

Análisis comparativo del modelo de red aleatoria y modelo de barabasi

1. Análisis comparativo de las redes

Para comenzar, cabe tener en cuenta que la realización de este análisis comparativo se ha realizado con los datos obtenidos a partir de la media de los resultados obtenidos tras repetir la generación de cada red 10 veces cada una, de cada uno de los modelos, tal y como se sugería para la realización de la parte opcional de la práctica. Dichos datos se pueden encontrar en el fichero “*metricas.xlsx*” de la entrega, con los valores obtenidos tras cada generación de cada red (pestañas dentro del Excel).

En el apartado 4 de este mismo documento trataremos las diferencias y similitudes entre los modelos generados y los modelos teóricos, asique en este apartado nos centraremos en cómo evolucionan las redes de cada modelo al generarlas nosotros.

Tomando las métricas del Excel como apoyo vemos como el número de aristas aumenta proporcionalmente en las redes aleatorias, siendo dependiente tanto del número de nodos a tratar como de la probabilidad seleccionada para cada etapa (resultando similares por ejemplo la red de tamaño 1000 en etapa subcrítica y la red de tamaño 500 en su etapa crítica, en lo que a número de aristas se refiere). En cambio, el grado medio en las redes aleatorias resulta similar en la misma etapa de cada red generada independientemente del tamaño de cada una de ellas, salvo en la etapa conectada de la red de tamaño 5000 que destaca por encima de las otras dos redes en esa etapa.

Siguiendo con el número de aristas, vemos como en las redes libres de escala generadas resulta mucho más matemático el resultado pues prácticamente coincide con el resultado de multiplicar el tamaño de la red con la m dada en cada caso.

Pasando a tratar la densidad de las redes, vemos cómo las redes aleatorias se vuelven más densas en función de la etapa en la que la situemos, siendo más densas cuanto más pequeñas son las redes pues existen menos conexiones potenciales que en las redes más grandes. De igual manera pasa con las redes libres de escala, pues son más densas las redes más pequeñas siendo prácticamente igual para $m=3$ y $m=4$ en cada una de ellas.

En referencia al hub de mayor grado vemos la gran diferencia entre las redes libres de escala y las redes aleatorias, pues mientras que en las redes aleatorias el hub con mayor grado ronda el grado 24 para la red con mayor número de nodos ($N=500$) en su etapa conectada (la que usa una probabilidad mayor), para las redes libres de escala el mayor grado de un hub llega a rondar el grado 268 lo que indica una clara diferencia entre ambas redes, confirmando la existencia de hubs en las redes libres de escala mientras que no es así en las redes aleatorias. De manera general, el largest hub degree aumenta según el tamaño de la red en ambos tipos de redes.

Acerca de la distancia media entre nodos, vemos que Gephi no tiene en cuenta los nodos inconexos mientras que la fórmula teórica si los tiene como trataremos en el apartado 4 de este documento. Asique en los datos obtenidos por Gephi tras la generación de las redes aleatorias, observamos como los datos obtenidos para las etapas subcrítica y crítica de estas redes no son demasiado coherentes, pues aumentan en función del número de nodos, sin embargo, para las etapas supercrítica y conectada la distancia media varía de una forma menos radical, hasta llegar a ser prácticamente igual independientemente del número de nodos en la etapa conectada.

Sin embargo, en las redes libres de escala la distancia media es prácticamente similar en todos los casos independientemente del número de nodos y sólo variando ligeramente en función de la m , siendo menor cuando la m es más grande.

Por último, fijándonos en el coeficiente de agrupamiento, observamos que en las redes aleatorias su valor aumenta en función de la etapa en la que se encuentre, siendo mayor cuanto mayor es la probabilidad, en cambio este coeficiente es más pequeño cuanto mayor es el tamaño de la red.

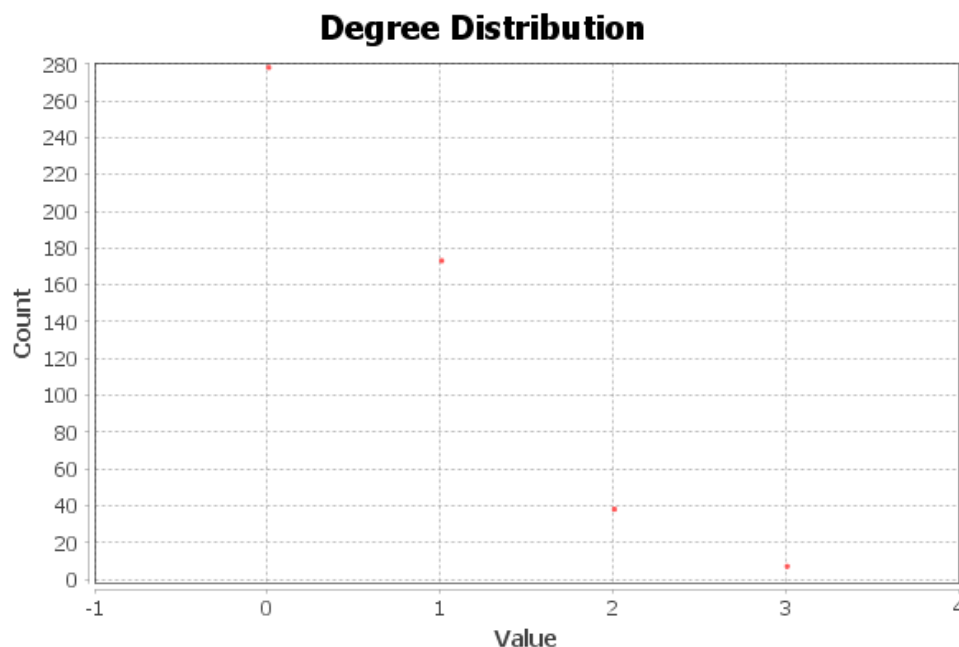
Para redes libres de escala vemos como el coeficiente de agrupamiento tiene valores más altos que las redes aleatorias en todos los casos, lo que nos indica que las redes libres de escala tienen unos nodos mucho más interconectados entre sí. En cambio, al igual que en las redes aleatorias los valores son más pequeños cuanto más grande es la red. Aunque para redes de mismo tamaño dicho coeficiente es más alto en función de cómo aumenta m , lo que es lógico pues una m más grande produce más aristas y por tanto hace una red con unos nodos más interconectados entre sí.

2. Representaciones gráficas de las distribuciones de grados

En este apartado situaremos las gráficas de las distribuciones de grados para cada una de las redes generadas.

2.1. Redes aleatorias.

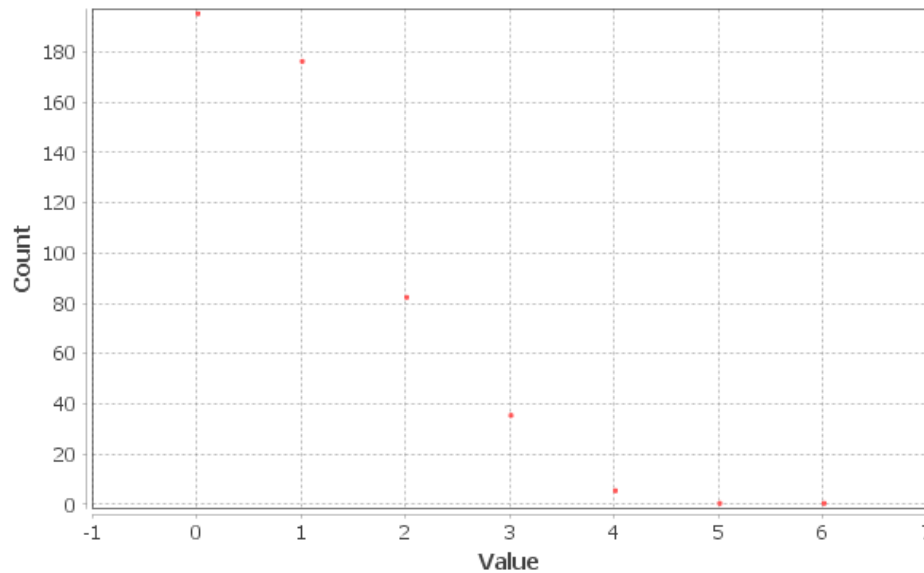
- Random Network, etapa subcrítica, $N=500$.



En el gráfico observamos como en la etapa subcrítica para $N=500$ abundan los nodos de grado bajo debido a que la probabilidad es muy baja, resultando muchos nodos inconexos o de grado 1, y la gráfica resultante es descendente.

- Random Network, punto crítico, N=500.

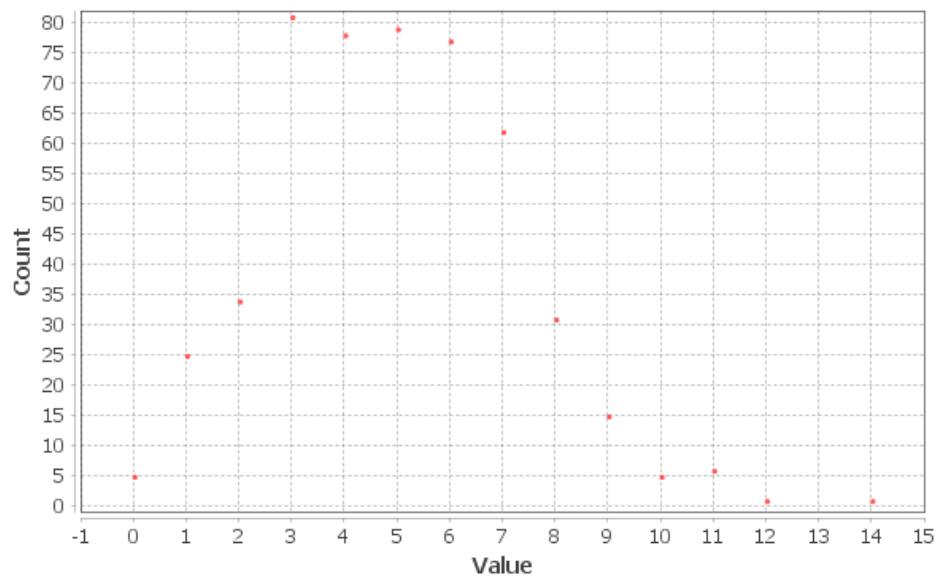
Degree Distribution



En el gráfico observamos como en el punto crítico para $N=500$ siguen abundando los nodos de grado bajo debido a que la probabilidad sigue siendo baja, aunque la gráfica resultante ya no muestra una caída tan pronunciada a pesar de seguir siendo descendente.

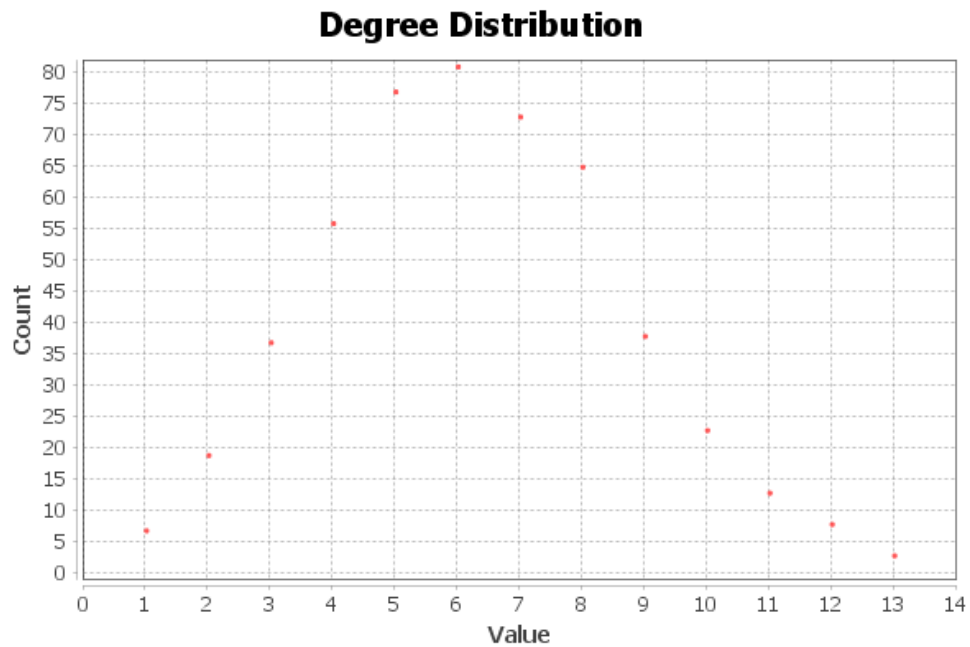
- Random Network, etapa supercrítica, N=500.

