El análisis de la nube de puntos y la regresión lineal obtenida muestran una relación inversamente proporcional entre  $k$  y  $P(k)$ , coherente con una ley de potencias del tipo:

$$P(k) \sim k^{-c}$$

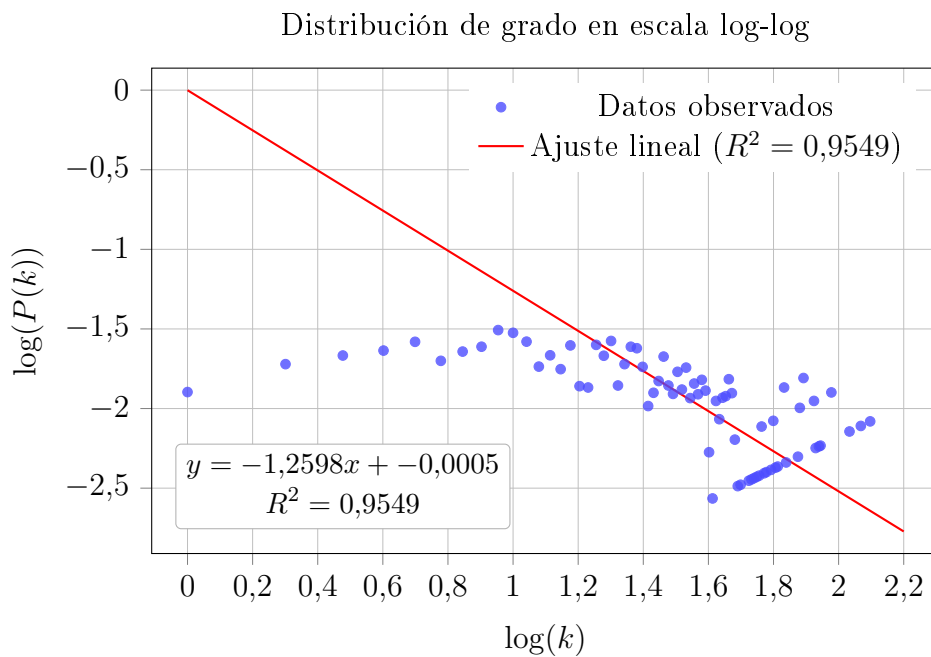


Figura 5: Representación log-log de la distribución de grados del grafo RetGraph. La pendiente obtenida ( $-1,2598$ ) y el coeficiente de determinación ( $R^2 = 0,9549$ ) confirman la existencia de una ley de potencias en la conectividad de la red.

En la Figura 5, se observa una tendencia lineal descendente, con una pendiente de  $-1,2598$  y un coeficiente de determinación  $R^2 = 0,9549$ , lo cual indica un ajuste muy sólido al modelo de ley de potencias. Este resultado confirma que la red presenta una distribución heterogénea de conexiones: la mayoría de los nodos poseen pocos enlaces, mientras que un número reducido de nodos actúa como centros altamente conectados o *hubs*. Por tanto, los resultados reflejan que el grafo EuroSiSGenerale sigue una estructura característica de las redes libres de escala, donde la conectividad no está uniformemente distribuida, sino concentrada en unos pocos nodos clave, lo que confiere a la red propiedades de robustez y vulnerabilidad propias de este tipo de topologías.

## 2. RetGraph.gephi

### 2.1. Descripción

El grafo **RetGraph** representa las co-ocurrencias entre palabras dentro de una misma frase, generadas a partir de un conjunto de documentos.

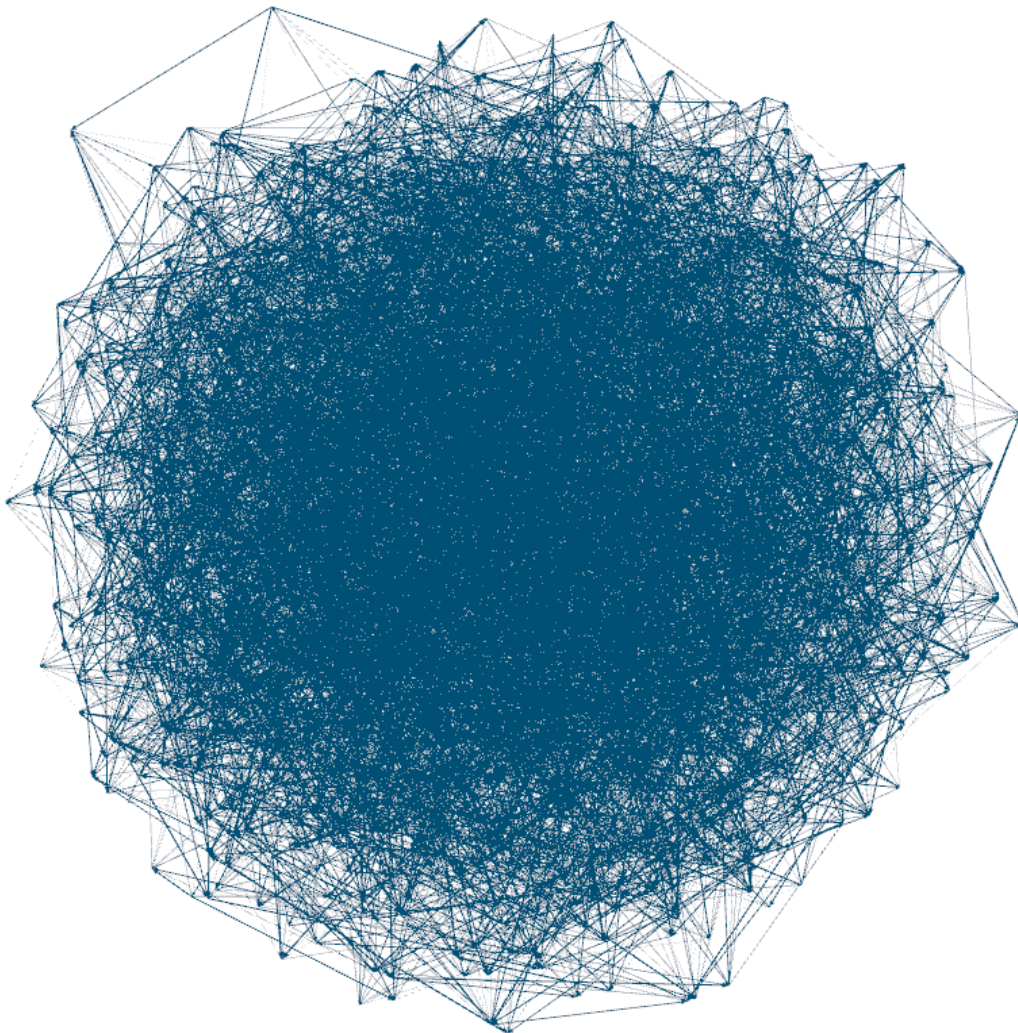


Figura 6: Visualización del grafo RetGraph en Gephi.

Cada nodo corresponde a una palabra y existe una arista dirigida entre dos nodos cuando ambas aparecen conjuntamente en una oración. El peso de cada arista indica la frecuencia de co-ocurrencia, lo que permite identificar asociaciones léxicas, agrupaciones semánticas y la estructura contextual del corpus.

