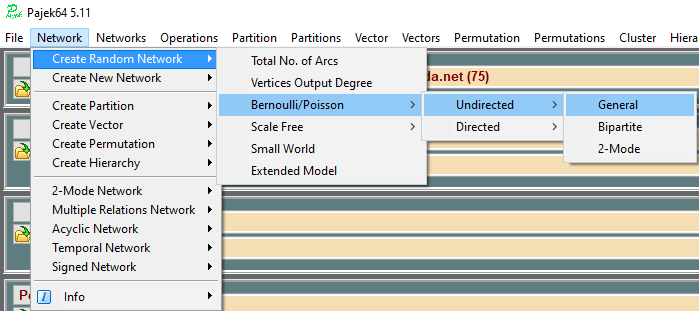
# **PRÁCTICA ANÁLISIS DE REDES SOCIALES**

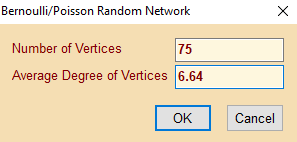
## **1. Obtener los parámetros de una red aleatoria con parámetros *N* y *p* que tenga un número de nodos y aristas esperadas similar a la red *Los Miserables* e indicar la ruta para hacerlo en Pajek**

Para generar los parámetros de una red aleatoria, dado que estamos ante un modelo de red **no dirigido**, utilizaremos la segunda opción disponible en Pajek: *Bernoulli/Poisson*. Para ello, nos dirigimos a *Network*, *Create Random Network, Bernoulli/Poisson, Undirected, General*:



A continuación, debemos introducir tanto el número de nodos como el grado medio de aristas. Dado que se trata de una red no dirigida, el grado medio se obtiene a través de la siguiente expresión:

Donde L es el **número de aristas** y N el **número de nodos**. Por tanto, el grado medio será:



Una vez configurados los parámetros, el modelo de red generado es el siguiente:

**Number of vertices (n): 75**

**----------------------------------------------------------**

**Arcs Edges**

**----------------------------------------------------------**

**Total number of lines 0 248**

**----------------------------------------------------------**

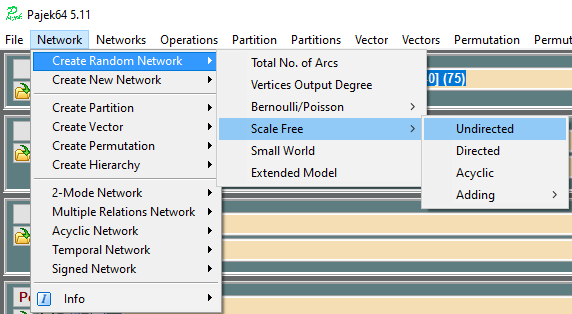
**Number of loops 0 0**

**Number of multiple lines 0 0**

**----------------------------------------------------------**

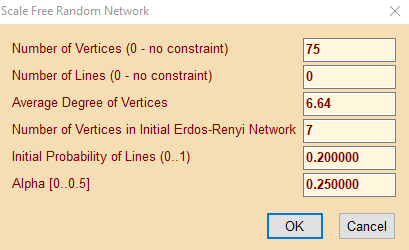
## **2. Obtener los parámetros de una red libre de escala que tenga un número de nodos y aristas esperadas similar a la red *Los Miserables* e indicar la ruta para hacerlo en Pajek**

En relación con una red libre de escala, nos dirigimos nuevamente a *Network*, *Create Random Network*, aunque esta vez debemos seleccionar la opción *Scale Free*, *Undirected*:



Una vez marcada la opción, debemos configurar los siguientes campos:

* *Number of Vertices*: fijamos el número de nodos para poder realizar la comparativa final con el resto de modelos (75).
* *Average Degree of Vertices*: nuevamente debemos introducir el grado medio de nuestra red, concretamente 6.64
* *Number of Vertices in Initial Erdos-Renyi Network* o Número de vértices a calcular de forma aleatoria: por defecto, está establecido a 10. No obstante, y dado que el tamaño de la red es muy pequeño, debemos reducir dicho parámetro. A *grosso modo*, el número de nodos es de 75 (cercano a 100), por lo que asignamos valor superior a 5 (mínimo), pero inferior a 10, concretamente 7:



Una vez configurados los parámetros, el modelo de red generado se muestra a continuación:

**Number of vertices (n): 75**

**----------------------------------------------------------**

**Arcs Edges**

**----------------------------------------------------------**

**Total number of lines 0 240**

**----------------------------------------------------------**

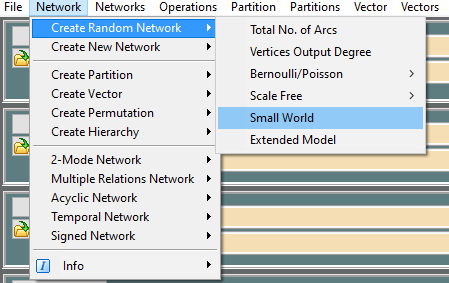
**Number of loops 0 0**

**Number of multiple lines 0 72**

**----------------------------------------------------------**

## **3. Obtener los parámetros de una red de pequeño mundo que tenga un número de nodos y aristas esperadas similar a la red *Los Miserables* e indicar la ruta para hacerlo en Pajek**

Por último, para generar una red de pequeño mundo basta con dirigirnos a *Network*, *Create Random Network*, *Small World*.



Una vez seleccionada la opción debemos configurar los siguientes parámetros:

1. *Number of vertices*: número de vértices de la red (75).
2. *Number of Linked Neighbors on each Side of a Vertex*: con el objetivo de decidir el tamaño de la *comunidad* de cada nodo, elegimos el número de vecinos situados a su izquierda y derecha, es decir, el grado medio.
3. *Replacement Probability* o probabilidad de reemplazo. Para el desarrollo de la práctica escogemos 0.2 como probabilidad de conexión con nodos “exteriores”.

En relación con la elección del número de vecinos, dado que el valor *k* obtenido en anteriores apartados ha sido de 6.64, podemos elaborar un modelo de red eligiendo como grado medio 3 (6.64 / 2 ~ 3):

|  |  |
| --- | --- |
| ***grado medio*** | ***numero de aristas*** |
| 2 | 150 |
| 3 | 225 ~ 249 |
| 4 | 300 |

Como podemos observar, con grado medio 3 (3 nodos tanto a la izquierda como a la derecha), el número de aristas es mucho más cercano a la de la red original (249):

**Number of vertices (n): 75**

**----------------------------------------------------------**

**Arcs Edges**

**----------------------------------------------------------**

**Total number of lines 0 225**

**----------------------------------------------------------**

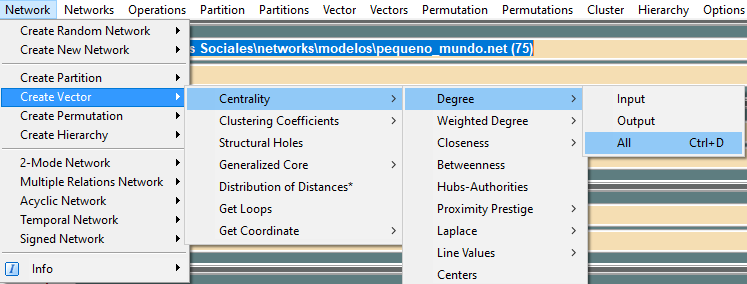
**Number of loops 0 0**

**Number of multiple lines 0 0**

**----------------------------------------------------------**

## **4. Para las cuatro redes (las tres de los apartados anteriores y la original *Los Miserables*), calcular su distribución de grado e indicar si son aleatorias o de libre escala cada una de ellas.**

Una vez creadas las tres redes anteriores, necesitamos estudiar lo importante que es cada nodo en la red, es decir, si está conectado a muchos nodos o no. Para ello, por cada red debemos calcular el grado a través de la opción *Network*, *Create Vector*, *Centrality*, *Degree*, *All* (dado que la red es **no dirigida**):



### **4.1 Red original**

Analizando el modelo de red original, nos encontramos ante una red que se asemeja en mayor medida a un modelo **libre de escala** que a un modelo aleatorio, dado que su distribución de grado **se basa en una ley potencial** en lugar de una distribución normal: en la parte izquierda nos encontramos con una alta concentración de personajes "con menor popularidad", esto es, con un menor número de conexiones; mientras que en la parte derecha se sitúa una larga cola, formado por personajes más populares (con un mayor número de conexiones en la historia con el resto de personajes), aunque en menor número.