Degree Distribution



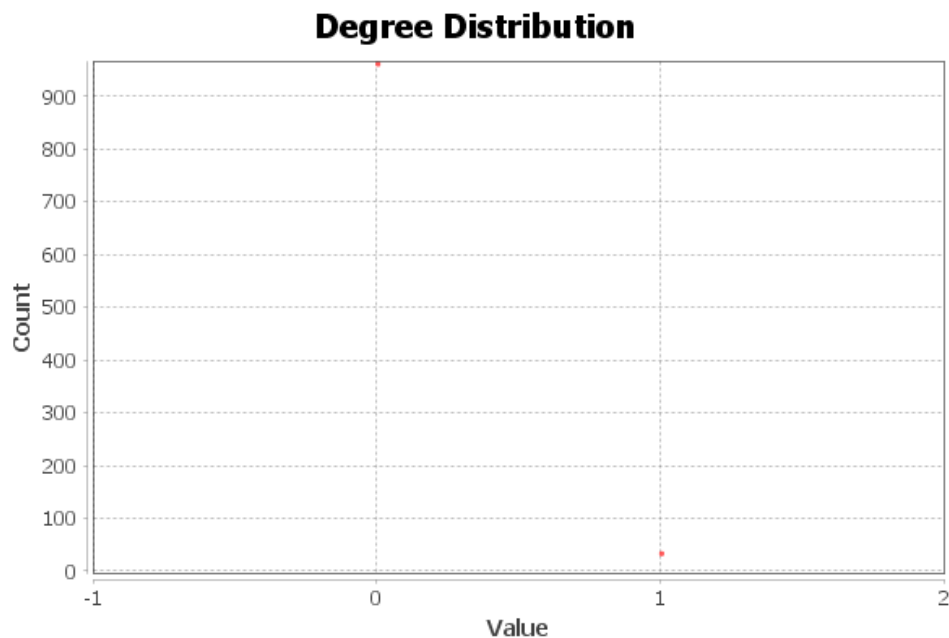
En el gráfico observamos como en la etapa supercrítica para $N=500$ se empieza a intuir la curva característica de las redes aleatorias, resultando una repartición de grados más igualada para los nodos de grado entre 3 y 6, apareciendo unos cuantos nodos con grados cercanos a esos.

- **Random Network, etapa conectada, N=500.**



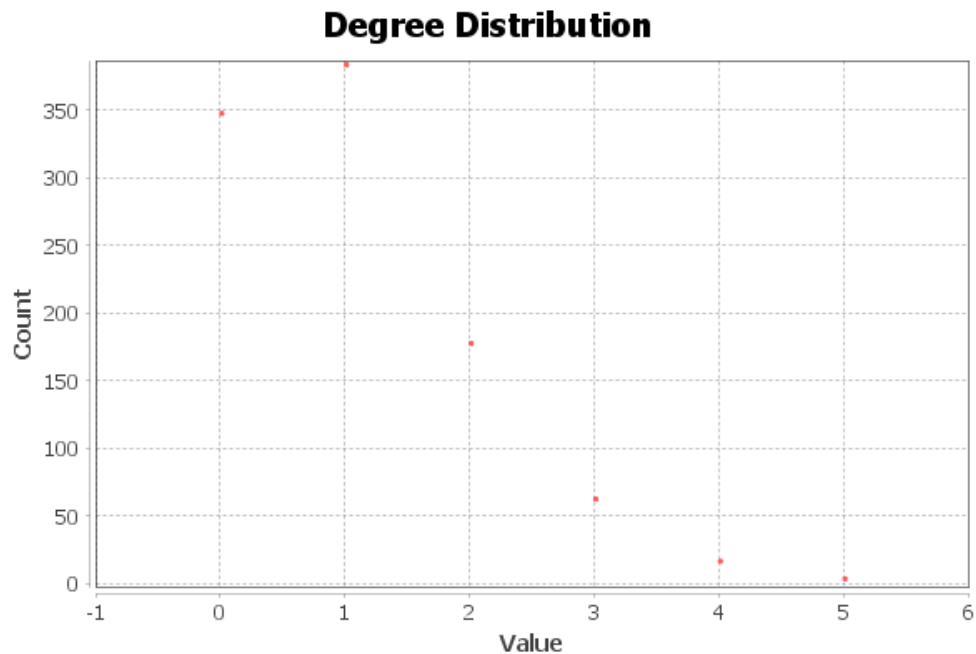
En el gráfico observamos como en la etapa conectada para $N=500$ se intuye totalmente la curva característica de las redes aleatorias, destacando como punto máximo el grado 6 que resulta ser el grado medio para la red en esta etapa, y la mayoría del resto de nodos se sitúa en grados cercanos a este.

- **Random Network, etapa subcrítica, N=1000.**



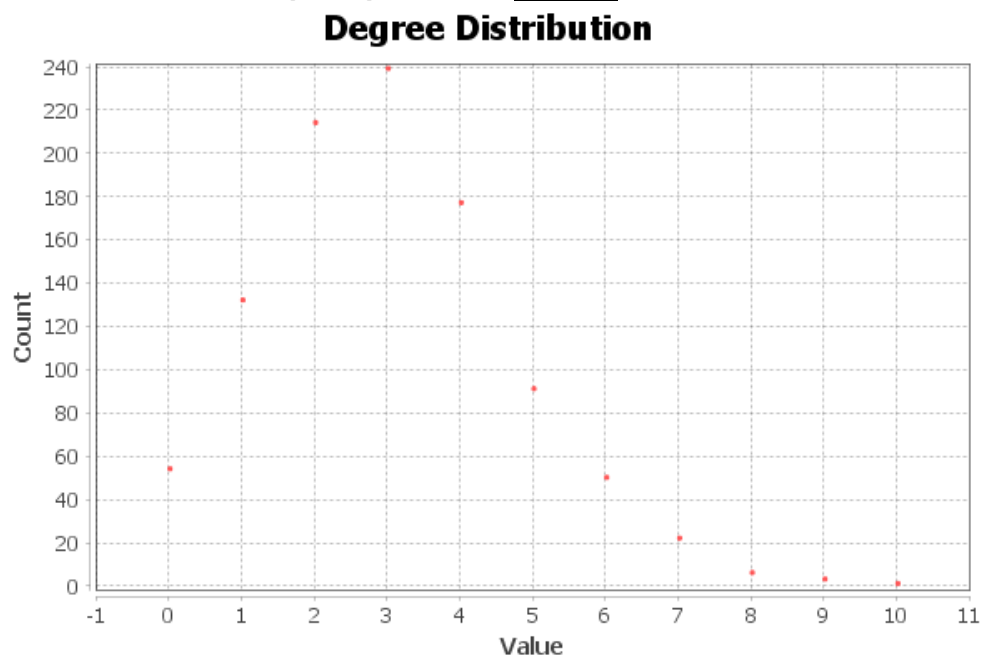
En el gráfico observamos como en la etapa subcrítica para $N=1000$ se sitúan todos los nodos en grados de entre 0 y 1 debido a que la probabilidad es muy baja y resultan muchos nodos inconexos, quedando la gráfica resultante descendente.

- **Random Network, punto crítico, $N=1000$.**



En el gráfico observamos como en el punto crítico para $N=1000$ siguen abundando los nodos de grado bajo, sobre todo de grado 1 lo que reduce el número de nodos inconexos, debido a que la probabilidad sigue siendo baja, aunque la gráfica resultante ya no muestra una caída tan pronunciada a pesar de seguir siendo descendente.

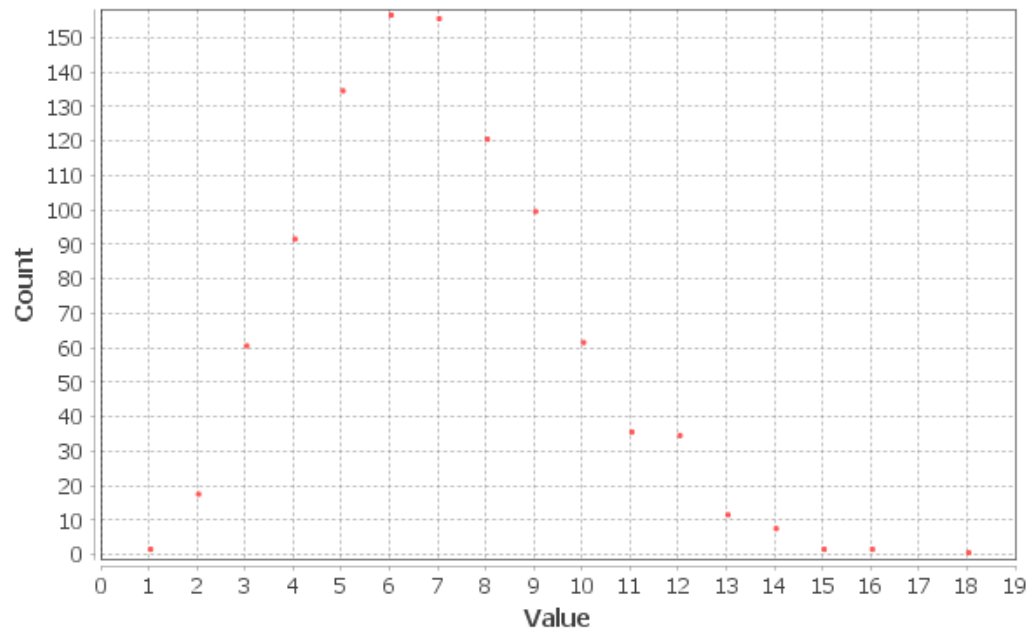
- **Random Network, etapa supercrítica, $N=1000$.**



En el gráfico observamos como en la etapa supercrítica para $N=1000$ se intuye la curva característica de las redes aleatorias, en la que destaca el grado 3, que resulta ser el grado medio para esta red, apareciendo la mayoría del resto de nodos en grados cercanos a ese.

- **Random Network, etapa conectada, N=1000.**

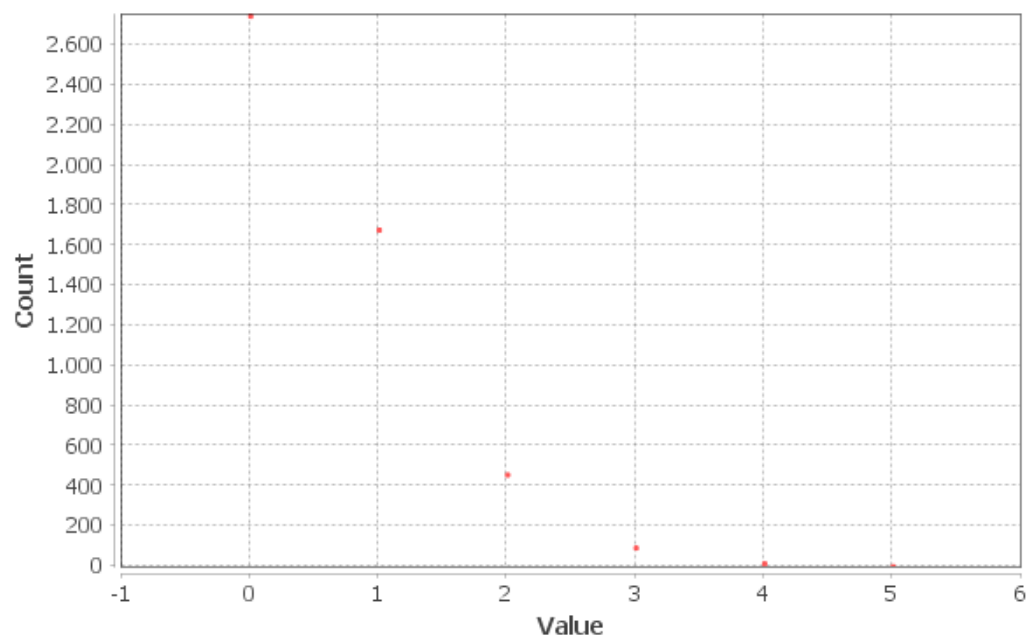
Degree Distribution



En el gráfico observamos como en la etapa conectada para $N=1000$ observamos cómo se sigue intuyendo la curva característica de las redes aleatorias, pero en este caso destacan los grados 6 y 7, situándose el grado medio entre estos dos y el resto de nodos quedan situados en grados cercanos a estos.

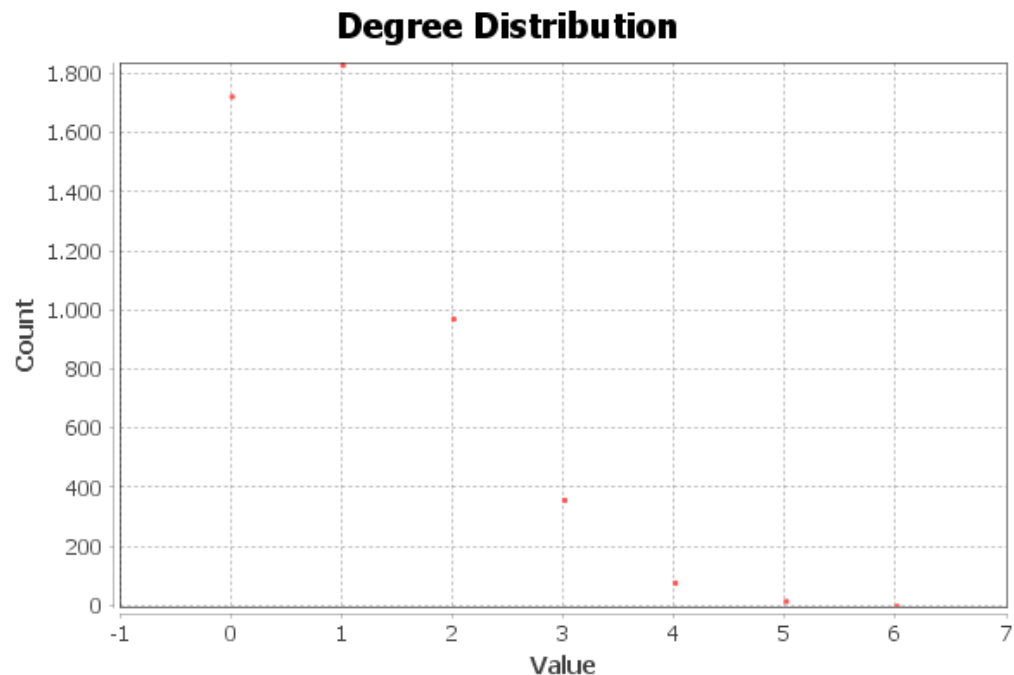
- **Random Network, etapa subcrítica, N=5000.**