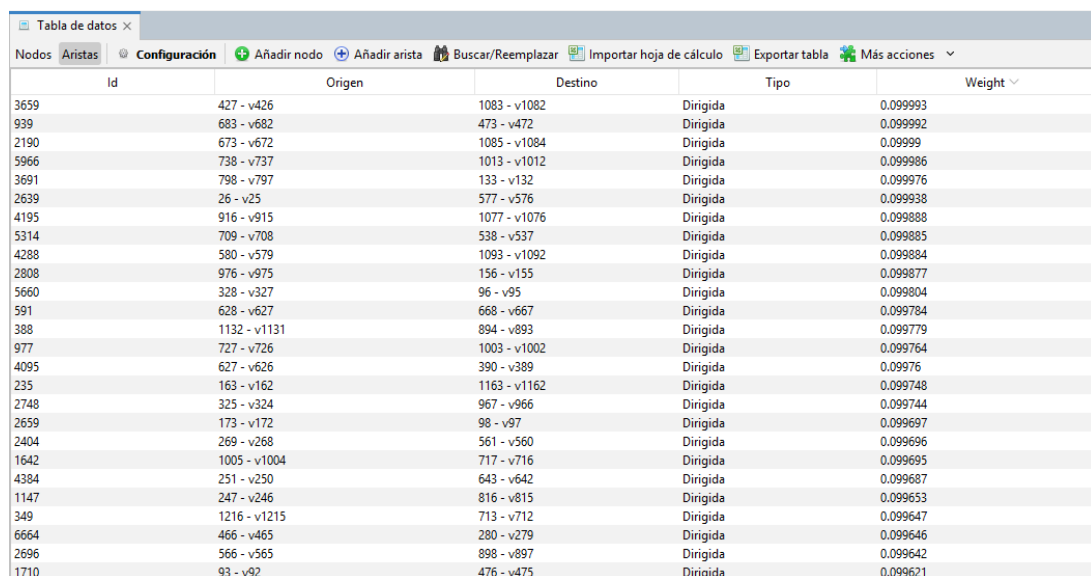
## 2.2. Datos generales

A partir de la ejecución de las estadísticas en Gephi, se han obtenido las siguientes métricas principales de la red:

Dato	Valor
Nodos	1300
Aristas	6708
Grafo dirigido	Sí
Grafo con pesos	Sí
Longitud media de los caminos	3.337
Coefficiente de clustering medio	0.009
Diámetro de la red	5
Grado medio	10.32
Grado medio (ponderado)	0.52
Componentes conexas fuertes (SCC)	No aplicable
Componentes conexas débiles (WCC)	1

Cuadro 2: Métricas generales del grafo RetGraph.

El grafo es ponderado pero no completamente dirigido, se cataloga como mixto al existir aristas que presentan relaciones simétricas, donde la conexión entre dos nodos es mutua.



Id	Origen	Destino	Tipo	Weight
3659	427 - v426	1083 - v1082	Dirigida	0.099993
939	683 - v682	473 - v472	Dirigida	0.099992
2190	673 - v672	1085 - v1084	Dirigida	0.09999
5966	738 - v737	1013 - v1012	Dirigida	0.099986
3691	798 - v797	133 - v132	Dirigida	0.099976
2639	26 - v25	577 - v576	Dirigida	0.099938
4195	916 - v915	1077 - v1076	Dirigida	0.099888
5314	709 - v708	538 - v537	Dirigida	0.099885
4288	580 - v579	1093 - v1092	Dirigida	0.099884
2808	976 - v975	156 - v155	Dirigida	0.099877
5660	328 - v327	96 - v95	Dirigida	0.099804
591	628 - v627	668 - v667	Dirigida	0.099784
388	1132 - v1131	894 - v893	Dirigida	0.099779
977	727 - v726	1003 - v1002	Dirigida	0.099764
4095	627 - v626	390 - v389	Dirigida	0.09976
235	163 - v162	1163 - v1162	Dirigida	0.099748
2748	325 - v324	967 - v966	Dirigida	0.099744
2659	173 - v172	98 - v97	Dirigida	0.099697
2404	269 - v268	561 - v560	Dirigida	0.099696
1642	1005 - v1004	717 - v716	Dirigida	0.099695
4384	251 - v250	643 - v642	Dirigida	0.099687
1147	247 - v246	816 - v815	Dirigida	0.099653
349	1216 - v1215	713 - v712	Dirigida	0.099647
6664	466 - v465	280 - v279	Dirigida	0.099646
2696	566 - v565	898 - v897	Dirigida	0.099642
1710	93 - v92	476 - v475	Dirigida	0.099621

Figura 7: Tabla de datos de RetGraph.gexf.

En el caso de RetGraph, esto significa que algunas co-ocurrencias entre palabras son bidireccionales ( $A \rightarrow B = B \rightarrow A$ ), mientras que otras reflejan una dependencia de aparición ( $A \rightarrow B$ ) más frecuente en un sentido que en el otro. La coexistencia de ambos tipos de

enlaces implica que el grafo podría captar asociaciones léxicas (palabras que coocurren en igual proporción) como relaciones secuenciales o direccionales (palabras que suelen preceder o seguir a otras en el texto).

El diámetro reducido (5) y la longitud media de camino corta (3.33) confirman que la red está globalmente conectada. Sin embargo, el coeficiente de clustering muy bajo ( $C = 0,009$ ) indica que la cohesión local entre los vecinos de un nodo es escasa, lo que sugiere una conectividad más difusa que organizada en clusters.

## 2.3. Principales características

Como ya hemos visto, las características principales las obtenemos de la vista 'Estadísticas' donde podremos ejecutar las funciones que nos permitirán obtener las métricas de nuestros grafos.

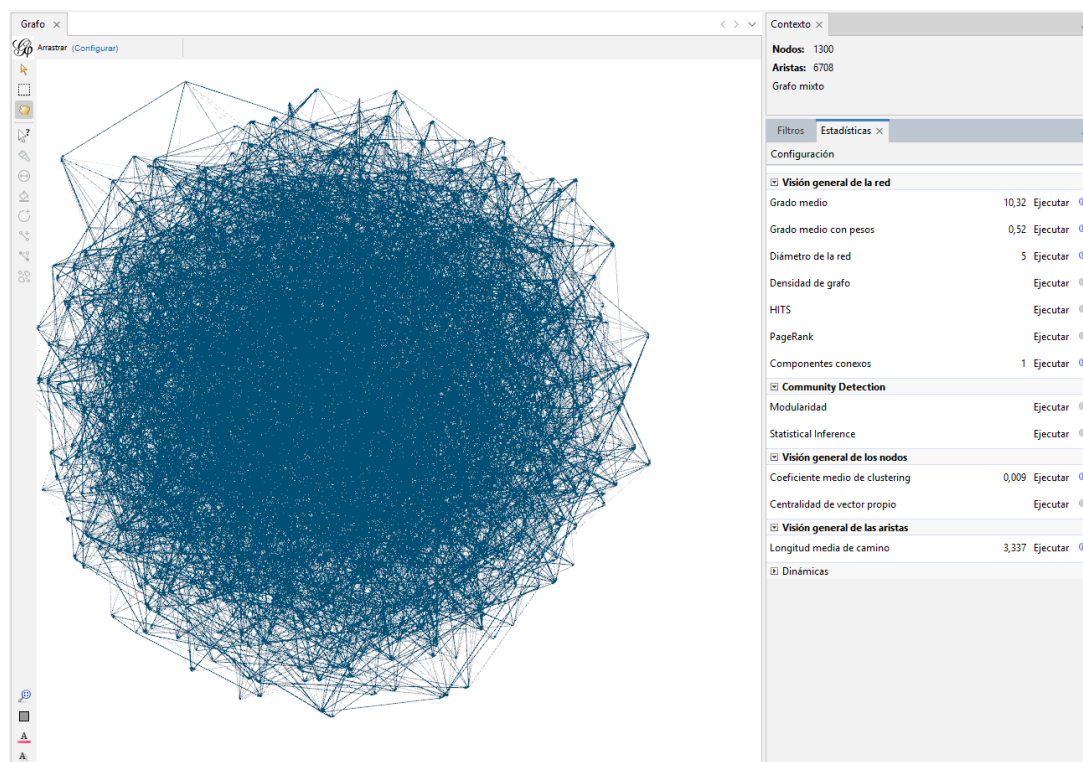


Figura 8: Imagen de datos estadísticos de RetGraph.gexf.