No obstante, el decrecimiento en la distribución de grado no sigue exactamente el modelo de ley potencial, asumiendo la lógica variabilidad que presenta entre los distintos miembros, como es el caso de los nodos 6 y 7 (naranja), cuya concentración en la distribución de grado es mayor con respecto a valores de grado anteriores como 4 o 5.

**4.2 Red aleatoria**

Continuando con el modelo aleatorio, podemos distinguir claramente **una distribución "aproximadamente" normal** en los valores de distribución de grado, nuevamente asumiendo la lógica variabilidad que presentan sus miembros, tales como los grados 4 y 5 (naranja).

### **4.3 Red de libre escala**

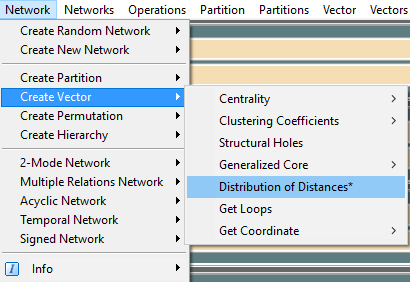
En relación con el modelo de libre escala, podemos comprobar como la distribución de los grados sigue aproximadamente una ley potencial, del mismo modo que la distribución de la red original, salvo por una diferencia: **en la red original la capacidad de atracción no supera a la proporcionalidad, es decir, no se trata de un modelo "más libre de escala"**. A diferencia del modelo de libre escala, la red aleatoria presenta personajes con un menor número de conexiones (máximo de 36 conexiones con otros personajes mientras que la red de libre escala alcanza un máximo de 50). Por ello, dado que existen nodos con un menor grado, dichos nodos no podrán crecer a una mayor velocidad y, como consecuencia, **no presentan una mayor** **capacidad de atracción**.

### **4.4 Red de pequeño mundo**

En el caso de la red de pequeño mundo, podemos identificar una distribución "aproximadamente" normal, incluso con mejor claridad que en el caso de la red aleatoria anterior, con una mayor concentración de los valores de grado en torno a la media. Por tanto, podemos asegurar que se trata de un **modelo de red aleatorio**.

## **5. Para las cuatro redes (las tres de los apartados anteriores y la original *Los Miserables*), calcular su distancia media e indicar si son de pequeño mundo cada una de ellas**

Inicialmente, para calcular la distancia media entre todos los pares alcanzables, seleccionamos la opción *Network*, *Create Vector*, *Distribution of Distances\**:



### **5.1 Red original**

**Number of unreachable pairs: 0**

**Average distance among reachable pairs: 2.62162**

En primer lugar, podemos comprobar que la red no presenta pares inalcanzables, por lo que no será necesario recalcular la distancia media entre los pares: **2.62162**. Por tanto, si calculamos nuevamente el grado medio de la red original:

Y calculamos el valor correspondiente:

Comprobamos que la distancia media es mayor al valor obtenido (2.62 > 2.28), aunque sus valores son aproximadamente iguales, **la cota de diferencia entre ambos no supera el 10 %**. Por tanto, podemos asegurar que **la red original es de pequeño mundo**.

### **5.2 Red aleatoria**

**Number of unreachable pairs: 0**

**Average distance among reachable pairs: 2.46306**

Nuevamente, el número de pares inalcanzables continua siendo cero, por lo que podemos seguir considerando como válida la distancia media obtenida a través de Pajek: **2.46306**. Dado que la red aleatoria presenta un número distinto, aunque aproximadamente igual, de arcos, debemos calcular el grado medio **empleando el número de arcos correspondiente**: 248.

Si calculamos el valor correspondiente:

Nuevamente, aunque la distancia media sea mayor al valor obtenido (aproximadamente similar al obtenido en la red original), el porcentaje de diferencia entre ambos valores no supera el 10 % de cota, por lo que decimos que el modelo de red aleatorio se trata, además, de una **red de pequeño mundo**.