Degree Distribution



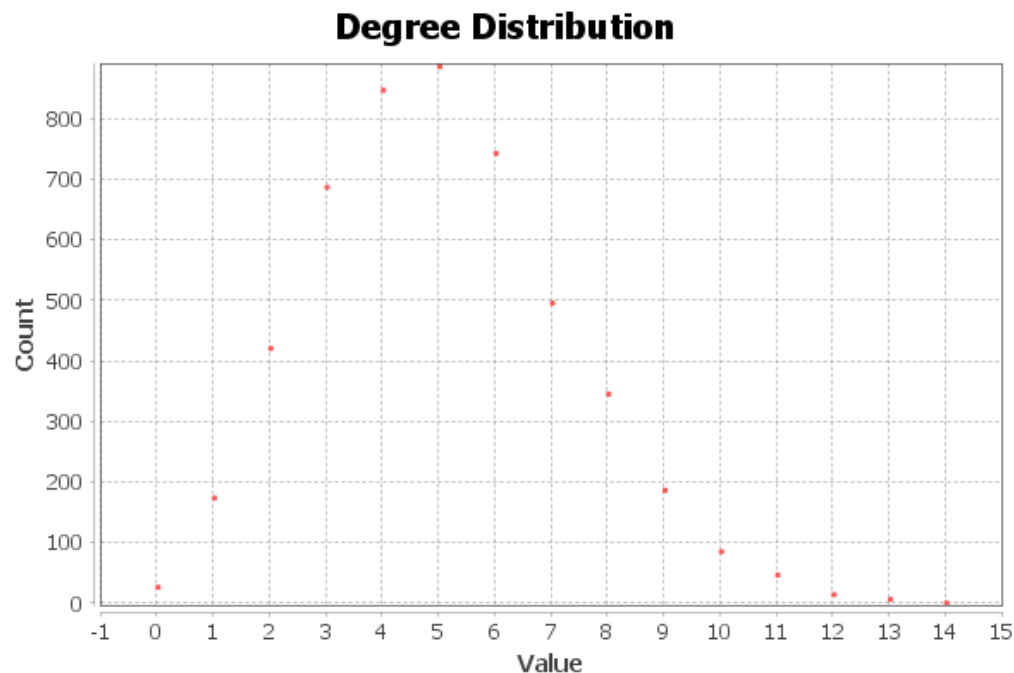
En el gráfico observamos como en la etapa subcrítica para $N=5000$ abundan los nodos de grado bajo debido a que la probabilidad es muy baja, resultando muchos nodos inconexos o de grado 1, y la gráfica resultante es descendente.

- **Random Network, punto crítico, $N=5000$.**



En el gráfico observamos como en el punto crítico para $N=5000$ siguen abundando los nodos de grado bajo, sobre todo de grado 1 lo que reduce el número de nodos inconexos, aunque siguen existiendo muchos nodos inconexos, debido a que la probabilidad sigue siendo baja, la gráfica resultante sigue siendo descendente.

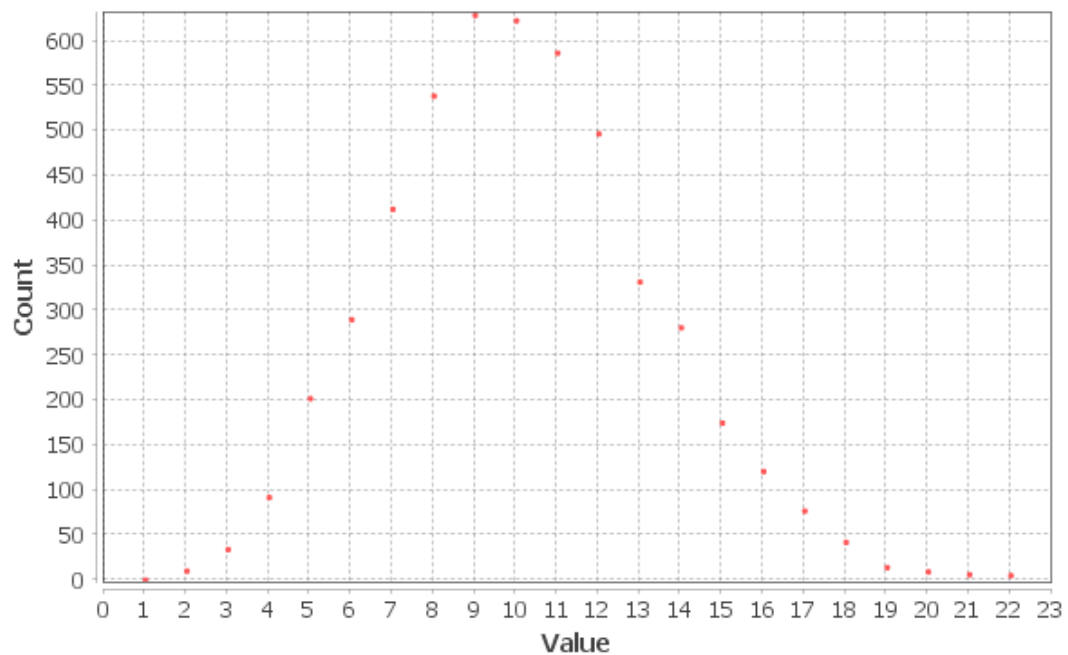
- **Random Network, etapa supercrítica, $N=5000$.**



En el gráfico observamos como en la etapa supercrítica para $N=5000$ se intuye la curva característica de las redes aleatorias, en la que destaca el grado 5, que resulta ser el grado medio para esta red, apareciendo la mayoría del resto de nodos en grados cercanos a ese.

- **Random Network, etapa conectada, $N=5000$.**

Degree Distribution

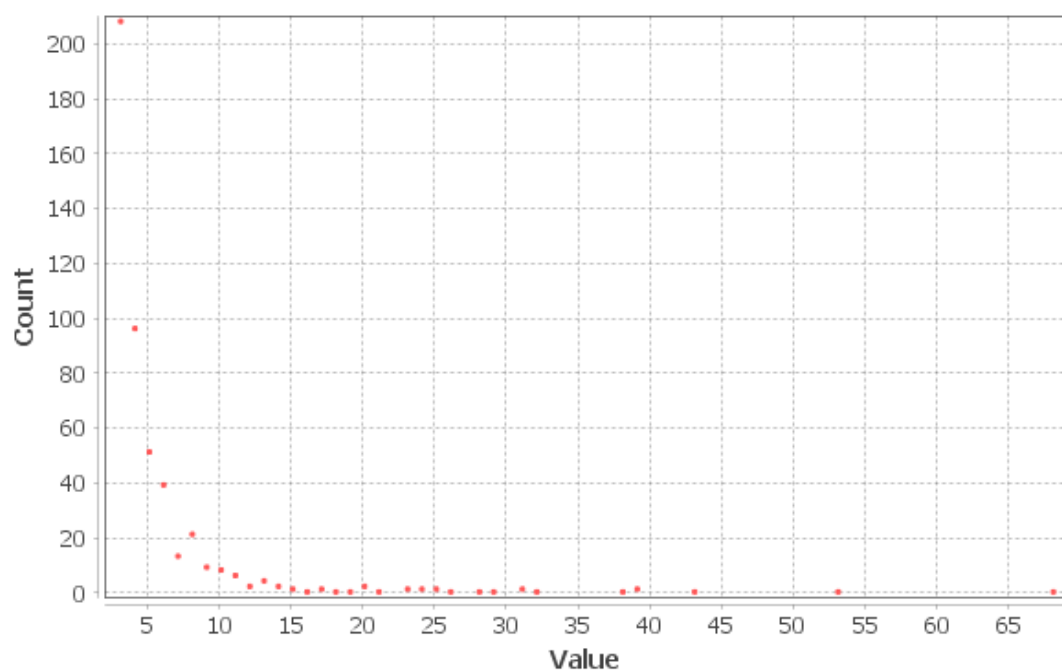


En el gráfico observamos como en la etapa conectada para $N=5000$ observamos cómo se sigue intuyendo la curva característica de las redes aleatorias, pero en este caso destacan los grados entre 9 y 11, situándose el grado medio cercano a 10 y el resto de nodos quedan situados en grados cercanos a estos.

2.2. Redes libres de escala (Barabasi-Albert).

- **Scale-free Network, $N=500$, $m=3$.**

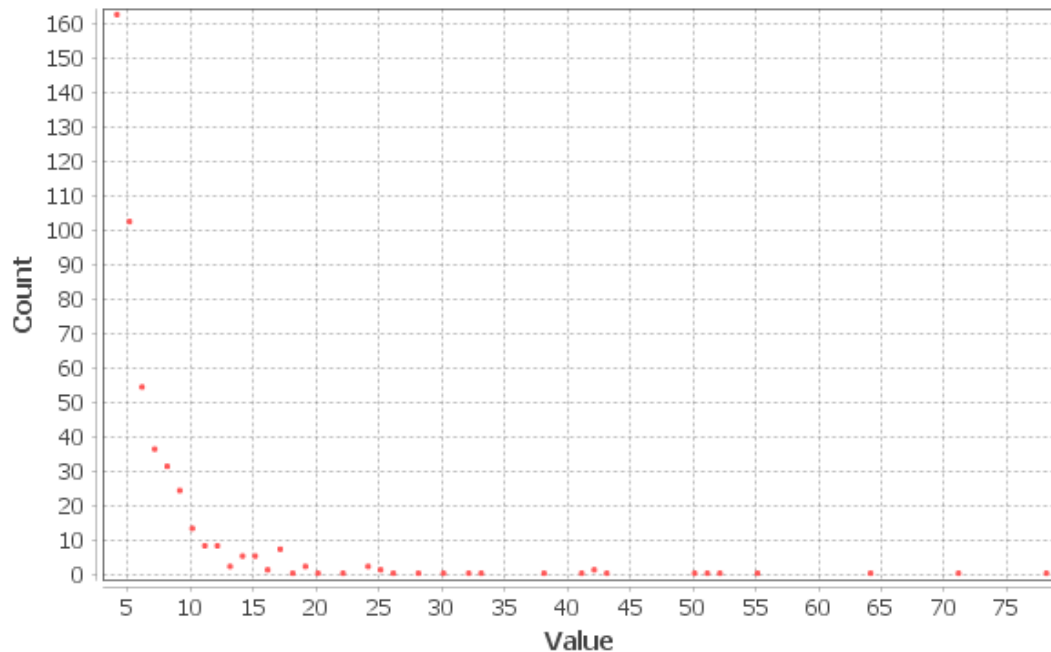
Degree Distribution



En el gráfico observamos como para $N=500$ y $m=3$ abundan los nodos de grado bajo (aproximadamente grado 3), aunque vemos como aparecen nodos de grado alto, en especial destacan algunos situados entre los grados 53 y 68 que se identificarán como los hubs de la red. Además de que la distribución de grados para este caso se ve mucho más repartida que en las redes aleatorias de este tamaño.

- **Scale-free Network, $N=500$, $m=4$.**

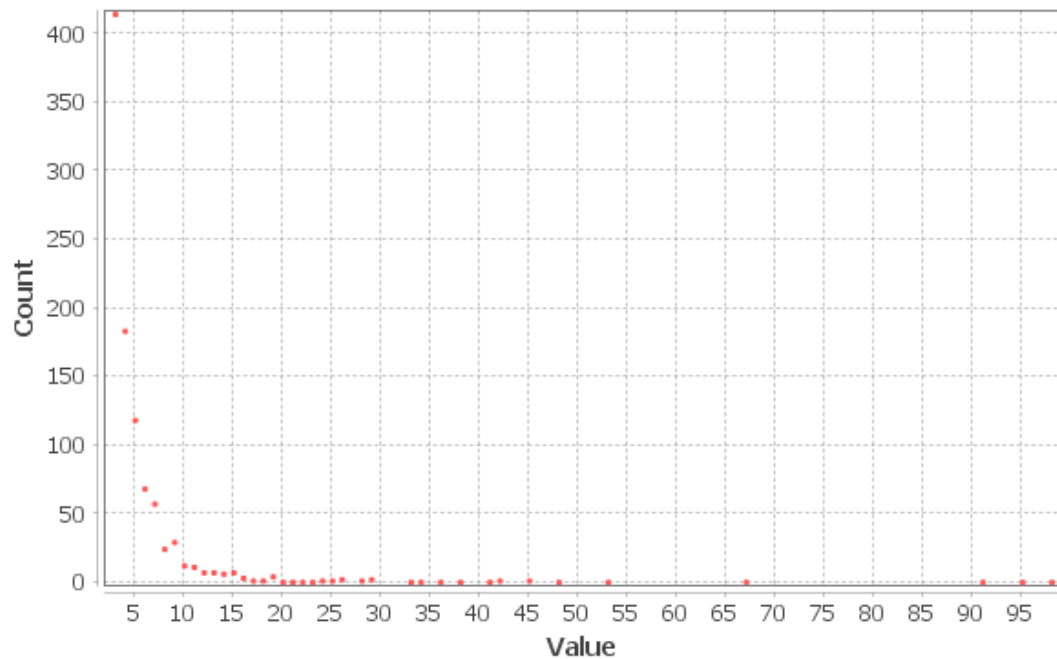
Degree Distribution



En el gráfico observamos como para $N=500$ y $m=4$ abundan los nodos de grado bajo (aproximadamente grado 4), aunque vemos como aparecen nodos de grado alto, en especial destacan algunos situados entre los grados 64 y 79 que se identificarán como los hubs de la red. Además de que la distribución de grados para este caso se ve mucho más repartida que en las redes aleatorias de este tamaño.