## 2.4. Distribución del grado

La distribución del grado  $P(k)$  describe la probabilidad de que un nodo seleccionado al azar tenga un número  $k$  de conexiones, y se define como:

$$P(k) = \frac{N_k}{N}$$

donde  $N_k$  es el número de nodos con grado  $k$ .

En el caso de un grafo mixto, el grado total de un nodo incluye tanto los enlaces entrantes y salientes (dirigidos) como los no dirigidos, de modo que:

$$k_i = k_i^{in} + k_i^{out} + k_i^{undirected}$$

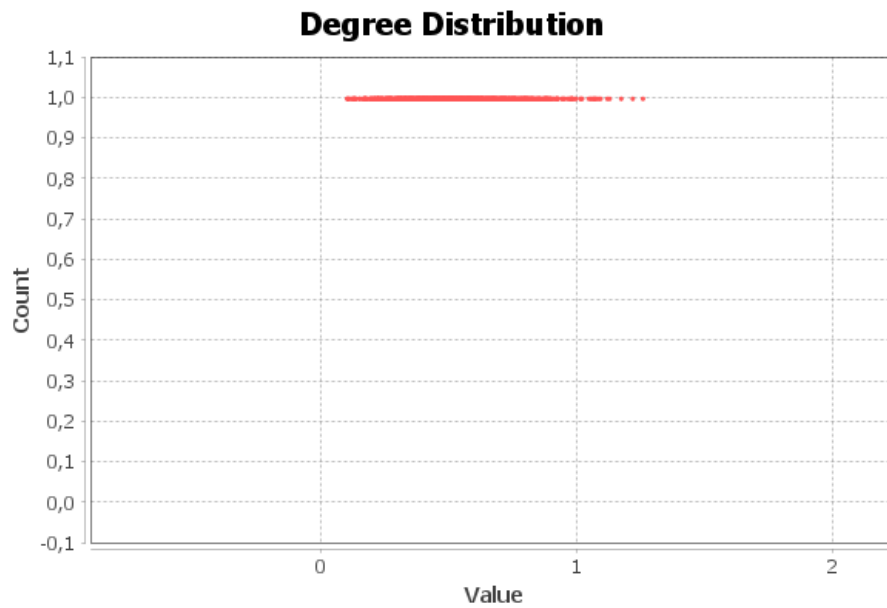


Figura 9: Representación de la distribución de los grados de RetGraph obtenida por Gephi.

El grado medio de 10.32 sugiere que cada palabra está conectada, en promedio, con unas diez más, aunque existen nodos con un grado muy superior correspondientes a términos de alta frecuencia en el corpus. No obstante, el conjunto muestra una fuerte dispersión: algunos nodos (palabras frecuentes) presentan un grado muy superior al promedio, mientras que la mayoría tienen pocos enlaces, lo que evidencia heterogeneidad estructural.

#### 2.4.1. Análisis de la heterogeneidad

La heterogeneidad se puede estimar mediante la varianza del grado:

$$\sigma_k^2 = \langle k^2 \rangle - \langle k \rangle^2$$

y mediante el coeficiente de heterogeneidad de Newman:

$$H = \frac{\langle k^2 \rangle}{\langle k \rangle^2}$$

Valores de  $H > 1$  indican que existen nodos con grados muy dispares. Aunque no se ha calculado explícitamente, las medidas observadas en Gephi y la forma de distribución del grado mostrado anteriormente, nos enseña una dispersión moderada, con algunos nodos fuertemente conectados y la mayoría con pocos enlaces, consistente con una red heterogénea pero no jerárquica.

## 2.5. Clasificación del grafo

Para analizar la posible naturaleza libre de escala de la red, se representó la distribución del grado  $P(k)$  en escala logarítmica doble (log-log).

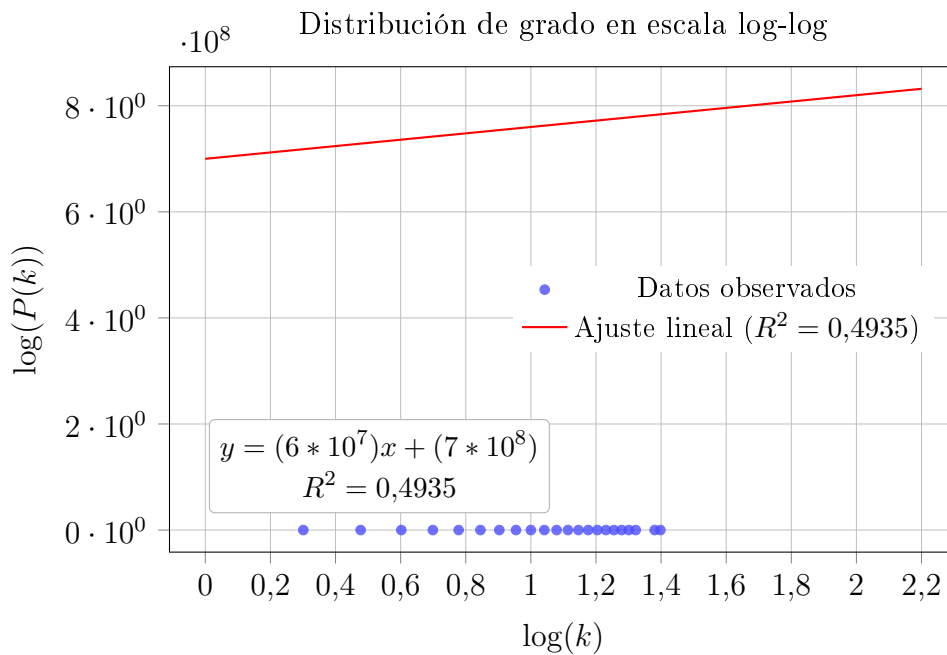


Figura 10: Representación log-log de la distribución de grados del grafo RetGraph. La leyenda se sitúa a la derecha para una mejor visualización de la recta de ajuste.

El coeficiente de determinación  $R^2 = 0,4935$  muestra una correlación moderada, por lo que los datos no se ajustan a una ley de potencias del tipo  $P(k) \sim k^{-c}$ . Aunque existen nodos más conectados que otros, la relación no sigue un patrón sistemático, de modo que no puede confirmarse una estructura libre de escala (*scale-free*) como la observada en el grafo EuroSiSGenerale.