### **5.3 Red de libre escala**

**Number of unreachable pairs: 2010**

**Average distance among reachable pairs: 2.58305**

¿Qué sucede con la red de libre escala? A diferencia del resto de modelos, **existe un elevado número de pares no alcanzables** (2010). De hecho, en la distribución de grado del apartado anterior pudimos comprobar como existen nodos cuyo grado de conexión es cero, esto es, personajes que no entablan ninguna conversación a lo largo de la historia, por lo que muchos de los pares quedarán sin alcanzar. Por ello, debemos **recalcular la distancia media, añadiendo un valor de peso a los pares no alcanzables**. Concretamente, dado que la máxima distancia entre pares es de 6, asignamos un valor de distancia de 7 a los pares no alcanzables:

|  |  |  |
| --- | --- | --- |
| ***Distance*** | ***Distribution of Distances*** | ***Distance\*Distribution*** |
| 1 | 336 | 336 |
| 2 | 1350 | 2700 |
| 3 | 1358 | 4074 |
| 4 | 448 | 1792 |
| 5 | 46 | 230 |
| 6 | 2 | 12 |
| 7 | 2010 | 14070 |
| **Total** | **5550** | **23214** |

|  |  |  |
| --- | --- | --- |
| Number | Distribution of Distances in N4 (6) |  |
| 1 | 336 | 336 |
| 2 | 1350 | 2700 |
| 3 | 1358 | 4074 |
| 4 | 448 | 1792 |
| 5 | 46 | 230 |
| 6 | 2 | 12 |
| 7 | 2010 | 14070 |

Por tanto, **el nuevo valor de distancia media será**:

Una vez calculado su valor, debemos obtener el grado medio empleando el número de arcos correspondiente a la red de libre escala (240):

Así como el valor correspondiente:

Observamos que la distancia media es **estrictamente mayor al valor obtenido** (6.4 >> 2.33), por lo que diremos que **se trata de un modelo que está lejos de ser una red de pequeño mundo**.

### **5.3 Red de pequeño mundo**

**Number of unreachable pairs: 0**

**Average distance among reachable pairs: 2.81297**

Por último, en el caso de la red de pequeño mundo volvemos a tener todos los pares de nodos alcanzables, por lo que podemos emplear la distancia media obtenida en Pajek: **2.81297**. Por tanto, si calculamos el grado medio con el número de arcos correspondiente (225):

Así como su valor correspondiente:

De nuevo, la cota de diferencia entre ambos valor no supera el 10 %, por lo que podemos asegurar que, efectivamente, **se trata de una red de pequeño mundo**.

## **6. Indicar la medida de centralidad que nos parece mejor en este caso y la motivación de esta opinión en esta red**

En relación con las medidas de centralidad, nos encontramos con las siguientes opciones:

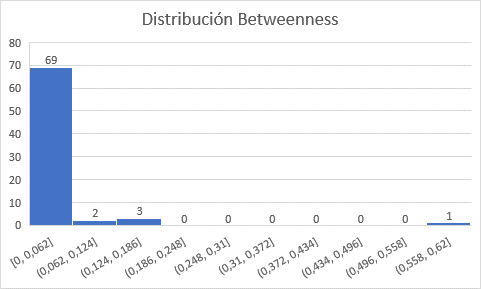
* **Grado (*Centrality Degree* y *Weighted Degree*)**
* **Cercanía o *Closeness***
* **Intermediación o *Betweenness***
* ***Hubs & Authorities***

En base al contexto de la red (conexiones y diálogos entre los diferentes personajes de una novela), consideramos que un personaje es **importante** en la obra si entabla numerosas conexiones con el resto de nodos y estas, a su vez, presentan elevados pesos. Esto supondría que la importancia reside en aquellos personajes que entablan conexiones con un alto porcentaje de personajes y, a su vez, un elevado número de conversaciones en la historia, es decir, aquel que hable prácticamente con todo el elenco de personajes. Por tanto, una medida de centralidad candidata sería el grado, tanto de centralidad como los pesos.