- **Scale-free Network, N=1000, m=3.**

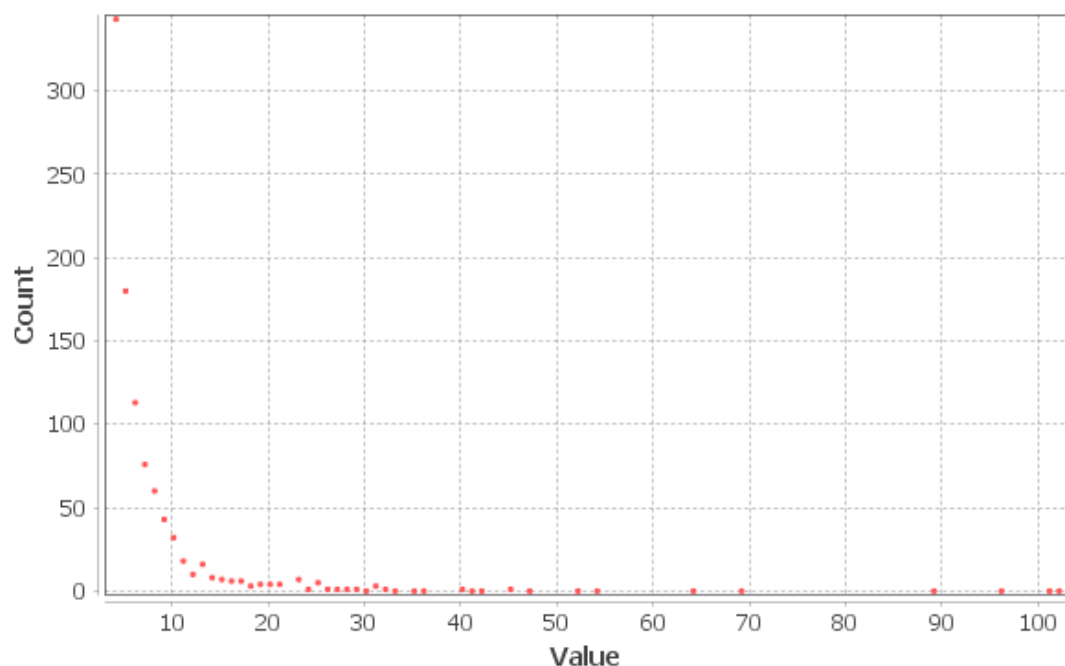
Degree Distribution



En el gráfico observamos como para $N=1000$ y $m=3$ abundan los nodos de grado bajo (aproximadamente grado 3), aunque vemos como aparecen nodos de grado mucho más alto que para las redes de $N=500$, en especial destacan algunos situados entre los grados 91, 95 y 98 que se identificarán como los hubs de la red. Además de que la distribución de grados para este caso se ve mucho más repartida que en las redes aleatorias de este tamaño.

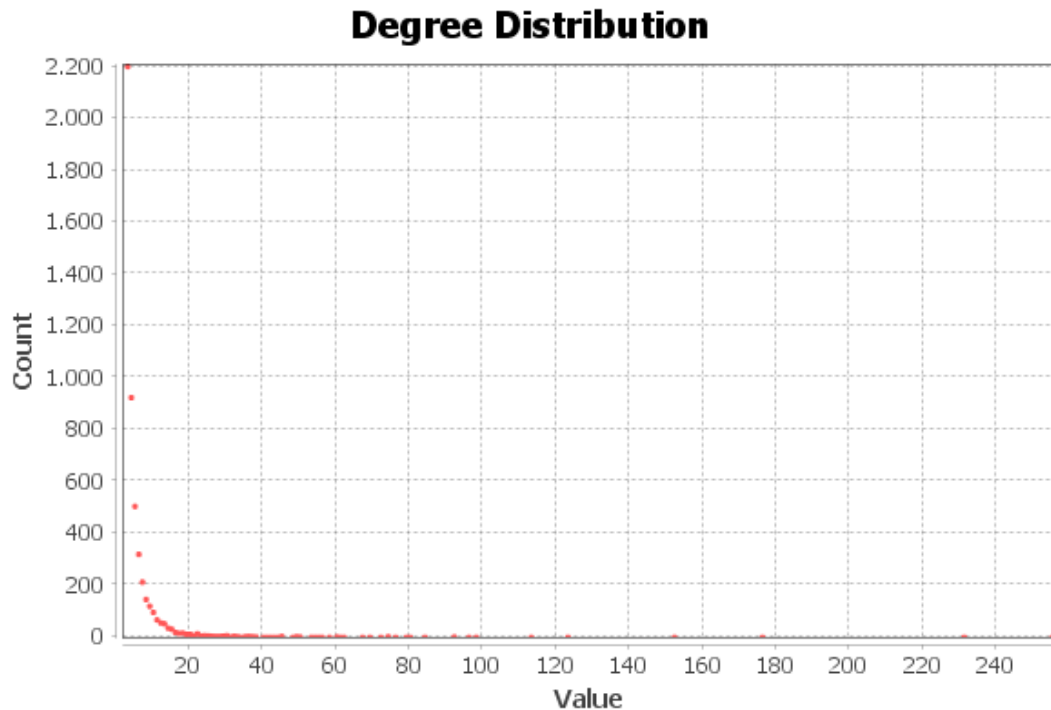
- **Scale-free Network, N=1000, m=4.**

Degree Distribution



En el gráfico observamos como para $N=1000$ y $m=4$ abundan los nodos de grado bajo (aproximadamente grado 4), aunque vemos como aparecen nodos de grado mucho más alto que para las redes de $N=1000$, en especial destacan algunos situados en grados superiores a 100 que se identificarán como los hubs de la red. Además de que la distribución de grados para este caso se ve mucho más repartida que en las redes aleatorias de este tamaño.

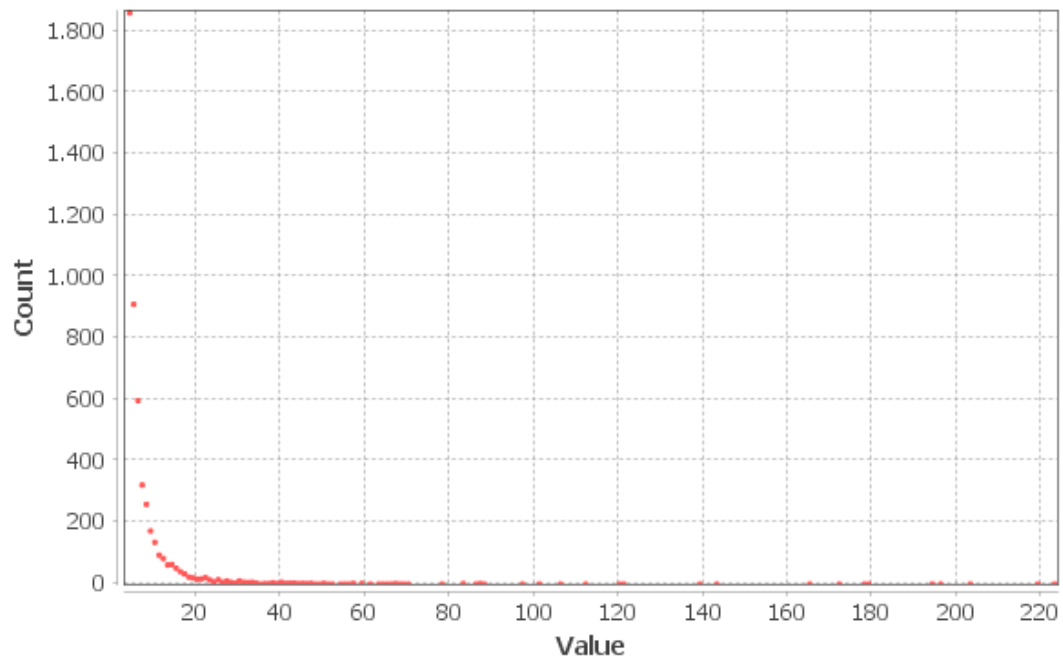
- **Scale-free Network, $N=5000$, $m=3$.**



En el gráfico observamos como para $N=5000$ y $m=3$ abundan los nodos de grado bajo (aproximadamente grado 3), aunque vemos como aparecen nodos de grado mucho más alto que para las redes de $N=500$ y $N=1000$, en especial destacan algunos situados en grados superiores a 230 que se identificarán como los hubs de la red. Además de que la distribución de grados para este caso se ve mucho más repartida que en las redes aleatorias de este tamaño.

- **Scale-free Network, $N=5000$, $m=4$.**

Degree Distribution



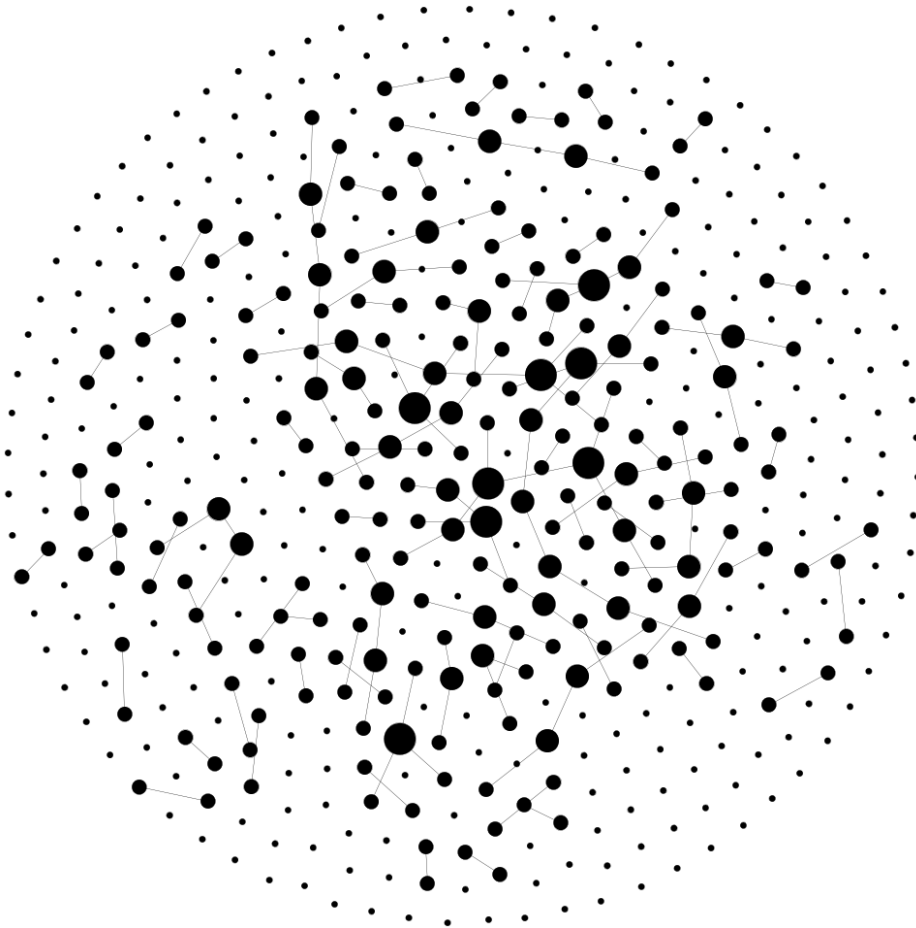
En el gráfico observamos como para $N=5000$ y $m=4$ abundan los nodos de grado bajo (aproximadamente grado 4), aunque vemos como aparecen nodos de grado mucho más alto que para las redes de $N=500$ y $N=1000$, en especial destacan algunos situados en grados superiores a 200 que se identificarán como los hubs de la red, aunque destaca como este último caso es muy parecido al de $m=3$ pues ya casi no afecta para nodos de alto grado. Además de que la distribución de grados para este caso se ve mucho más repartida que en las redes aleatorias de este tamaño.

3. Representaciones gráficas de las redes de menor tamaño

En este apartado situaremos las representaciones gráficas de las redes de menor tamaño ($N=500$) de cada uno de los modelos generados. Cada una de las representaciones ha sido realizada siendo el tamaño de cada nodo dependiente de su grado.

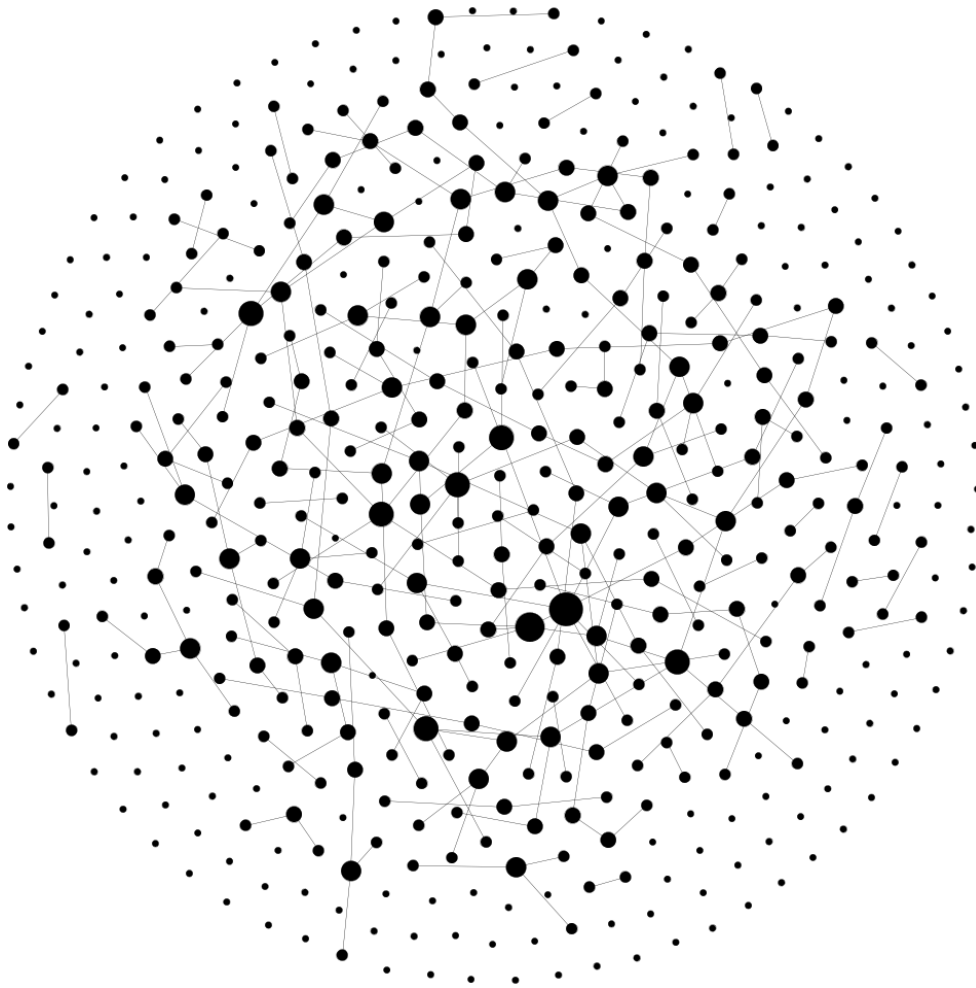
3.1. Redes aleatorias.