El grafo RetGraph tampoco cumple las propiedades de una red de mundo pequeño, ya que su coeficiente de clustering es muy bajo. Tampoco presenta una ley de potencias clara,



por lo que no puede clasificarse como red libre de escala. En el contexto lingüístico, esto refleja una red donde las palabras se coocurren de forma estadística más que semántica, sin formar comunidades léxicas densas ni jerarquías de conexión pronunciadas.

En consecuencia, el grafo RetGraph se aproxima más a una **red aleatoria dirigida y ponderada**, caracterizada por:

- Caminos cortos entre nodos (longitud media baja).
- Clustering local muy reducido ( $C \approx 0,009$ ).
- Distribución del grado dispersa pero sin dependencia de potencia.

### 3. LesMiserables.gexf

#### 3.1. Descripción

El grafo **LesMiserables** representa la red de coaparición entre los personajes de la novela *Los Miserables* de Victor Hugo. Cada nodo corresponde a un personaje y las aristas indican posiblemente la relación entre ellos. El peso de las aristas puede representar la frecuencia con la que dos personajes coexisten en el texto, por lo que las conexiones más fuertes corresponden a vínculos narrativos más estrechos.

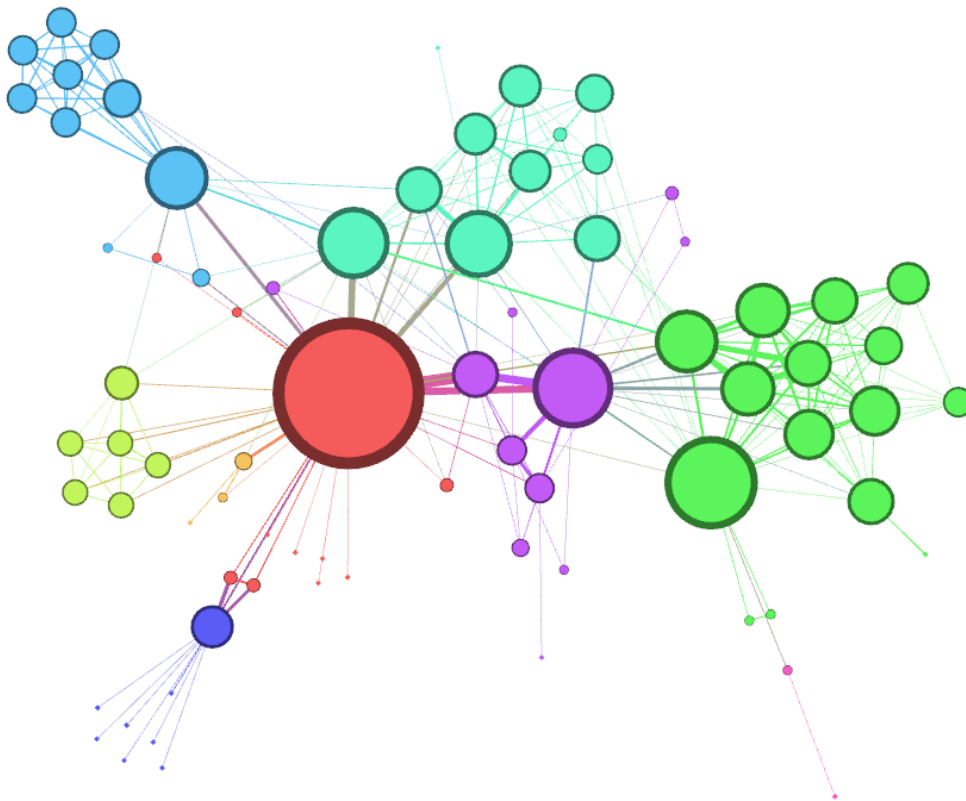


Figura 11: Visualización del grafo LesMiserables en Gephi.

#### 3.2. Datos generales

A partir de la ejecución de las estadísticas en Gephi, se han obtenido las siguientes métricas principales de la red:

Dato	Valor
Nodos	77
Aristas	254
Grafo dirigido	No
Grafo con pesos	Sí
Longitud media de los caminos	2.641
Coefficiente medio de clustering	0.736
Diámetro de la red	5
Grado medio	6.597
Grado medio (ponderado)	21.299
Componentes conexas fuertes (SCC)	No aplicable
Componentes conexas débiles (WCC)	1

Cuadro 3: Métricas generales del grafo LesMiserables.

El grafo es **no dirigido y ponderado**, lo que implica que las conexiones entre personajes son recíprocas y poseen un peso asociado a la frecuencia de interacción. La red presenta una única componente conexas, por lo que todos los personajes están relacionados directa o indirectamente dentro de la misma estructura narrativa. El diámetro bajo (5) y la longitud media de camino (2.641) indican que cualquier personaje puede conectarse con otro mediante pocos intermediarios, lo que refleja una red social compacta. El coeficiente de clustering muy alto ( $C = 0,736$ ) señala una fuerte tendencia de los personajes a formar grupos cerrados o comunidades locales dentro de la trama.

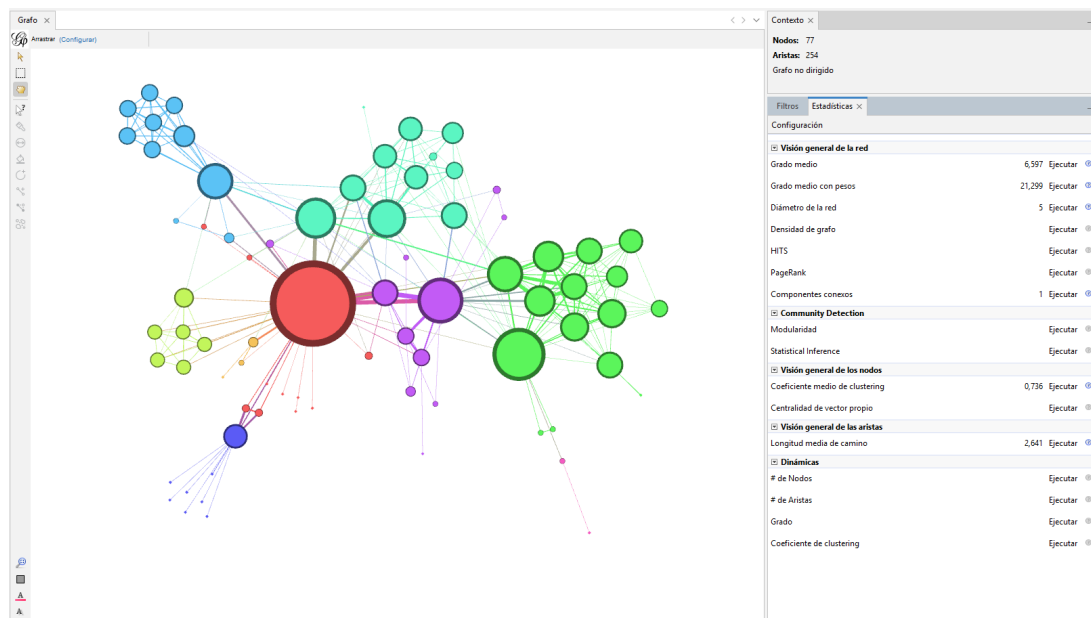
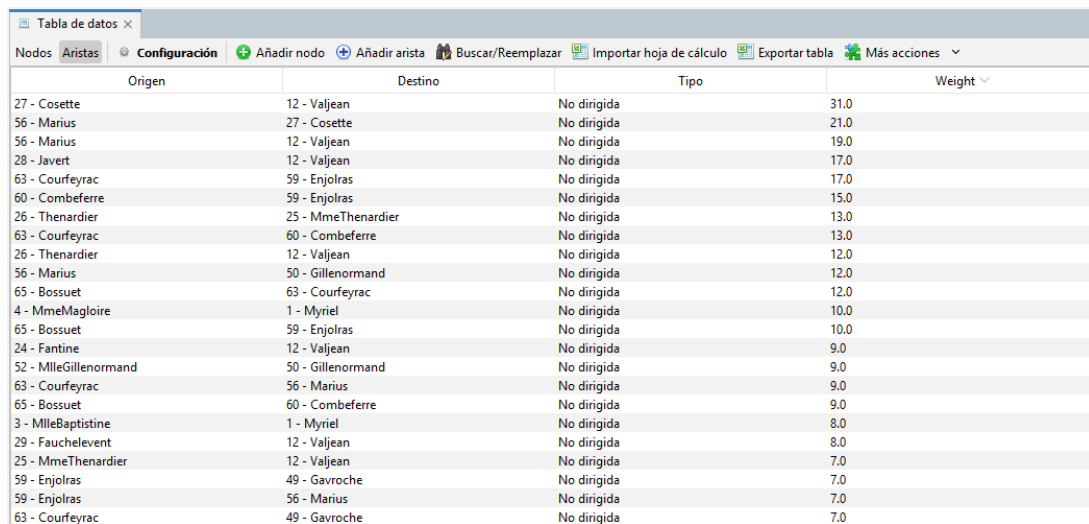