Sin embargo, ¿Es la única medida a tener en cuenta? A modo de ejemplo, supongamos que un personaje en la obra entabla 10 conexiones con otros personajes y cuya suma de los pesos es de 40. Por el contrario, otro personaje entabla 8 conexiones y la suma de sus pesos es de 30. Aparentemente, las medidas de centralidad escogidas apuntarían al primer personaje como el nodo de mayor importancia. Sin embargo, resulta que el primer personaje presenta un valor de c*loseness* muy bajo (0.3), mientras que el segundo presenta un valor significativamente mayor (0.7). Esto último supondría que, pese a que el primer personaje entable un mayor número de conversaciones, se sitúa mucho más lejos del resto de personajes, es decir, tan solo se relaciona con los nodos de su grupo y rara vez entabla conversación con otros nodos. Por el contrario, el segundo nodo está mucho más cerca del centro, es decir, **está en medio de todos los personajes**, por lo que no parece relacionarse con un solo grupo en particular.

Por tanto, **no solo importa que un nodo entable numerosas conversaciones con sus nodos más habituales**, **sino además que lo haga, en la medida de lo posible, con otros personajes que no sean de su grupo**, por lo que consideramos como medida de centralidad tanto el grado como la *closeness*.

Sin embargo, ¿Qué sucede con la *betweenness*? Efectivamente, también se trata de una medida importante, puesto que un personaje con una elevada capacidad de intermediación supondría ser un personaje que se ve involucrado en las tramas más importantes de la historia, ya que es un nodo fundamental por el que transcurren la mayor parte de los diálogos entre personajes. Sin embargo, al analizar el valor de *betweenness* en la red, nos encontramos con el siguiente problema:



Podemos comprobar que la mayoría de los valores giran en torno a cero. De hecho, si calculamos el número de nodos cuyo valor de *betweenness* es cero:

|  |  |
| --- | --- |
| ***betweenness*** | ***número de nodos*** |
| 0 | 41 (~55 %) |

Observamos que alrededor del 55 % de los nodos no presentan capacidad de intermediación alguna con el resto de nodos, mientras que tan solo un nodo concentra la mayor parte del grado de intermediación. Por otro lado, si comparamos el número de valores únicos tanto en *closeness* como en *betweenness*:

|  |  |
| --- | --- |
| ***medida centralidad*** | ***número de nodos*** |
| *betweenness* | 32 |
| *closeness* | 42 |

El número de valores únicos en el caso de *closeness* es mayor al de *betweenness*, lo que nos permitiría **descartar un mayor número de nodos en caso de empate**. Por tanto, aportaremos un mayor peso de la centralidad para la *closeness*. Por último, en relación con el *Hubs & Authorities*, aunque se trate de una medida especialmente orientada a redes dirigidas, aplicaremos un pequeño porcentaje de importancia de cara a la agregación, con el objetivo de deshacer posibles empates entre diferentes nodos.

En conclusión, siguiendo el orden de importancia concentramos el mayor peso de la agregación sobre los grados tanto de centralidad y peso (35 % en cada uno), seguido del valor de *closeness* (25 %). En menor medida, aportamos un peso del 3 % al valor *betweenness*, dado el elevado número de nodos cuya capacidad de intermediación es prácticamente nula, además de un 2 % al *Hub & Authority* como **valor residual**:

## **7. Busca los cinco nodos más importantes en la red y los cinco nodos con una interpretación más diferente en función del criterio de centralidad usado**

Inicialmente, comenzamos buscando los cinco nodos más importantes en la red, tanto en la **agregación por órdenes**:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **N** | **Label** | **OrdenD** | **OrdenW** | **OrdenC** | **OrdenB** | **OrdenH&A** | **Agregacion** | **Dispersion** |
| *12* | *12* | *1* | *1* | *1* | *1* | *10* | *1,18* | *9* |
| *54* | *56* | *3* | *2* | *2* | *5* | *6* | *2,52* | *4* |
| *49* | *49* | *2* | *4* | *5* | *3* | *5* | *3,54* | *3* |
| *57* | *59* | *6,5* | *3* | *6* | *8* | *1* | *5,085* | *7* |
| *26* | *26* | *5* | *7,5* | *3,5* | *6* | *15* | *5,73* | *11,5* |