- Random Network, etapa subcrítica, $N=500$.



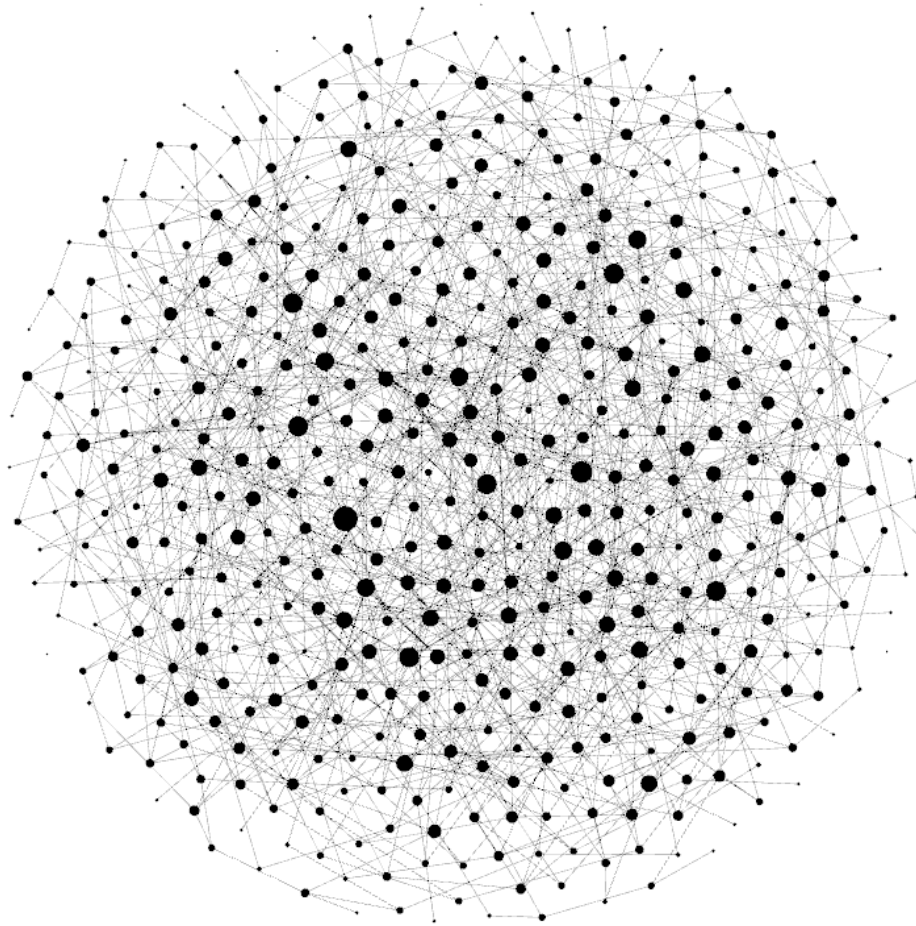
En esta imagen vemos como para la etapa subcrítica de la red aleatoria de $N=500$ existen muchos nodos inconexos y unos cuantos más grandes que indican que hay algunos con unos grados un poco más alto a pesar de que apenas se observan aristas que los unan.

- Random Network, punto crítico, $N=500$.



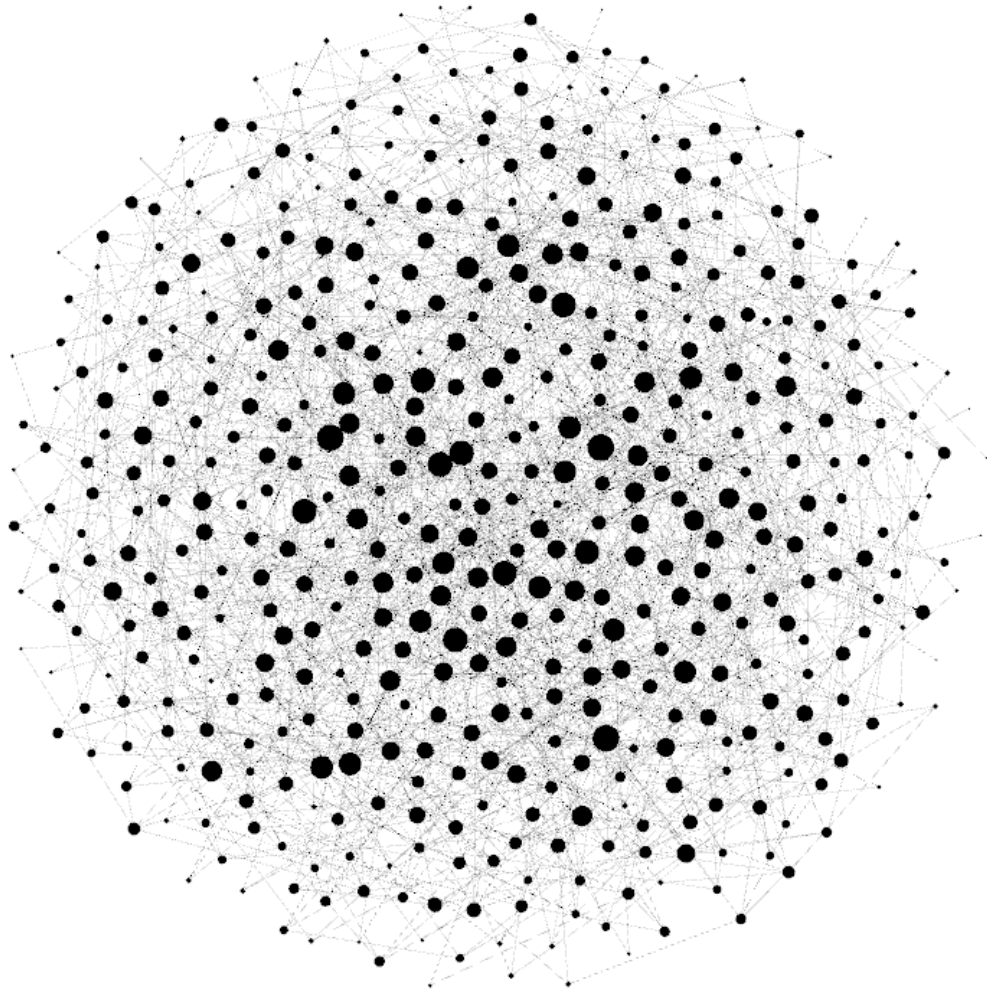
En esta imagen vemos como para el punto crítico de la red aleatoria de $N=500$ siguen existiendo muchos nodos inconexos, aunque ya se comienzan a apreciar más aristas que unen los nodos por la parte central de la imagen, apreciándose que no destacan nodos con un grado alto.

- **Random Network, etapa supercrítica, N=500.**



En esta imagen vemos como para la etapa supercrítica de la red aleatoria de $N=500$ ya casi no existen nodos inconexos y sin embargo siguen sin aparecer nodos que destaquen por su alto grado, estando los grados de los nodos muy repartidos, tal y como se supone en una red aleatoria.

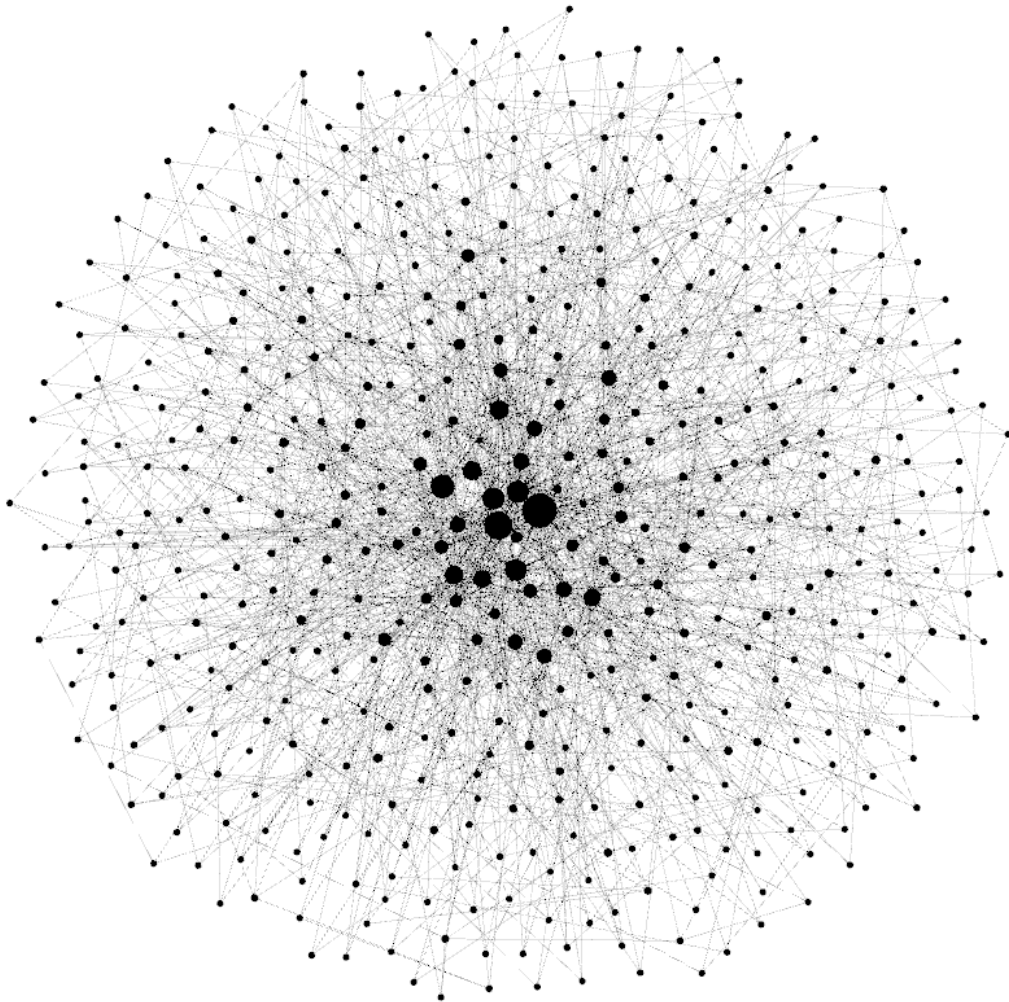
- **Random Network, etapa conectada, N=500.**



En esta imagen vemos como para la etapa conectada de la red aleatoria de $N=500$ ya no existen nodos inconexos (se aprecia sólo una componente conexa en apariencia) y sin embargo siguen sin aparecer nodos que destaquen por su alto grado, estando los grados de los nodos muy repartidos, tal y como se supone en una red aleatoria en esta etapa.

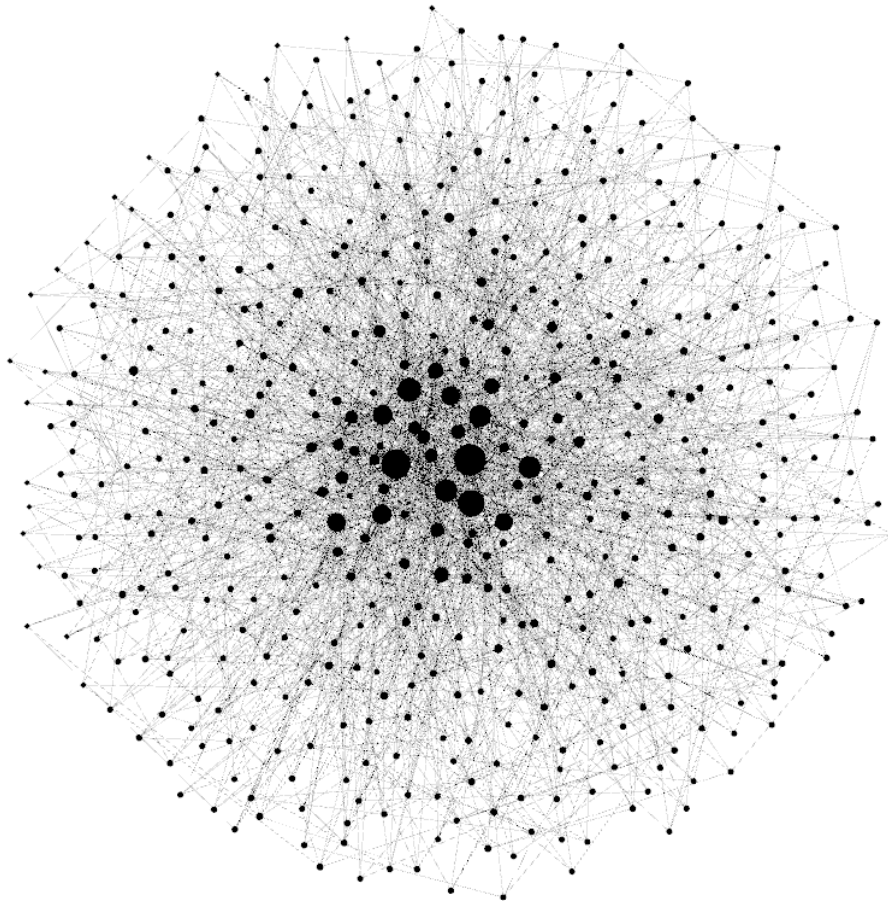
3.2. Redes libres de escala (Barabasi-Albert).