Figura 12: Imagen de datos estadísticos de LesMiserables.gexf.

### 3.3. Principales características

En la vista de estadísticas de Gephi se han ejecutado las métricas que conforman las características principales del grafo. Dado que las aristas poseen diferentes pesos, se confirma que el grafo es ponderado. Cada valor de peso refleja la frecuencia de interacción entre los personajes, por lo que una mayor ponderación indica una relación más estrecha o recurrente.



Origen	Destino	Tipo	Weight
27 - Cosette	12 - Valjean	No dirigida	31.0
56 - Marius	27 - Cosette	No dirigida	21.0
56 - Marius	12 - Valjean	No dirigida	19.0
28 - Javert	12 - Valjean	No dirigida	17.0
63 - Courfeyrac	59 - Enjolras	No dirigida	17.0
60 - Combeferre	59 - Enjolras	No dirigida	15.0
26 - Thenardier	25 - MmeThenardier	No dirigida	13.0
63 - Courfeyrac	60 - Combeferre	No dirigida	13.0
26 - Thenardier	12 - Valjean	No dirigida	12.0
56 - Marius	50 - Gillenormand	No dirigida	12.0
65 - Bossuet	63 - Courfeyrac	No dirigida	12.0
4 - MmeMagloire	1 - Myriel	No dirigida	10.0
65 - Bossuet	59 - Enjolras	No dirigida	10.0
24 - Fantine	12 - Valjean	No dirigida	9.0
52 - MilleGillenormand	50 - Gillenormand	No dirigida	9.0
63 - Courfeyrac	56 - Marius	No dirigida	9.0
65 - Bossuet	60 - Combeferre	No dirigida	9.0
3 - MilleBaptistine	1 - Myriel	No dirigida	8.0
29 - Fauchelevent	12 - Valjean	No dirigida	8.0
25 - MmeThenardier	12 - Valjean	No dirigida	7.0
59 - Enjolras	49 - Gavroche	No dirigida	7.0
59 - Enjolras	56 - Marius	No dirigida	7.0
63 - Courfeyrac	49 - Gavroche	No dirigida	7.0

Figura 13: Tabla de datos de LesMiserables.gexf.

### 3.4. Distribución del grado

La probabilidad de que un nodo tomado al azar tenga grado  $k$  se define como:

$$P(k) = \frac{N_k}{N}$$

donde  $N_k$  es el número de nodos con grado  $k$ .

El grado medio del grafo es 6,597, lo que indica que cada personaje se relaciona en promedio con unos seis o siete más. Sin embargo, el grado ponderado medio (21,299) muestra que la intensidad de las conexiones varía considerablemente, ya que algunos personajes parecen relacionarse con mayor frecuencia. En la siguiente figura se observa que la mayoría de los nodos tienen bajo grado y sólo unos pocos presentan un número elevado de conexiones, lo que refleja una red social con clara desigualdad en las relaciones.

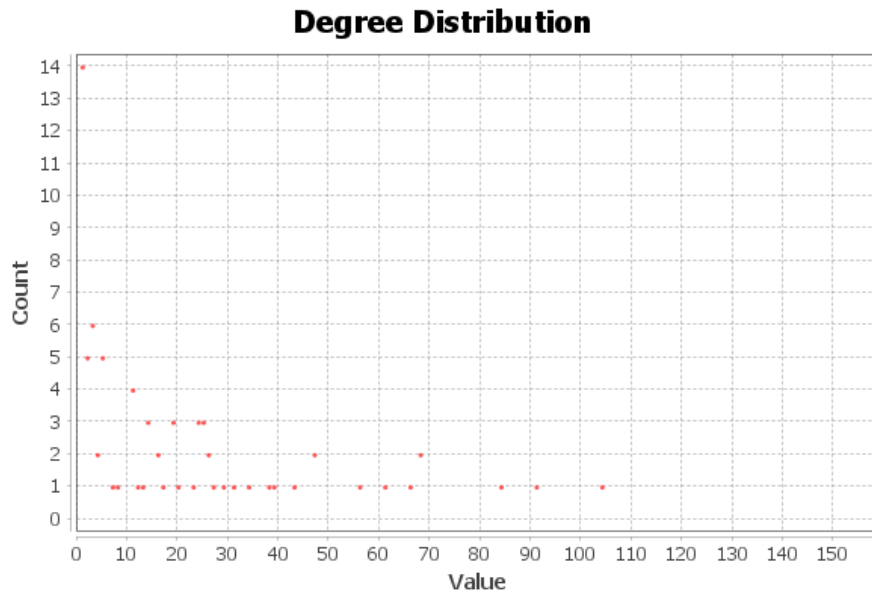


Figura 14: Distribución de grados del grafo LesMiserables obtenida por Gephi.

### 3.4.1. Análisis de la heterogeneidad

La heterogeneidad de la red puede expresarse mediante la varianza del grado:

$$\sigma_k^2 = \langle k^2 \rangle - \langle k \rangle^2$$

y mediante el coeficiente de heterogeneidad de Newman:

$$H = \frac{\langle k^2 \rangle}{\langle k \rangle^2}$$

En redes sociales, valores elevados de  $H$  implican la existencia de personajes con muchos más enlaces que la media, es decir, de individuos centrales o protagonistas. La forma de la distribución observada confirma una dispersión importante de los grados, coherente con una red heterogénea donde algunos personajes desempeñan un papel estructural clave.

## 3.5. Clasificación del grafo

El grafo **LesMiserables** muestra una estructura social coherente con los patrones descritos por el modelo de *mundo pequeño*. La elevada cohesión local ( $C = 0,736$ ) combinada con caminos globales cortos ( $L = 2,641$ ) evidencia la existencia de comunidades densas representadas por los grupos de personajes más cercanos entre sí, unidas por nodos clave de alta conectividad. Esta topología refleja fielmente la estructura narrativa de la novela, donde unos pocos personajes principales actúan como nexo entre tramas secundarias,

asegurando la cohesión global de la red de relaciones y el sentido de una trama continua.