Como también por **porcentajes** **e índices**:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **N** | **Label** | **PorcentajeCD** | **PorcentajeWD** | **PorcentajeC** | **PorcentajeB** | **PorcentajeH&A** | **Agregacion** | **Dispersion** |
| *12* | *12* | *0,072289157* | *0,06937799* | *0,022062264* | *0,345091608* | *0,037671111* | *0,066205238* | *0,323029344* |
| *49* | *49* | *0,044176707* | *0,040271132* | *0,017588099* | *0,102401555* | *0,063786697* | *0,038301549* | *0,084813456* |
| *54* | *56* | *0,036144578* | *0,044657097* | *0,017964987* | *0,071924504* | *0,063014263* | *0,036189854* | *0,053959517* |
| *57* | *59* | *0,030120482* | *0,043859649* | *0,01643855* | *0,026705454* | *0,076754769* | *0,032338942* | *0,060316219* |
| *26* | *26* | *0,032128514* | *0,034290271* | *0,017711959* | *0,046336205* | *0,020849408* | *0,029481639* | *0,028624247* |

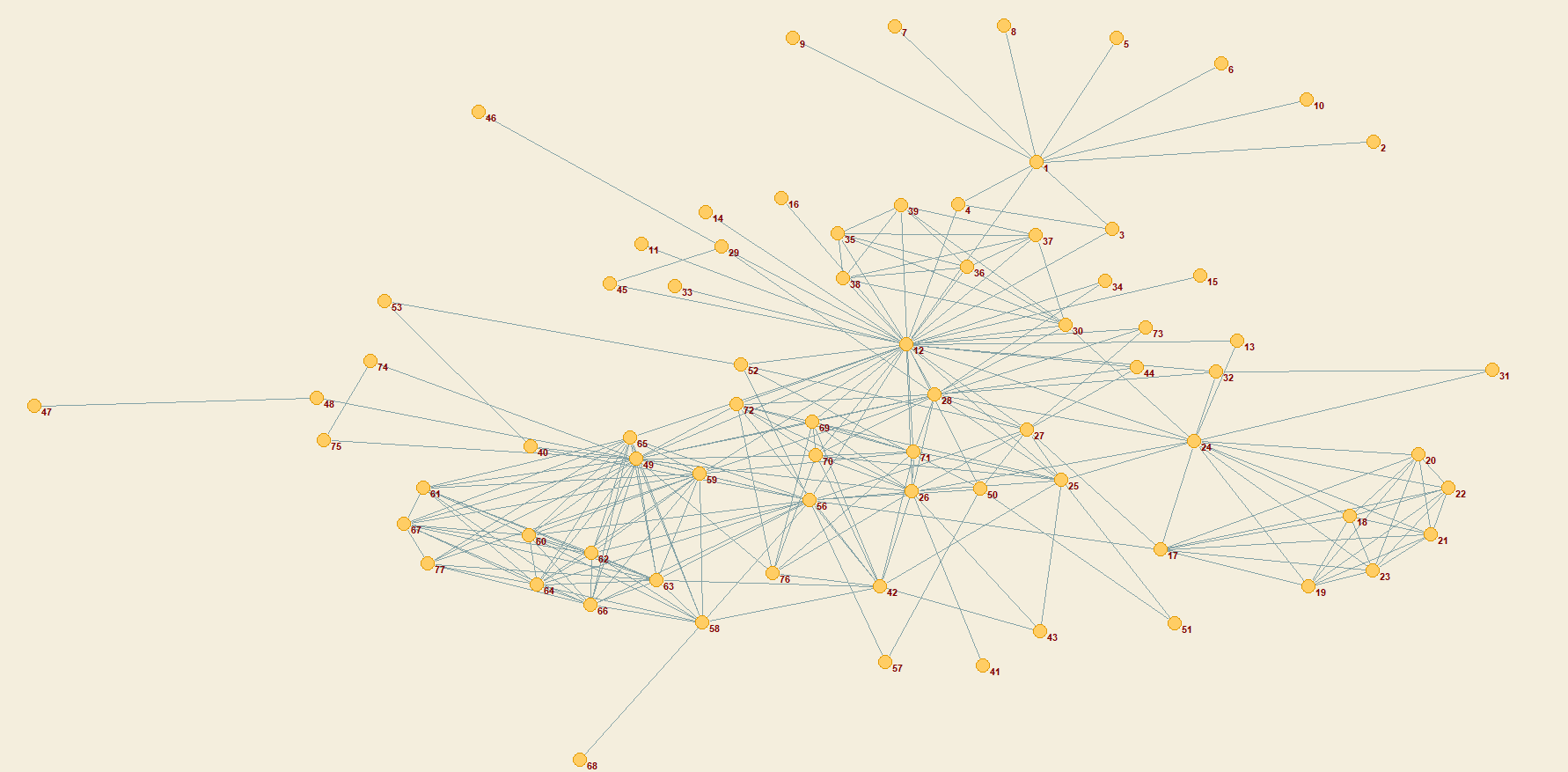
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **N** | **Label** | **IndiceCD** | **IndiceWD** | **IndiceC** | **IndiceB** | **IndiceH&A** | **Agregacion** | **Dispersion** |
| *12* | *12* | *5,421686747* | *5,203349282* | *1,654669826* | *25,88187059* | *2,825333324* | *4,965392851* | *24,22720077* |
| *49* | *49* | *3,313253012* | *3,020334928* | *1,319107412* | *7,680116588* | *4,784002239* | *2,872616175* | *6,361009175* |
| *54* | *56* | *2,710843373* | *3,349282297* | *1,347374002* | *5,394337805* | *4,726069689* | *2,714239013* | *4,046963803* |
| *57* | *59* | *2,259036145* | *3,289473684* | *1,232891243* | *2,002909039* | *5,756607686* | *2,425420676* | *4,523716442* |
| *26* | *26* | *2,409638554* | *2,571770335* | *1,328396903* | *3,475215407* | *1,563705635* | *2,211122912* | *2,146818503* |