- Scale-free Network, $N=500$, $m=3$.



En esta imagen vemos como para la red libre de escala de $N=500$ y $m=3$, no existen nodos inconexos (pues su grado mínimo es 3), como sí que se apreciaban en las redes aleatorias de este tamaño. Además, se pueden ver nodos de grado muy alto en el centro de la imagen, que resultan ser los hubs característicos de este tipo de redes.

- **Scale-free Network, $N=500$, $m=4$.**



En esta imagen vemos como para la red libre de escala de $N=500$ y $m=4$, no existen nodos inconexos (pues su grado mínimo es 4), como sí que se apreciaban en las redes aleatorias de este tamaño. Además, se pueden ver nodos de grado muy alto en el centro de la imagen, que resultan ser los hubs característicos de este tipo de redes. En cambio, vemos que a pesar de apreciarse muchas más aristas que en la imagen anterior, ambas son bastantes similares.

4. Diferencias y similitudes entre resultados reales y predichos

En este apartado vamos a comparar los resultados obtenidos tras aplicar las fórmulas vistas en la teoría con los resultados reales obtenidos tras ejecutar 10 veces cada tipo de red generada de cada modelo.

Vemos como el número de aristas prácticamente coinciden los resultados teóricos obtenidos con los resultados reales obtenidos en las redes aleatorias. Por otro lado, el grado medio en estas redes también resulta similar tras comparar los resultados obtenidos con Gephi de las redes generadas con la aplicación de la fórmula teórica. Lo que nos indica que estos resultados se asemejan bastante al modelo teórico predicho.

En las redes libres de escala el número de aristas coincide totalmente, aunque hubo que modificar la fórmula teórica para el cálculo ya que la que estaba incluida en los apuntes no

tenía en cuenta que la red inicial estuviera totalmente conectada (como es el caso de la práctica), pero tras realizar esta modificación los resultados coinciden con los obtenidos con Gephi.

Siguiendo con los resultados de la densidad obtenidos tras aplicar la fórmula observamos como son muy similares tanto en los casos de las redes aleatorias como en los de las redes libres de escala. Lo que nos vuelve a indicar que nuestras redes generadas se asemejan bastante a los resultados predichos para estos modelos.

Al analizar los resultados obtenidos tras aplicar las fórmulas de obtención del Largest Hub Degree comenzaron nuestros problemas, no para el caso de las redes libres de escala que los resultados fueron muy aproximados a los obtenidos en las redes generadas, pero sí para el caso de las redes aleatorias ya que la fórmula teórica tiene en cuenta para este cálculo una división entre el grado medio y además una suma con el grado mínimo de la red, pero claro, al ir aumentando el grado medio según aumenta el número de nodos y la probabilidad se va haciendo cada vez más pequeño el valor resultante, lo que no tiene demasiado sentido pues en los datos reales obtenidos con Gephi estos valores van en aumento según aumenta el número de nodos y la probabilidad. En la fórmula esto podría solucionarse con la suma del grado mínimo de la red que debería ser cada vez un poco más grande (la gráfica de distribución de grados debería ir desplazándose a la derecha en cada etapa), sin embargo, hemos obtenido en todos los casos un grado mínimo de 1 en las redes aleatorias, lo que hace prácticamente nula la aportación de esta componente en la fórmula aplicada.

Continuando el análisis de los datos, llegamos al cálculo de la distancia media, que para el caso de las redes libres de escala observamos como la fórmula teórica sólo tiene en cuenta el tamaño de la red por lo tanto da igual la m que se use, que el cálculo da lo mismo para redes con la misma N , mientras que los resultados reales obtenidos por Gephi, aunque se asemejan bastante a los predichos, resultaron que para las redes con una m mayor nos queda una distancia media entre nodos menor, que realmente es lo que la lógica nos indica, pues a mayor número de enlaces debería reducirse la distancia media entre nodos.

En cambio, para las redes aleatorias la fórmula teórica usa el valor del grado medio lo que tiene en cuenta los nodos inconexos y que lleva a resultados negativos para las etapas subcrítica y crítica, sin embargo Gephi no tiene en cuenta estos nodos inconexos para sus cálculos del grado medio (nos cercioramos de ello viendo la api de Gephi que pudimos encontrar en este [enlace](#)), lo que permite que Gephi no saque números negativos para este cálculo. Para las siguientes etapas (supercrítica y conectada) como el número de nodos inconexos cada vez es más reducido los resultados obtenidos en ambos cálculos se asemejan en mayor medida.

En referencia al coeficiente de agrupamiento tanto para las redes aleatorias como para las redes libres de escala el cálculo teórico y los resultados reales obtenidos con Gephi se asemejan bastante, siendo valores más altos para las redes libres de escala tal y como hablamos de ello en el apartado 1, pues son redes con mayor número de enlaces.

Por último, tenemos en cuenta que para el número de componentes conexas no existe una fórmula teórica que aproxime su cálculo, aunque lo que sabemos seguro es que para la etapa conectada de las redes aleatorias debe existir tan solo una componente conexa y para estos casos Gephi nos devolvió valores aproximados al esperado.

5. Análisis redes de Twitter España y redes equivalentes

En este último apartado vamos a generar redes equivalentes de las redes de Twitter España, Twitter Global. Para cada una de ellas hemos generado redes equivalentes con el mismo número de nodos y el número de aristas más aproximado posible.

- Red de Twitter España y equivalentes:

Para el caso de la red de Twitter España que constaba de 100 nodos y 781 aristas hemos generado una red aleatoria con $N = 100$ y $p = 0.15777$, y una red barabasi con $N = 100$ y $m = 8$. Parámetros calculados con las fórmulas teóricas para obtener el número de aristas más aproximado posible. Obteniendo finalmente las siguientes estadísticas:

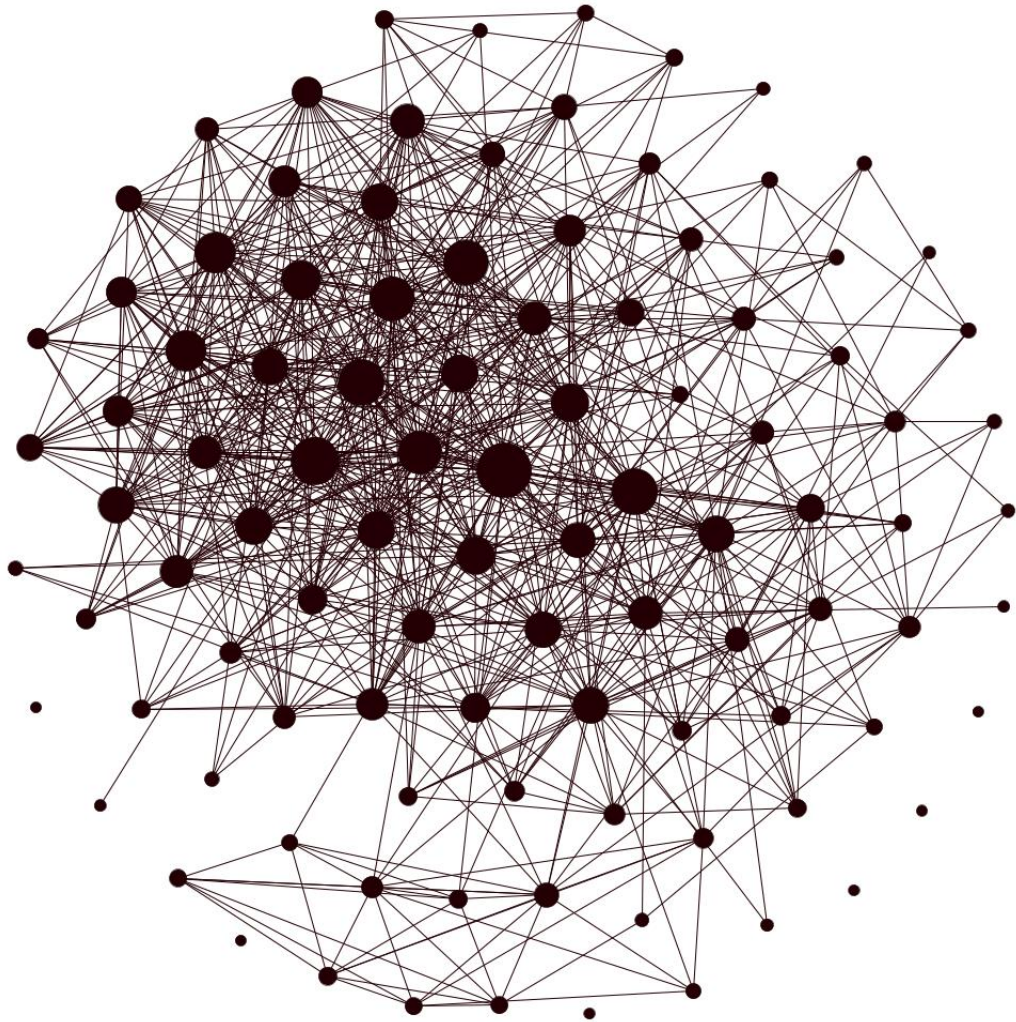
	Twitter España	Red Aleatoria Generada	Red Barabasi Generada
Número de nodos	100	100	100
Número de aristas	781	792	764
Grado Medio	15,62	15,84	15,28
Densidad	0,158	0,16	0,154
Hub de mayor grado	49	26	48
Distancia media	2,104	1,911	1,936
Coefficiente de clustering	0,502	0,164	0,283
Componentes conexos	7	1	1

Vemos como a nivel estadístico la red de Twitter de España se asemeja tanto a la red aleatoria como a la red de Barabasi, aunque si nos fijamos en los valores del Hub de mayor grado, ese dato se asemeja más con el obtenido en la red de Barabasi, ya que en twitter existen usuarios que son muy seguidos y actúan como hubs de la red como si de una red libre de escala se tratara.

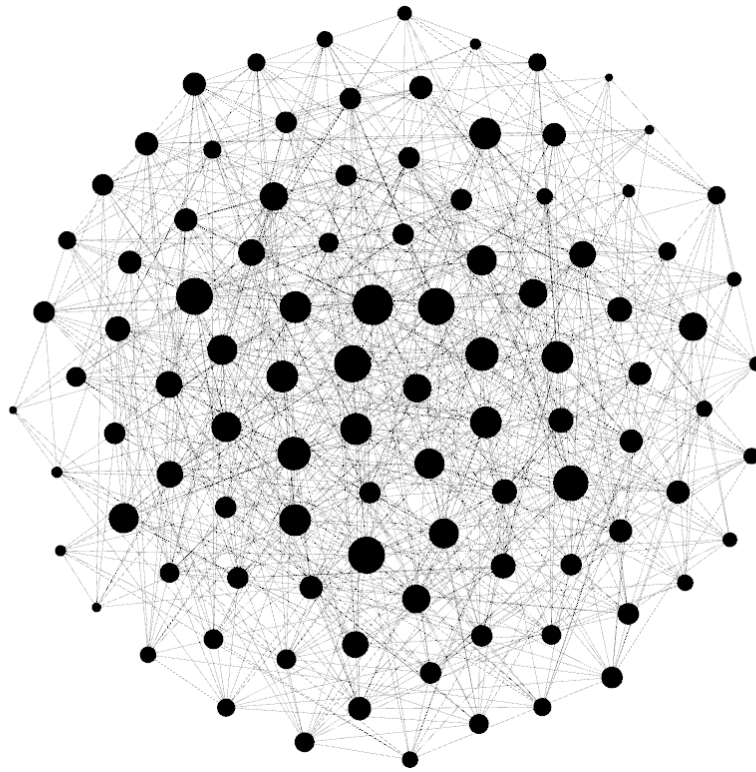
La diferencia más clara que vemos entre las redes generadas y la red de Twitter España es que ésta primera tiene 7 componentes conexos mientras que una red Barabasi siempre tiene un solo componente conexo y para el caso de la red aleatoria generada también resulta con solo un componente conexo debido a la alta probabilidad usada para poder obtener el número de aristas aproximado.

Sin embargo, si nos fijamos en las visualizaciones, la red de Twitter España se asemeja mucho más a una red aleatoria pues tiene una distribución de enlaces mucho más repartida y no está tan concentrada en los hubs como si se apreciara en una red Barabasi, como vemos a continuación.