### *Red Small-World*

El grafo cumple las condiciones de una red de *mundo pequeño*, ya que combina un coeficiente de clustering alto ( $C = 0,736$ ) con una longitud media de los caminos corta ( $L = 2,64$ ). Esto implica que los personajes forman comunidades densas localmente conectadas, mientras que algunos personajes principales actúan como intermediarios entre otros. La estructura refleja la organización típica de las redes sociales reales, donde los subgrupos están unidos por conexiones interpersonales fuertes y por intermediarios que enlazan distintas comunidades.

### *Red libre de escala (scale-free)*

Para examinar si la red sigue una ley de potencias, se representó la distribución del grado  $P(k)$  en escala logarítmica doble (log-log).

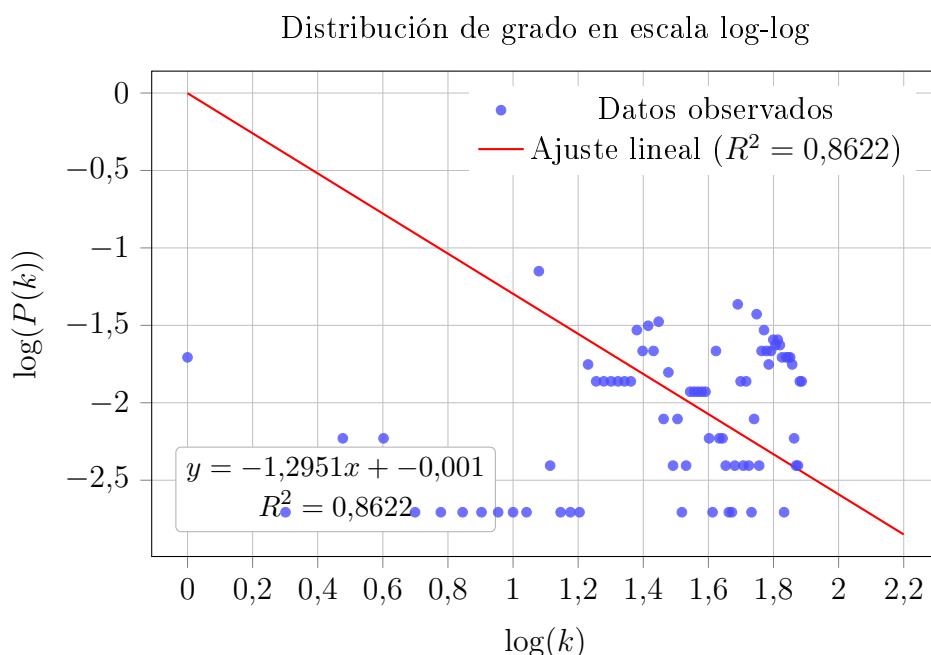


Figura 15: Representación log-log de la distribución de grados del grafo LesMiserables. La pendiente obtenida ( $-1,2951$ ) y el coeficiente de determinación ( $R^2 = 0,8622$ ) confirman la existencia de una ley de potencias en la conectividad de la red.

El coeficiente de determinación  $R^2 = 0,8622$  muestra un ajuste considerable al modelo de potencia, lo que indica que la mayoría de los personajes tienen pocas conexiones, mientras

que unos pocos concentran un gran número de relaciones. Aunque el exponente no alcanza valores típicos de redes masivas, la tendencia es coherente con un comportamiento *scale-free*, donde existen nodos centrales que sostienen la conectividad global. Por tanto, el grafo puede considerarse una **red de mundo pequeño y parcialmente libre de escala**, lo que refleja la estructura social narrativa del texto y un ejemplo de red social.

## 4. BoundaryCountries.gephi

### 4.1. Descripción

El grafo **BoundaryCountries** fue generado en el marco de un proyecto europeo cuyo objetivo era asignar países a distintos distritos administrativos o de gestión fronteriza. Cada nodo representa un país o distrito, mientras que las aristas indican una relación de asignación o pertenencia entre ellos. Todas las aristas poseen el mismo peso, por lo que se trata de un grafo no ponderado. Su estructura refleja relaciones jerárquicas o de dependencia entre entidades, evidenciando un flujo direccional claro entre niveles administrativos.

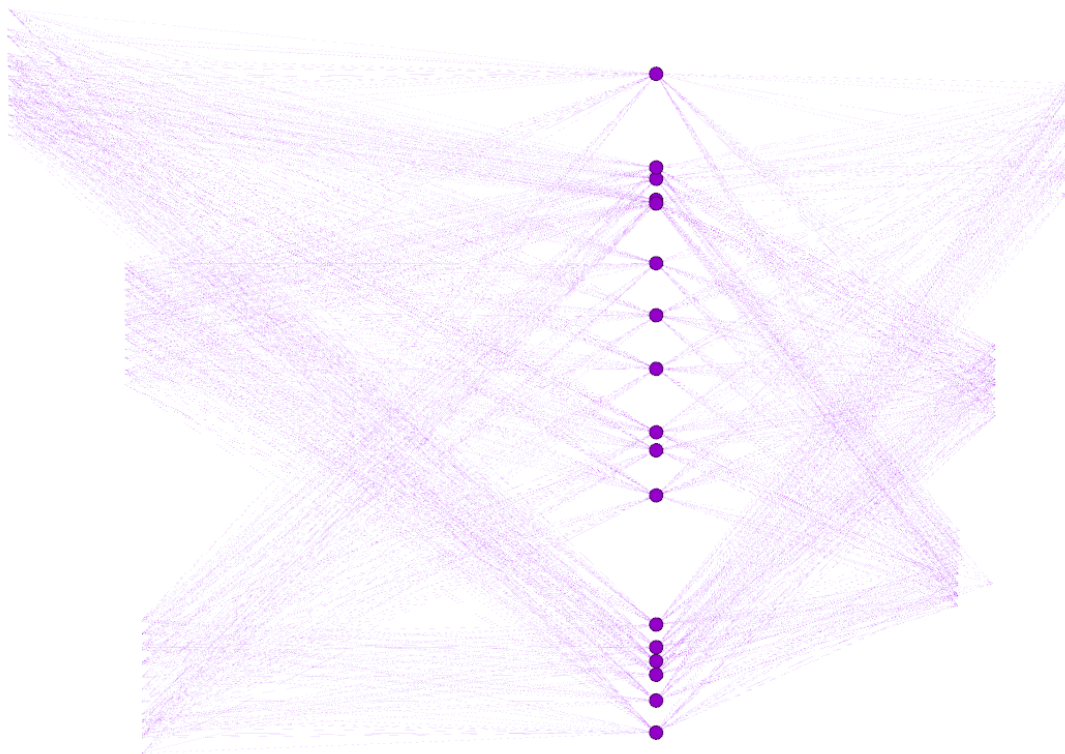


Figura 16: Visualización del grafo BoundaryCountries en Gephi.

### 4.2. Datos generales

A partir de la ejecución de las estadísticas en Gephi, se han obtenido las siguientes métricas principales de la red:



Dato	Valor
Nodos	146
Aristas	1290
Grafo dirigido	Sí
Grafo con pesos	No
Grado medio	8.836
Grado medio (ponderado)	8.836
Diámetro de la red	1
Longitud media de los caminos	1
Coefficiente medio de clustering	0
Componentes conexas fuertes (SCC)	1
Componentes conexas débiles (WCC)	1

Cuadro 4: Métricas generales del grafo BoundaryCountries.

El grafo es completamente **dirigido y no ponderado**, lo que indica que las conexiones poseen sentido de dirección (origen–destino) pero no varían en intensidad. La existencia de una única componente conexas, tanto fuerte como débil, implica que todos los nodos están conectados entre sí de manera directa o indirecta. El diámetro y la longitud media de los caminos iguales a uno reflejan que la red es extremadamente densa, donde prácticamente todos los nodos están interconectados sin intermediarios.

Origen	Destino	Tipo	Weight
164 - Azerbaijan	150 - 4	Dirigida	1.0
164 - Azerbaijan	151 - 5	Dirigida	1.0
164 - Azerbaijan	152 - 6	Dirigida	1.0
164 - Azerbaijan	153 - 7	Dirigida	1.0
164 - Azerbaijan	154 - 8	Dirigida	1.0
164 - Azerbaijan	155 - 9	Dirigida	1.0
164 - Azerbaijan	156 - 10	Dirigida	1.0
164 - Azerbaijan	157 - 11	Dirigida	1.0
164 - Azerbaijan	158 - 12	Dirigida	1.0
164 - Azerbaijan	159 - 13	Dirigida	1.0
165 - 19	151 - 5	Dirigida	1.0
165 - 19	152 - 6	Dirigida	1.0
165 - 19	153 - 7	Dirigida	1.0
165 - 19	154 - 8	Dirigida	1.0
165 - 19	155 - 9	Dirigida	1.0
165 - 19	156 - 10	Dirigida	1.0
165 - 19	157 - 11	Dirigida	1.0

Figura 17: Tabla de datos de BoundaryCountries.gexf.

### 4.3. Principales características

Como puede observarse en la vista de estadísticas de Gephi, los valores de grado medio y grado medio ponderado coinciden, confirmando que el grafo no es ponderado. El coeficiente medio de clustering es nulo, lo que significa que no existen triángulos formados entre los nodos; es decir, si un país A está conectado con B y B con C, no necesariamente existe una conexión directa entre A y C. Esto indica ausencia de comunidades locales o agrupaciones cerradas dentro de la red.

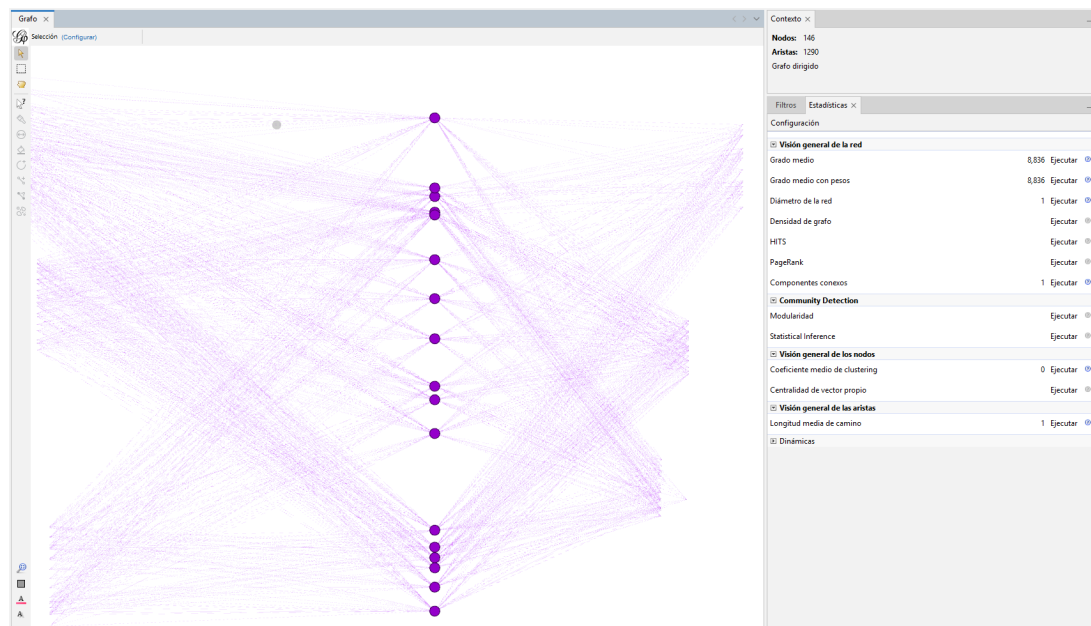


Figura 18: Imagen de datos estadísticos de BoundaryCountries.gexf.

#### 4.4. Distribución del grado

La distribución del grado  $P(k)$  describe la probabilidad de que un nodo seleccionado al azar tenga un número  $k$  de conexiones, definida por:

$$P(k) = \frac{N_k}{N}$$

donde  $N_k$  es el número de nodos con grado  $k$ .

En este grafo, el grado medio es 8,836, y la distribución del grado muestra que la mayoría de los nodos tienen un número similar de conexiones, con pocos nodos que alcanzan valores más elevados. Este comportamiento refleja una red bastante homogénea, donde la conectividad está distribuida de forma uniforme entre los países o distritos.

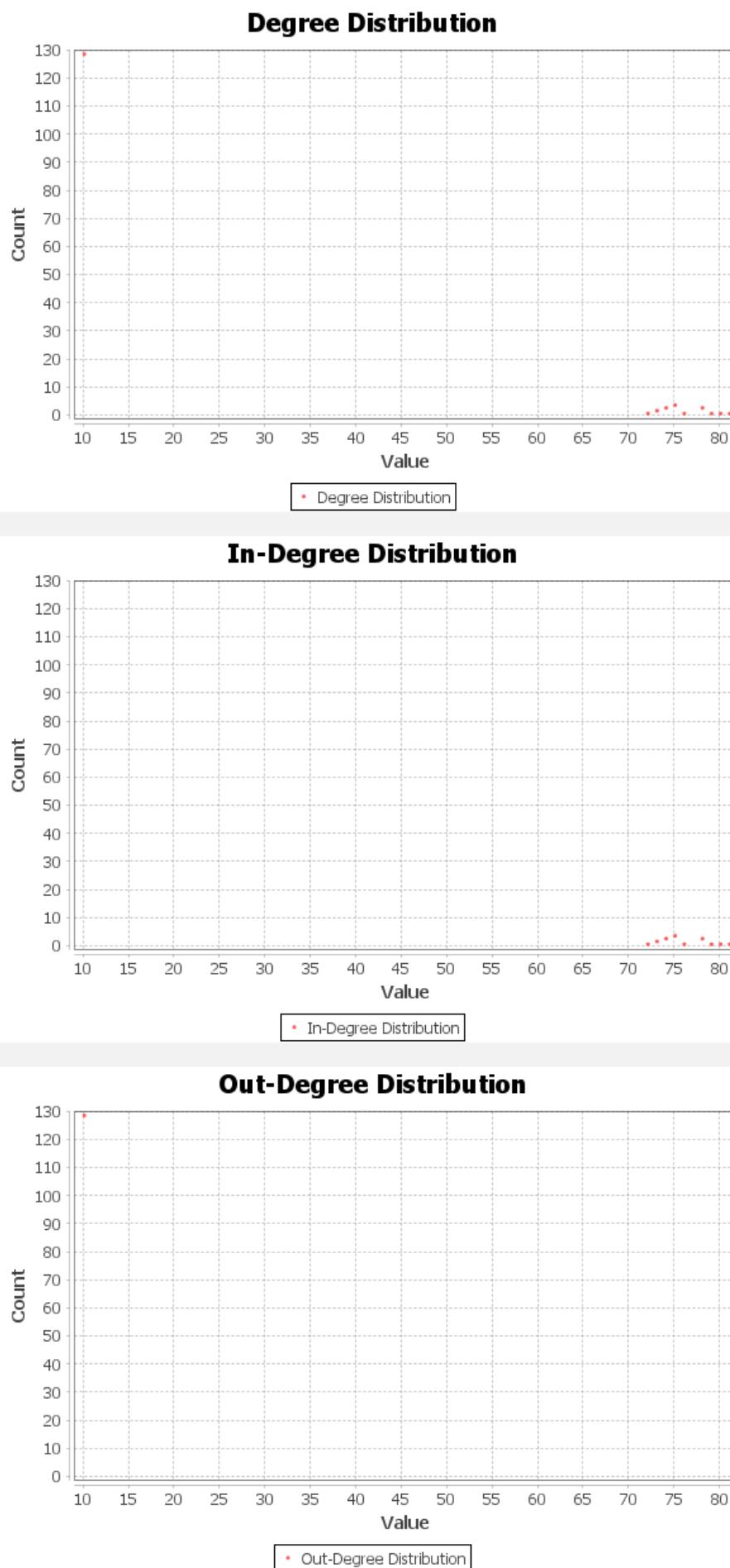


Figura 19: Distribución de grados del grafo BoundaryCountries obtenida por Gephi.

#### 4.4.1. Análisis de la heterogeneidad

La heterogeneidad se puede estimar mediante la varianza del grado:

$$\sigma_k^2 = \langle k^2 \rangle - \langle k \rangle^2$$

y mediante el coeficiente de heterogeneidad de Newman:

$$H = \frac{\langle k^2 \rangle}{\langle k \rangle^2}$$

En este caso, el grafo presenta valores de grado muy próximos entre sí, por lo que la varianza es baja y  $H$  se aproximaría a 1, lo cual corresponde a una red homogénea y estructuralmente regular.

### 4.5. Clasificación del grafo

El grafo **BoundaryCountries** presenta una estructura extremadamente conectada, con caminos mínimos y sin agrupaciones locales pero para que sea de tipo *mundo pequeño*, debe presentar caminos cortos y un alto coeficiente de clustering. En este caso, aunque la longitud media de los caminos ( $L = 1$ ) es mínima, el coeficiente de clustering nulo ( $C = 0$ ) descarta clasificarlo como tal. La red no posee la estructura local característica de las redes de *mundo pequeño*, sino una conectividad puramente direccional y uniforme.

#### *Red libre de escala (scale-free)*

Para analizar su posible naturaleza libre de escala, se representó la distribución de grados en escala logarítmica doble (log-log).

Aunque el ajuste lineal presenta un coeficiente de determinación alto ( $R^2 = 0,9892$ ) y una pendiente cercana a  $-1$ , esta relación se debe al reducido número de valores distintos de grado presentes en el grafo, más que a una verdadera distribución tipo potencia.

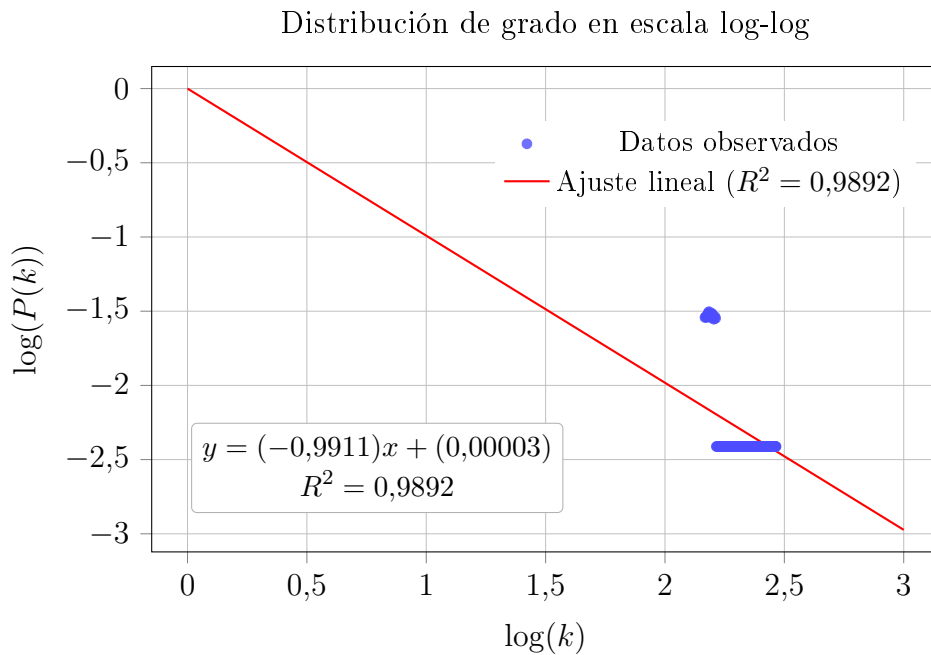


Figura 20: Representación log-log de la distribución de grados del grafo BoundaryCountries.

Una red libre de escala debe cumplir dos condiciones esenciales: (1) que su distribución de grados siga una ley de potencias del tipo  $P(k) \sim k^{-\gamma}$  con  $2 < \gamma < 3$ , y (2) que exista una “cola larga” en la que unos pocos nodos concentren la mayoría de las conexiones. Estas propiedades no se observan en el grafo **BoundaryCountries**, donde la mayoría de nodos poseen grados similares entre 8 y 10, y sólo unos pocos alcanzan valores extremos sin continuidad intermedia.

Su estructura es más coherente con una **red aleatoria densamente conectada**, caracterizada por:

- Un grado medio elevado y homogéneo ( $\langle k \rangle = 8,836$ ).
- Un diámetro mínimo ( $D = 1$ ) y longitud media de camino igualmente mínima ( $L = 1$ ).
- Ausencia de clustering local ( $C = 0$ ).

Estas propiedades describen un grafo fuertemente conectado, regular y no jerárquico, donde todos los nodos mantienen relaciones casi uniformes sin presencia de nodos dominantes o *hubs*, alejándose del comportamiento típico de las redes complejas de tipo *scale-free*.

## Conclusión

El análisis realizado ha permitido comprender mejor la estructura y comportamiento de diferentes tipos de redes a través del uso de *Gephi*. La herramienta ha resultado muy intuitiva y visual, facilitando la exploración de grafos complejos y la interpretación de sus propiedades mediante métricas y gráficas automáticas.

Entre las principales dificultades se encontró la correcta comprensión teórica de los conceptos estudiados, especialmente los relacionados con los temas 2, 13 y 20, que fue necesario complementar con lecturas adicionales para entender con precisión qué caracteriza a una red libre de escala y cómo representar su distribución en escala log-log.

Lo más gratificante del proceso ha sido comprobar cómo grafos de gran tamaño, que en un principio resultaban complejos, se pueden descomponer y analizar con claridad, permitiendo descubrir patrones, relaciones y estructuras que no son evidentes a simple vista.