Analizando cada uno de los criterios de agregación, a excepción de los nodos 54 y 49, los cuales varían en una posición en función del tipo de agregación, en los tres casos coinciden en cuales son los cinco personajes más importantes de la red:

* **Personaje 12**
* **Personaje 49**
* **Personaje 56**
* **Personaje 59**
* **Personaje 26**

Para comprender el orden de agregación, analicemos los valores de centralidad de cada nodo:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **N** | **Label** | **Centrality Degree** | **Weighted Degree** | **Closeness** | **Betweenness** | **H&A** |
| *12* | *12* | *36* | *87* | *0,649122807* | *0,574939739* | *0,176098496* |
| *54* | *56* | *18* | *56* | *0,528571429* | *0,119829792* | *0,294568346* |
| *49* | *49* | *22* | *50,5* | *0,517482517* | *0,17060607* | *0,298179189* |
| *57* | *59* | *15* | *55* | *0,483660131* | *0,044492611* | *0,358800127* |
| *26* | *26* | *16* | *43* | *0,521126761* | *0,077198417* | *0,097463265* |

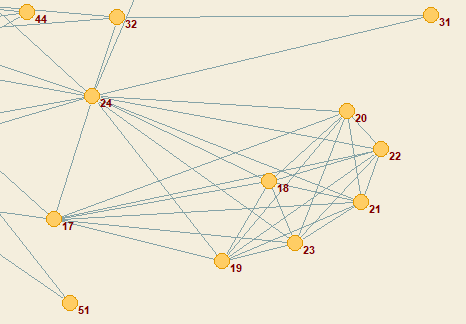


* **Personaje 12**: sin duda alguna, parece tratarse del personaje con mayor relevancia en la novela, un nodo con el mayor número de conexiones y diálogos en la historia (36 y 87, respectivamente), cuya *closeness* se sitúa cerca de 1, lo que indica que se trata de un personaje muy próximo al resto del elenco. Además el valor de *betweennees* que presenta (0.57) indica que se trata de un personaje por el que transcurre, mayoritariamente, la trama o tramas principales de la historia, en comparación con el resto de personajes.
* **Personajes 54 y 49**: aunque en menor medida, son personajes con un elevado número de conexiones y conversaciones con el resto de nodos, así como unos valores de *closeness* muy cercanos a los del personaje 12. No obstante, la capacidad de intermediación en ambos casos disminuye significativamente (0.11 y 0.17, respectivamente), es decir, pese a tener un amplio número de conexiones y diálogos en la historia, no se utilizan en gran medida como intermediario en las conversaciones entre dos personajes, o dicho de otro modo, el "grueso" de la trama no recae sobre estos personajes con especial relevancia.
* **Personajes 59 y 26**: por último, estos personajes destacan principalmente por un alto grado, tanto de centralidad como de pesos, aunque menor en comparación con el resto de personajes anteriores. Por el contrario, tanto el valor de *closeness* como de intermediación o *betweenness* no destacan especialmente, sobretodo en este último, cuyo valor es muy próximo a cero en ambos nodos.