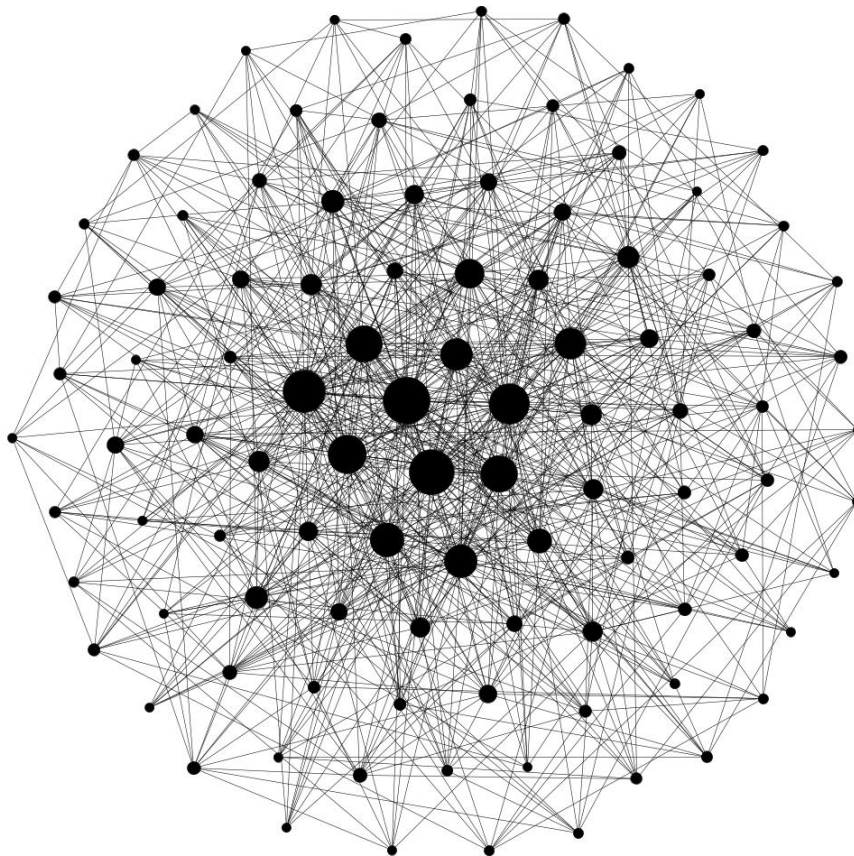
- Twitter España:



- Red aleatoria 100 nodos:

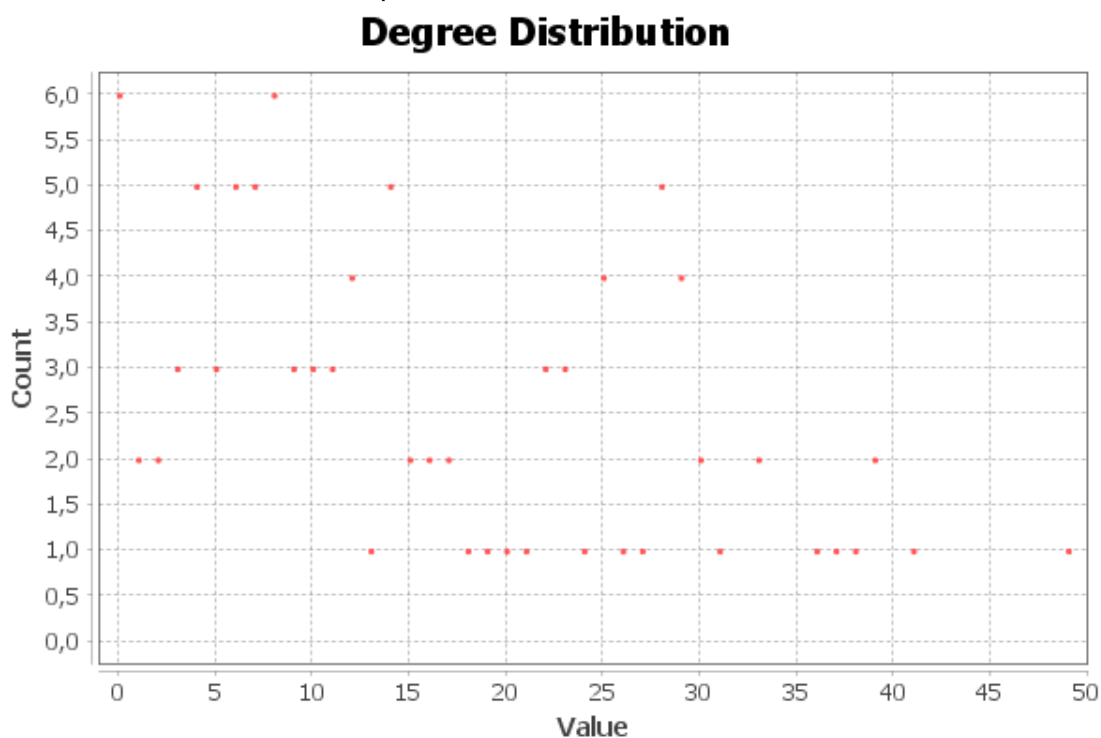


- Red Barabasi 100 nodos:

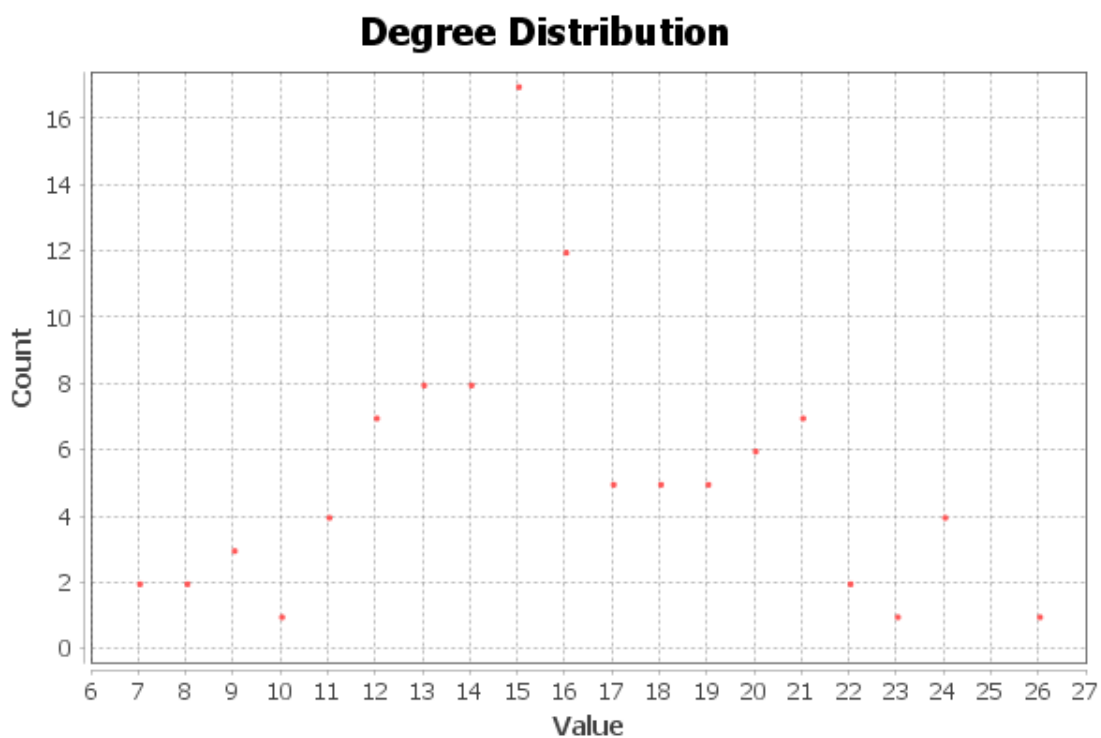


Por último, si nos fijamos en la distribución de grados vemos como la red de Twitter España no es parecida a ninguna de las generadas pues no se aprecia ninguna de las curvas características de las redes creadas, siendo casi una mezcla de las dos.

- Distribución Grados Twitter España:

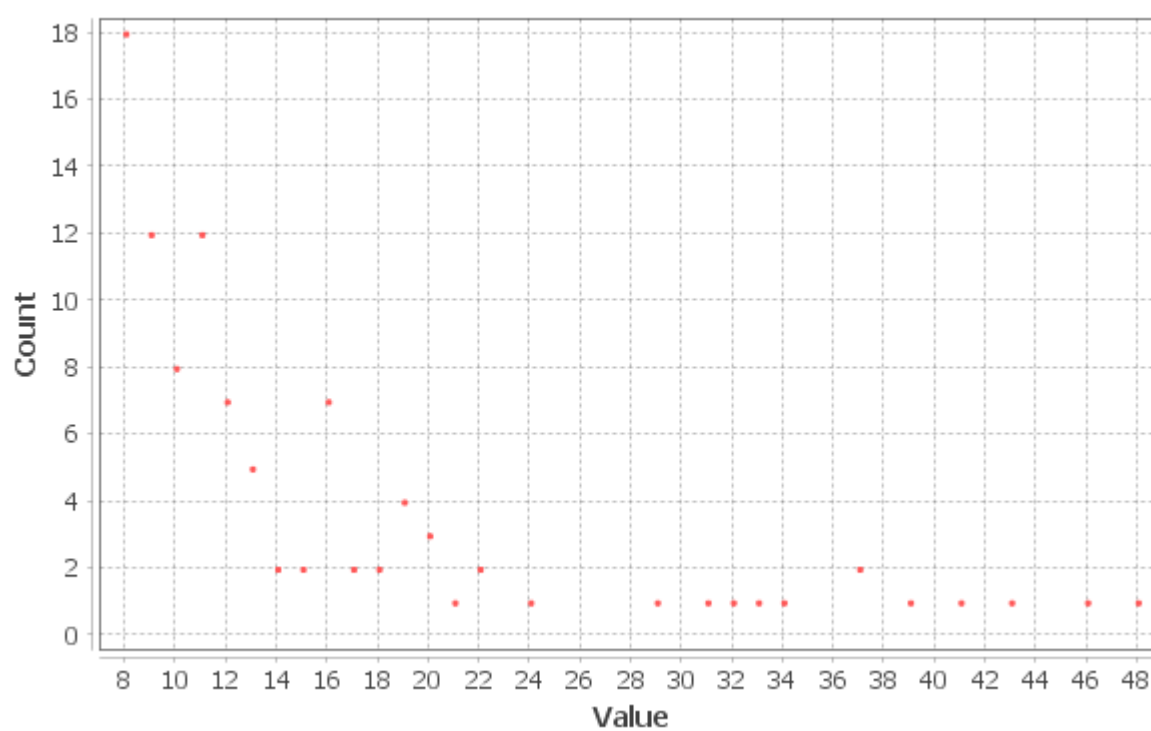


- Distribución Grados red aleatoria 100 nodos:



- Distribución Grados red Barabasi 100 nodos:

Degree Distribution



En resumen, la red de Twitter España es una mezcla de ambos modelos de redes pues estadísticamente parece una red libre de escala y visualmente se asemeja más a una red aleatoria, y a la vez su distribución de grados no se parece a la de ninguno de los modelos.

- **Red de Twitter Global y equivalentes:**

Para el caso de la red de Twitter Global que constaba de 120 nodos y 330 aristas hemos generado una red aleatoria con $N = 120$ y $p = 0.0462$, y una red barabasi con $N = 120$ y $m = 3$. Parámetros calculados con las fórmulas teóricas para obtener el número de aristas más aproximado posible. Obteniendo finalmente las siguientes estadísticas:

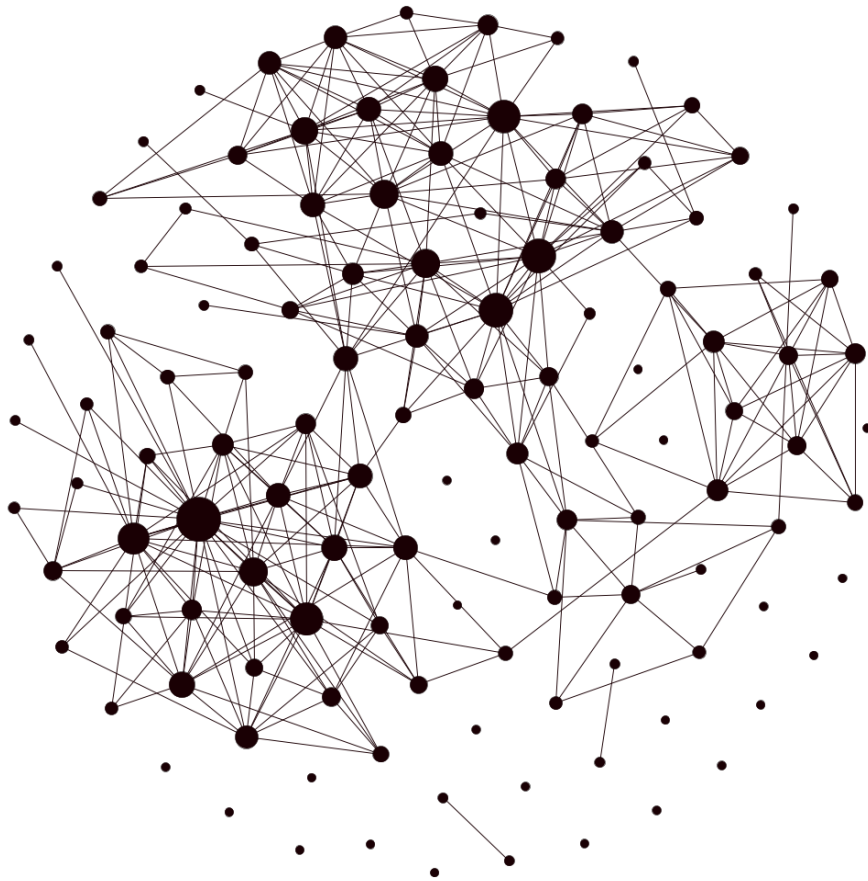
	Twitter Global	Red Aleatoria Generada	Red Barabasi Generada
Número de nodos	120	120	120
Número de aristas	330	353	354
Grado Medio	5,5	5,883	5,9
Densidad	0,046	0,049	0,050
Hub de mayor grado	25	13	28
Distancia media	3,457	2,901	2,68
Coeficiente de clustering	0,587	0,061	0,129
Componentes conexos	25	1	1

En el caso de la red de Twitter Global vemos cómo al igual que en el caso de la comparación de la red de Twitter España, la red de la práctica 1 se parece estadísticamente a la red aleatoria y la red de barabasi, aunque destacando una vez más un resultado muy similar en el Hub de mayor grado con la red de barabasi.

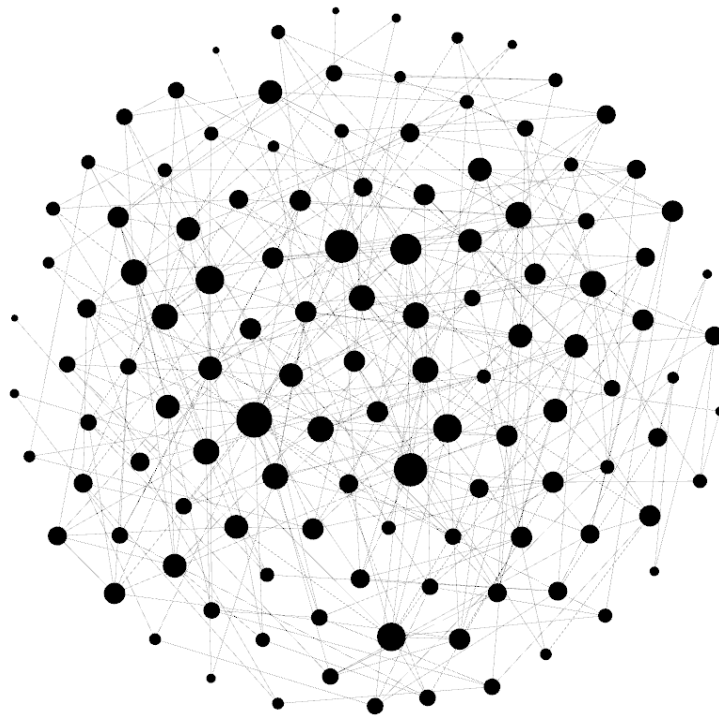
En cambio, al igual que en el caso anterior detectamos como existe una gran diferencia entre los componentes conexos de la red de Twitter y de las redes generadas, siguiendo las mismas razones del caso anterior.

A pesar de ello, nuevamente nos encontramos con la situación de que una red de Twitter visualmente se parece más a una red aleatoria, ya que tiene los grados de los nodos algo más distribuidos como vemos a continuación.

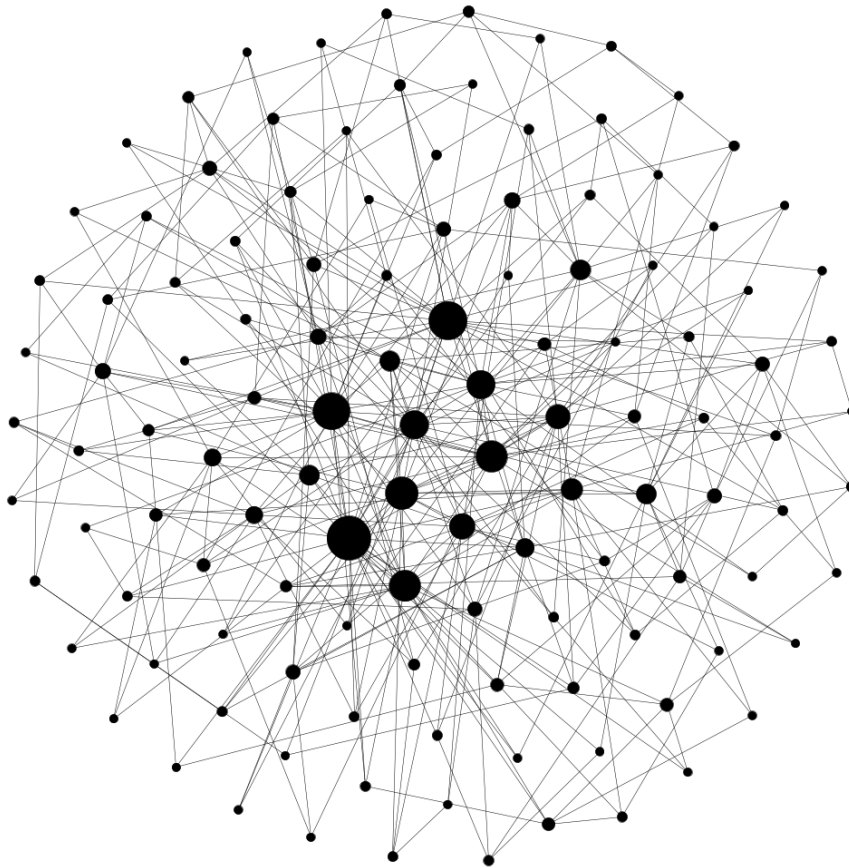
- Twitter Global:



- Red aleatoria 120 nodos:

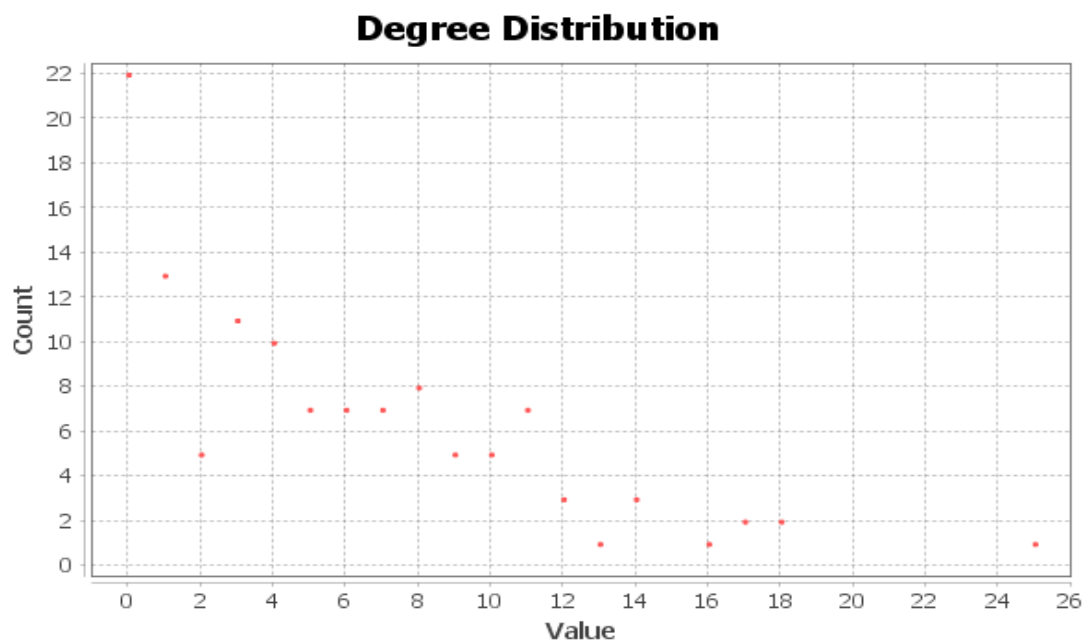


- Red Barabasi 120 nodos:

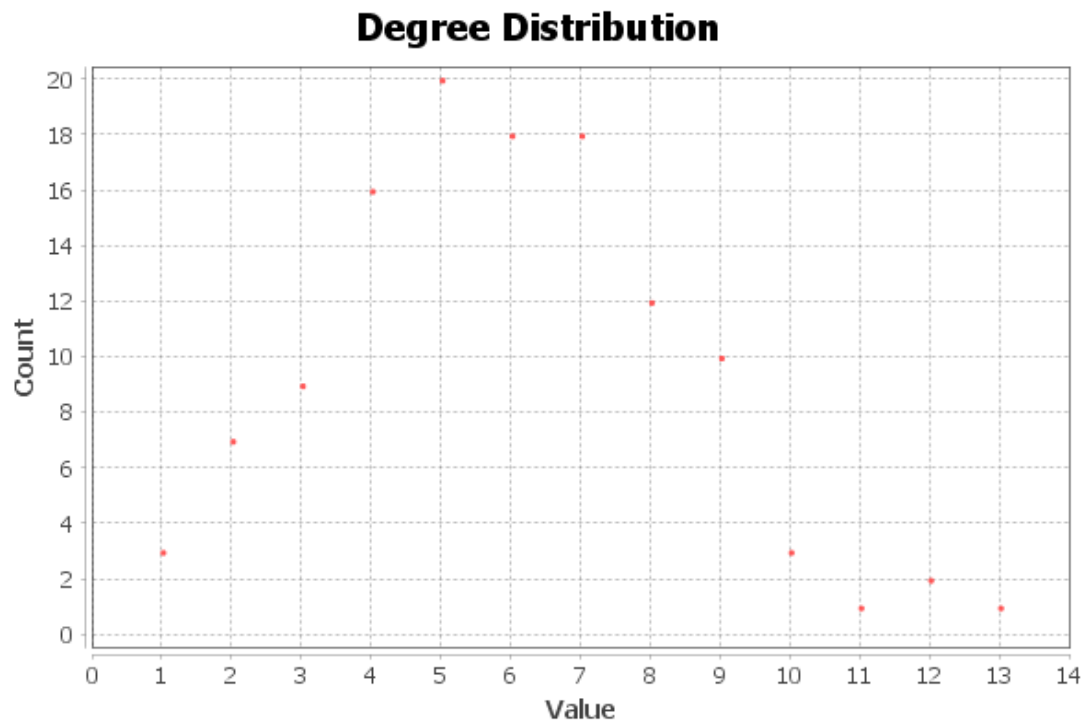


Por último, si nos fijamos en la distribución de grados vemos como la red de Twitter Global vuelve a no ser parecida a ninguna de las generadas pues no se aprecia ninguna de las curvas características de las redes creadas, aunque nuevamente se asemeja algo más a una red Barabasi.

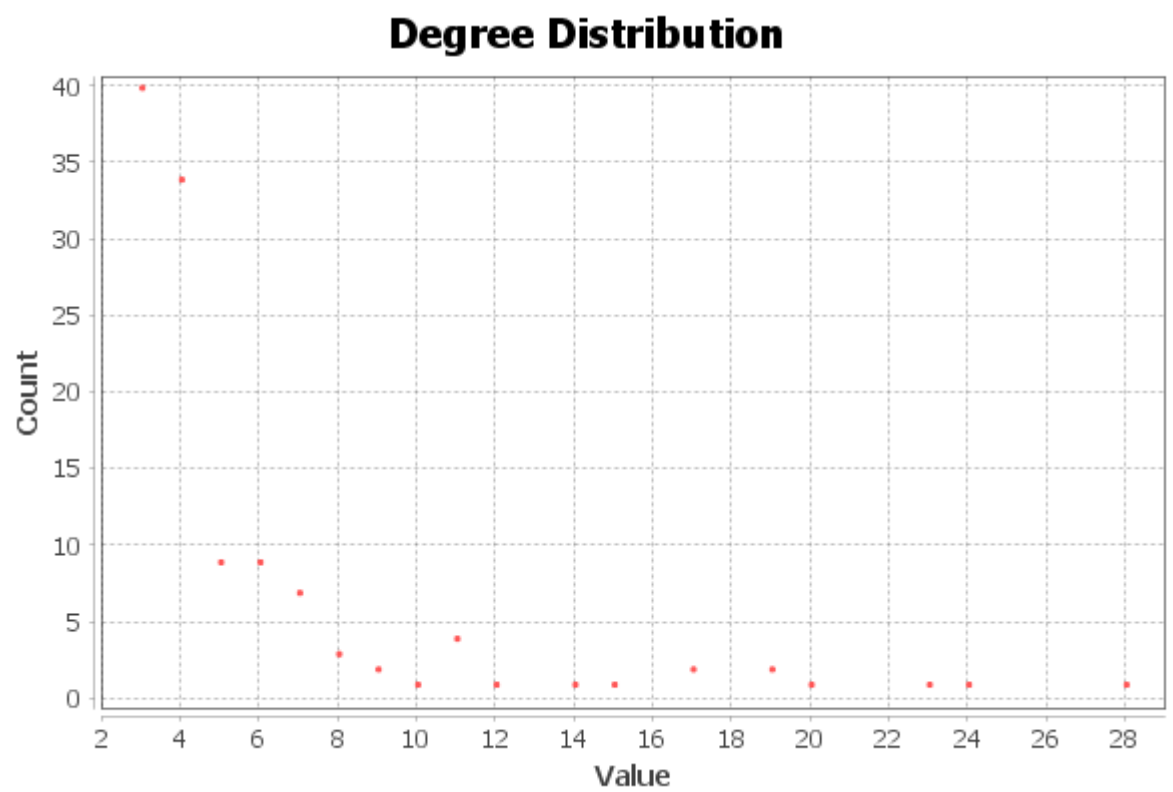
- Distribución Grados Twitter Global:



- Distribución Grados red aleatoria 120 nodos:



- Distribución Grados red Barabasi 120 nodos:



En resumen, la red de Twitter España es una mezcla de ambos modelos de redes pues estadísticamente parece una red libre de escala y visualmente se asemeja más a una red aleatoria, y a la vez su distribución de grados es más parecida a una red Barabasi. Aunque tenga más puntos a favor la red de Barabasi, no nos terminamos de decantar por ella ya que una Barabasi siempre tiene solo una componente conexa y vemos como en la red de Twitter Global existen hasta 25.