Una vez analizados los nodos más importantes, ¿Qué nodos presentan una interpretación diferente en función del criterio de centralidad empleado? Para ello, estudiemos los nodos con mayor dispersión obtenida con la agregación por órdenes:

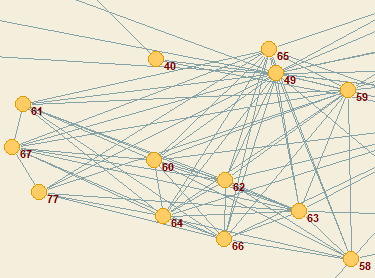
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **N** | **Label** | **OrdenCD** | **OrdenWD** | **OrdenC** | **OrdenB** | **OrdenH&A** | **Agregacion** | **Dispersion** |
| *21* | *21* | *30,5* | *15* | *59,5* | *55* | *45* | *33,35* | *44,5* |
| *20* | *20* | *30,5* | *15* | *59,5* | *55* | *50* | *33,45* | *44,5* |
| *59* | *61* | *24* | *27* | *49* | *55* | *11* | *31,97* | *44* |
| *48* | *48* | *54,5* | *53* | *53* | *14* | *33* | *51,955* | *40,5* |
| *58* | *60* | *14* | *9* | *44* | *29,5* | *4* | *20,015* | *40* |

* **Personajes 21 y 20**: como podemos comprobar en la tabla, ambos nodos presentan el mismo orden tanto en número de conexiones como en, peso total, *closenness*, *betweenness* y por ende, pese a un valor en *Hub & Authority* superior en el nodo 21, **un valor de agregación muy similar**. Si observamos la dispersión, destaca especialmente un contraste entre el valor de *closenness* (59.5 en ambos nodos) y el *Weighted Degree* (15 en ambos nodos). Por tanto, nos encontramos con dos personajes **muy alejados del resto de nodos** (orden de *closenness* situado en el tercio inferior), y como consecuencia **de la trama principal**, pero que conversan especialmente con los personajes de su entorno, aunque el número de conexiones que presenten no sea demasiado alto (30.5):



* **Personaje 61**: a diferencia de los personajes anteriores, el contraste se produce entre un valor de la *betweenness* situado en el tercio inferior, lo que se traduce en un personaje con escasa o nula capacidad de intermediación entre nodos, pero con un elevado *Hub & Authority*, es decir, pese a no relacionarse con demasiados nodos, con los que si lo hace son con personajes de gran importancia. De hecho, si nos fijamos en la agregación de los nodos conexos, observamos que algunos de ellos se sitúan en el tercio superior, incluso ciertos nodos como el 49 y el 59 se sitúan entre los 5 primeros:

|  |  |  |
| --- | --- | --- |
| **N** | **Label** | **Agregacion** |
| *63* | *65* | *7,185* |
| *49* | *49* | *3,54* |
| *57* | *59* | *5,085* |



* **Personaje 48**: destaca especialmente el contraste entre el valor de *betweenness* y los grados de centralidad y peso, es decir, **un personaje con una alta capacidad de intermediación** **o difusión de las principales tramas de la historia** (cuyo orden, 11, se sitúa sobre el tercio superior), **aunque se relaciona con muy pocos nodos**, destacando más por su capacidad de difusión que por sus diálogos, además de estar **muy alejado** con respecto al resto de los personajes, dado que su valor de *closenness* se sitúa sobre el tercio inferior (53). Además, si nos fijamos en la siguiente captura podemos comprobar que el alto grado de intermediación es debido principalmente a que es el **único puente de diálogo entre los nodos 47 y 49**:

