

Práctica 2: Análisis de Modelos de Redes



Grupo 13

Trabajo realizado por:

Alejandro Valencia

Alonso Mata

Introducción

En Matemáticas se denomina grafo aleatorio a un grafo que es generado por algún tipo de proceso aleatorio. La teoría de los grafos aleatorios cae en la intersección entre la teoría de grafos y la teoría de probabilidades y se fundamenta en el estudio de ciertas propiedades de los grafos aleatorios. Uno de los modelos matemáticos más aplicados en la generación de redes aleatorias es modelo Erdős–Rényi.

Una de las ramas más estudiadas en el área de las redes aleatorias es el de la teoría de la percolación (nivel de percolación) relacionado con el estudio de la fiabilidad en las redes de comunicación. Un campo de estudio inicial fue el de redes sociales, estudios sobre la topología de redes evolutivas como puede ser internet, etc. Se ha visto que algunas de las redes actuales crecen según modelos predefinidos en sus distribuciones de grado, como puede ser las redes libres de escala.

Un grafo aleatorio $G(N,p)$ es un grafo no dirigido donde N es el número de nodos y p es la probabilidad de que dos nodos estén conectados aleatoriamente. Esta aleatoriedad permite que pueda haber muchas instanciaciones distintas con un mismo valor de N y p cada uno con un número de enlaces L distinto. La probabilidad de generar un grafo aleatorio $G(N,p)$ con L enlaces es:

$$P(G(N, p)) = p^L (1 - p)^{(N(N-1)/2) - L}$$

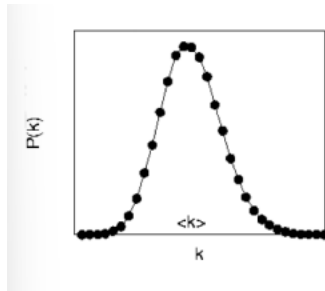
La probabilidad de obtener una red con L enlaces viene dada por esta otra función:

$$P(L) = \binom{\binom{N}{2}}{L} p^L (1 - p)^{\frac{N(N-1)}{2} - L}$$

Los grados de los nodos se son caracterizados en la distribución de grados p_k . Esta distribución representa la probabilidad de que un nodo aleatorio de la red tenga grado k .

$P(k)$ es la probabilidad de tener exactamente k enlaces en una red de N nodos la cual se define a partir del producto de los siguientes términos:

- p^k es la probabilidad de que k de sus enlaces estén presentes
- $(1-p)^{N-1-k}$ es la probabilidad de que los $N-1-k$ enlaces restantes falten.
- $\binom{N-1}{k}$ Es el factor combinacional que cuenta el número de combinaciones distintas en la que podemos elegir k enlaces de entre los $(N-1)$ enlaces que puede tener un nodo



Binomial distribution

$$P(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k}$$

Select k nodes from N-1 probability of having k edges probability of missing N-1-k edges

Teóricamente para valores altos de N y bajos para $\langle k \rangle$, $P(k)$ de las redes generadas aleatoriamente debería seguir distribución de Poisson:

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

Por otra parte, el coeficiente de clustering en redes aleatorias es bajo ya que los enlaces son independientes y tienen la misma probabilidad entre sí. El coeficiente de clustering es independiente del grado del nodo y decremente a medida que N aumenta. Esto no sucede en las redes reales. Teniendo esto en cuenta el coeficiente de clusteing C en una red aleatoria:

$$\langle L_i \rangle \cong p \frac{k_i(k_i - 1)}{2} \quad \Rightarrow \quad C_i = \frac{2\langle L_i \rangle}{k_i(k_i - 1)} = p = \frac{\langle k \rangle}{N}.$$

En esta práctica hemos realizado una labor comparativa entre distintos modelos de generación de redes y sus propiedades teóricas. Para esta comparación hemos generado redes con distintos modelos y la comparamos con la red sintética adaptada lo más posible a las características de la red utilizada en la Práctica 1.

En esta práctica hemos tenido que comparar distintas redes modelizadas de diferentes formas para poder analizarlas de la forma óptima y aprovechar así lo que cada modelo nos ofrece.

Para ello hemos hecho uso de librerías de Python como puede ser Networkx. El entorno en el que hemos trabajado ha sido en un notebook en Jupyter.

La teoría de grafos es una rama de las matemáticas. Formalmente, un grafo $G=(V,E)$ es una pareja ordenada en la que V es un conjunto no vacío de vértices y E es un conjunto de aristas.

Los datos de nuestro grafo de la práctica 1 con los que vamos a tener que comparar son:

Número de nodos: 445.0

Número de aristas: 1334

Camino mínimo medio: 3,64

Grado medio: 5.996

Radio de la red 1

Diámetro medio de la red 7.2

Número de componentes conexas: 22

Media del coeficiente de clustering 0.733

Más adelante voy a exponer de forma individual cada modelo y ventajas de cada uno.

Configuration Model

En ciencia de redes, el modelo de configuración es un método para generar redes aleatorias a partir de una secuencia de grados determinada. Es ampliamente utilizado como modelo de referencia para las redes sociales de la vida real, ya que permite al modelador incorporar distribuciones de grado arbitrario.

Generamos una red según la secuencia de grados de la red de la Práctica 1 a partir de una secuencia de grados aleatoria que sigue una distribución normal. Para realizar la distribución utilizaremos la función de ***numpy.random.normal(loc, scale, size)*** con los siguientes valores de desviación típica(*scale*) y media(*loc*) extraídos de la Práctica 1.

```
degrees = [val for (node, val) in G.degree()]  
np.std(degrees)
```

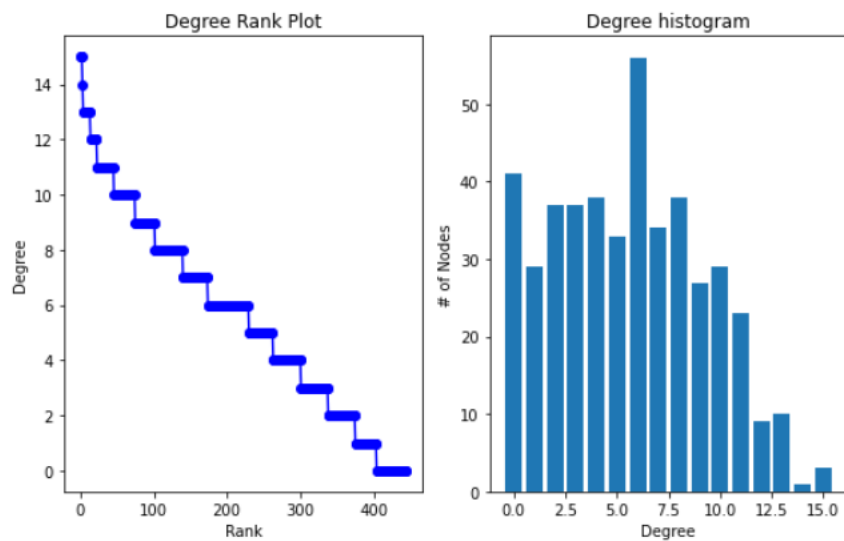
4.15216201974372

```
np.mean(degrees)
```

6.0

Cuando modelizamos las redes con esta configuración obtenemos los siguientes resultados:

Distribución de grado de la Red Nº 46



RED 1

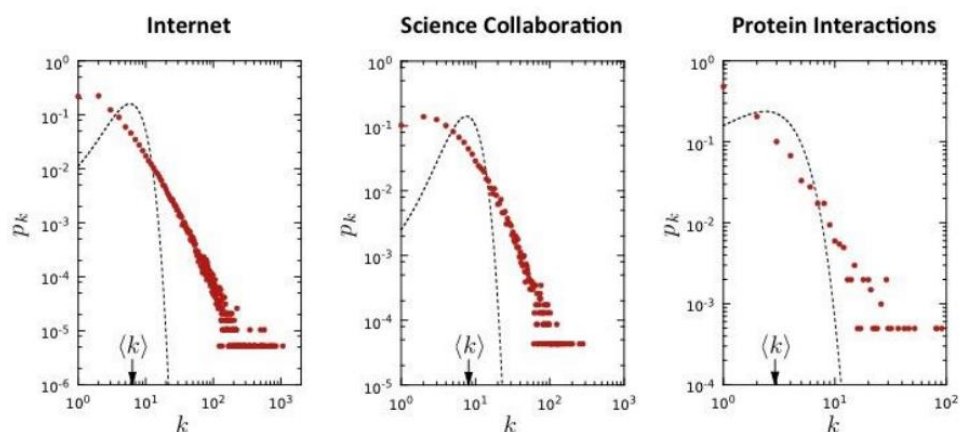
Número de nodos: 445.0
 Número de aristas: 1240.16
 Camino mínimo medio: 403.7
 Grado medio: 5.57
 Radio medio de la red 5.0
 Diámetro medio de la red 7.3
 Número de componentes conexas: 42
 Media del coeficiente de clustering 0.0164
 Coeficiente de clustering global 0.01939

El modelo de red aleatoria difiere de las redes reales en dos características importantes:

Crecimiento: Mientras que el modelo de red aleatoria asume que el número de nodos es fijo (invariante en el tiempo), las redes reales son el resultado de un proceso de crecimiento que aumenta continuamente. En este caso el número de nodos es un valor constante tomado de la red que definía el comportamiento de los tejedores sociales.

Conexión preferencial: Mientras que los nodos en redes aleatorias eligen aleatoriamente su compañero de interacción, en redes reales los nuevos nodos prefieren vincularse a los nodos más conectados.

Las redes reales se no siguen una distribución de Poisson si no que siguen más una forma logarítmica:



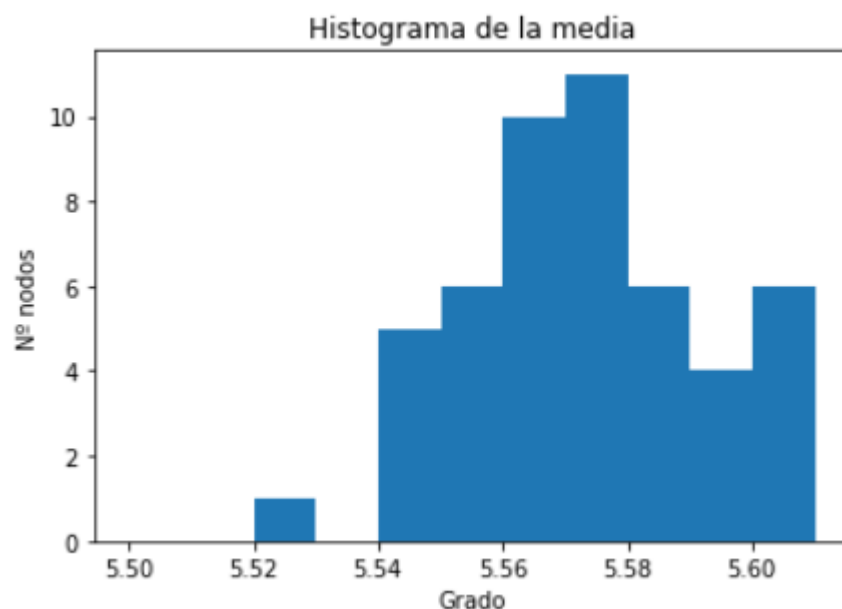
Como podemos ver las características de esta red no distan mucho de la vista en la P1. El número de aristas ($1270 < 1334$) es algo más escaso pero muy aproximado y en consecuencia el grado medio de la red ($5.71 < 6.00$). El camino medio es mayor. En cambio, choca con el diámetro de la red que es igual.

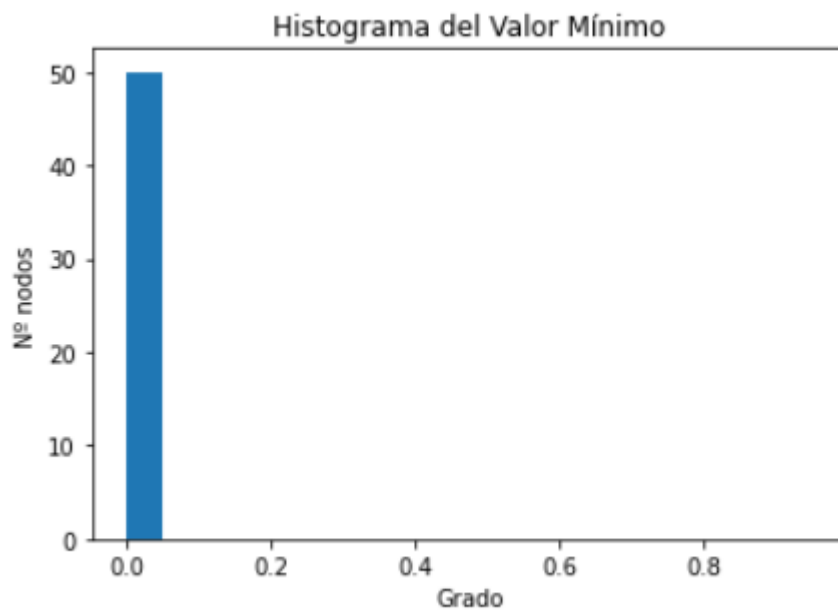
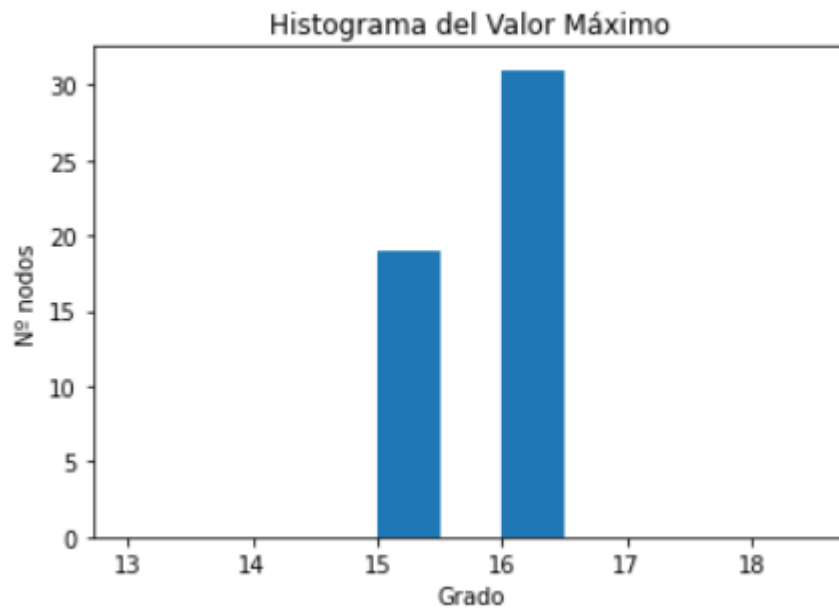
El camino mínimo medio es mucho mayor que el obtenido en la P1 lo que coincide con el menor valor en el coeficiente de clustering como se espera. Esto tiene sentido ya que al tener menos enlaces y nodos relevantes en la red afecta a la hora de encontrar caminos óptimos entre dos nodos de la red.

La distribución de grado se parece a la de las redes reales lo que puede significar que el grado medio $\langle k \rangle$ es elevado respecto al valor de N .

Esta red se asemeja al modelo teórico para el configuration model.

Datos de las 50 redes





Modelo Erdős–Rényi

En teoría de grafos el modelo Erdős–Rényi es uno de los métodos empleados en la generación de grafos aleatorios. En este modelo se tiene que un nuevo nodo se enlaza con igual probabilidad con el resto de la red, es decir posee una independencia estadística con el resto de los nodos de la red. Hoy en día se emplea como una base teórica en la generación de otras redes.

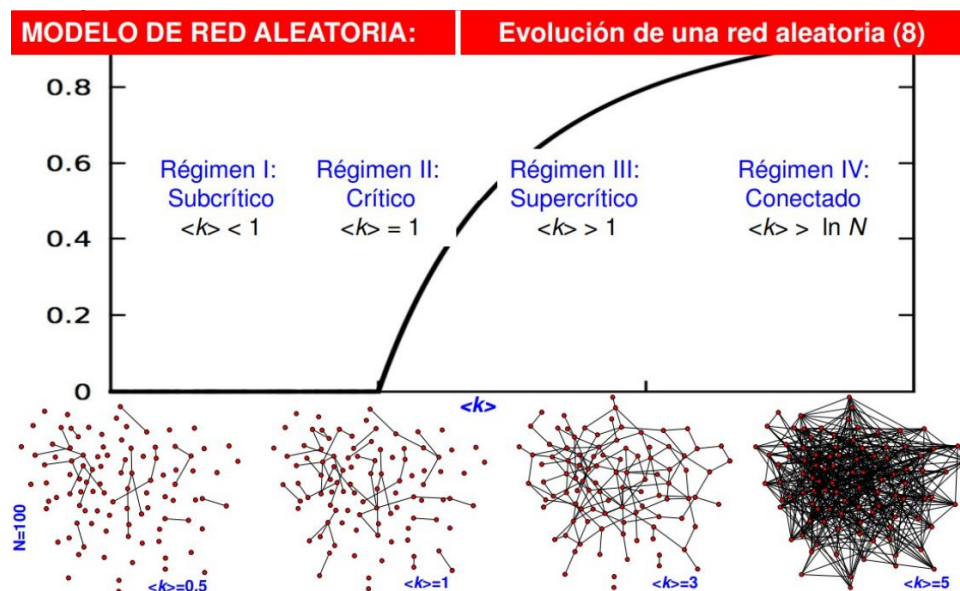
Si consideramos N nodos de una red sin conectar y distribuidos de forma aleatoria, podemos imaginar que en un instante inicial enlazamos dos cualesquiera, de esta forma en pasos sucesivos vamos enlazando aleatoriamente de dos a dos nodos. Los nodos que se encuentren enlazados se

descartan. Si repetimos el proceso M veces eligiendo un par de nodos en cada turno al final habremos establecido como máximo M enlaces entre parejas de nodos. Si M es un valor pequeño con respecto al valor total de nodos muchos de los nodos estarán desconectados, mientras que por el contrario otros nodos estarán formando pequeñas islas.

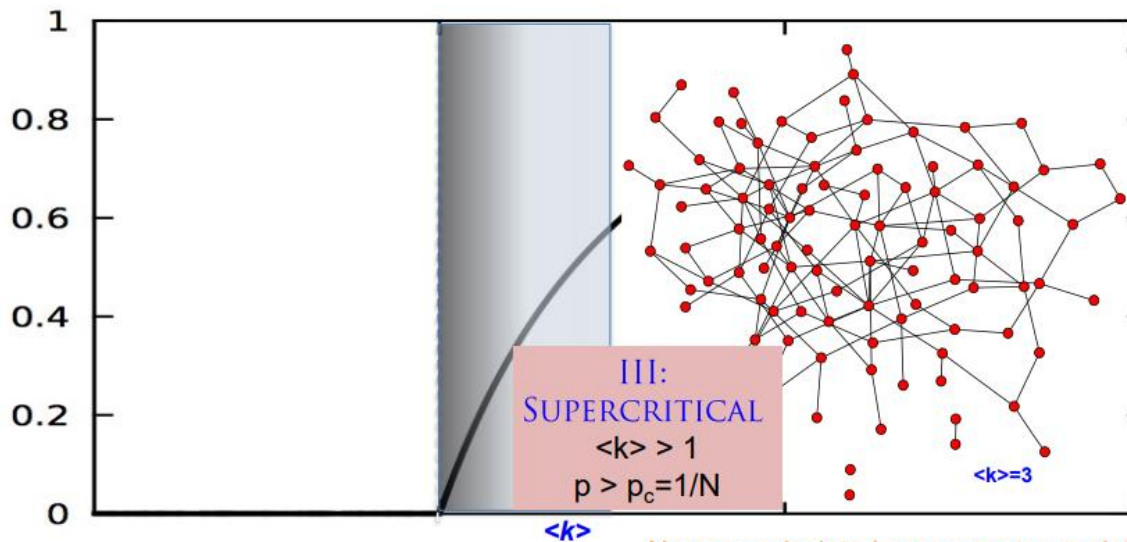
Por el contrario, si M es grande en comparación con N el número total de nodos es muy posible que casi todos los nodos estén enlazados entre sí. Cuando se enlazan los nodos de esta forma aparecen propiedades específicas en la distribución de grado $P(k)$. Durante muchas décadas a partir de los años 1950 se pensó que las redes con esta característica eran las más adecuadas para describir ciertas redes complejas y pronto se vio que no era del todo cierto.

Erdos and Renyi determinaron que el valor de $\langle k \rangle$ crítico es 1. Lo cual significa que a partir de ese valor el grado de la red aumenta rápidamente con lo que emerge un cluster grande denominado componente gigante.

En la primera práctica vimos como la red tenía varias componentes gigantes con una notablemente más grande. Por tanto, para la comparación la haremos tomando la probabilidad como parámetro de distintas regiones de la evolución de la red aleatoria. En esta ocasión tomaremos el régimen supercrítico y conectado ya que son los casos que más se asemejan a la red de la práctica 1 sobre todo el primero, ósea $\langle k \rangle > 1$ y $\langle k \rangle > \ln N$. Para la primera red $\ln(1499) = 7,313$. Para la segunda red $\ln(444) = 6,096$. Tomaremos $\langle k \rangle = \ln(N)/2$ para el régimen supercrítico y $\langle k \rangle = \ln(N)+1$.



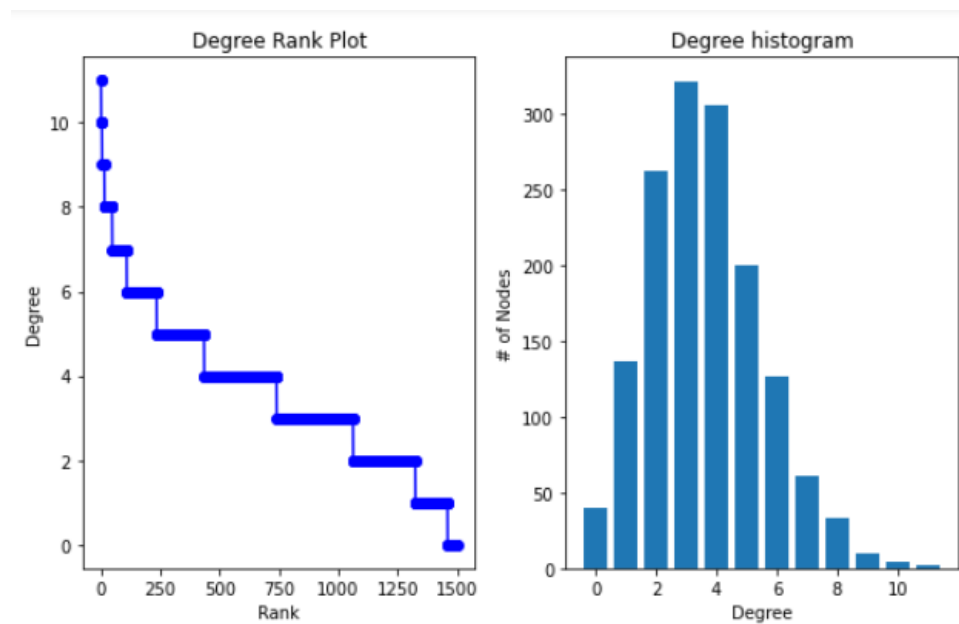
Comparación de distintas redes en etapa supercrítica



Unique giant component: $N_G \sim (p - p_c)N$

→ GC has loops.

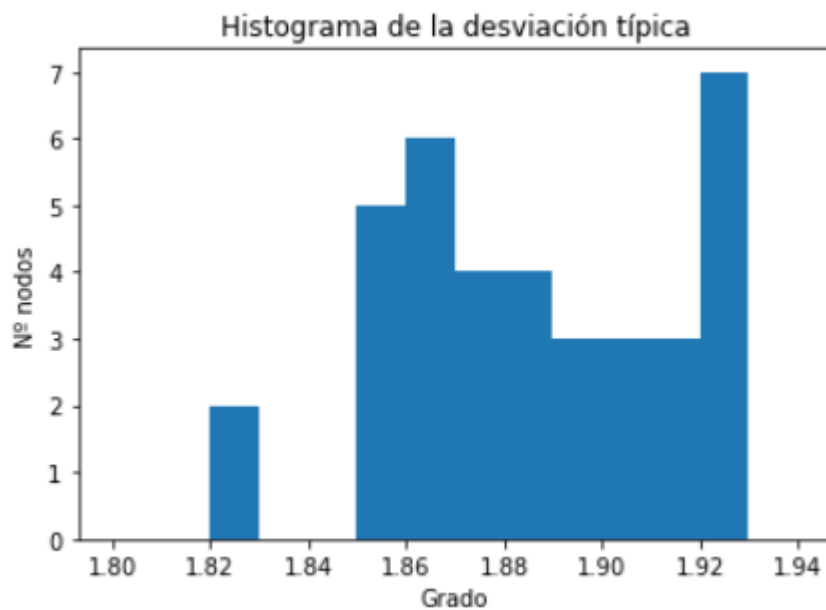
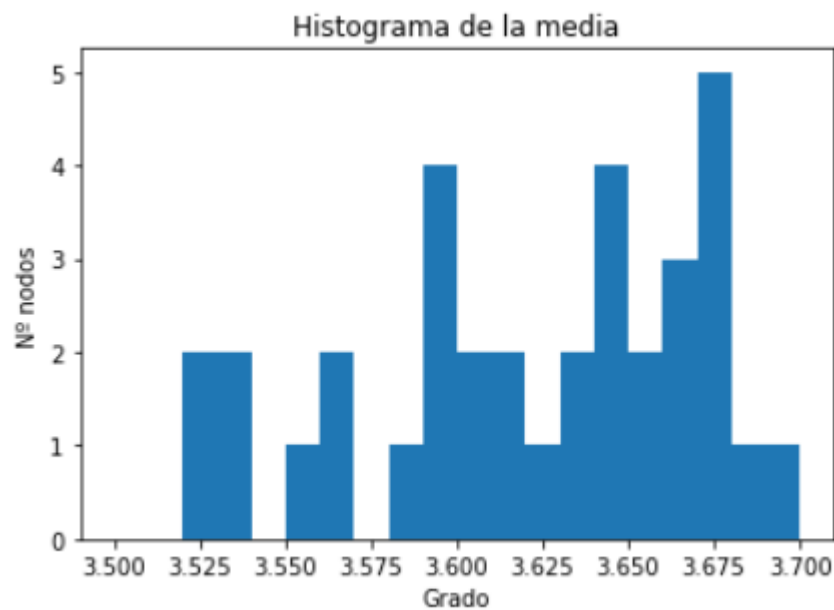
Numerous isolated components coexist with the giant component (which increases its size as $\langle k \rangle$ increases)

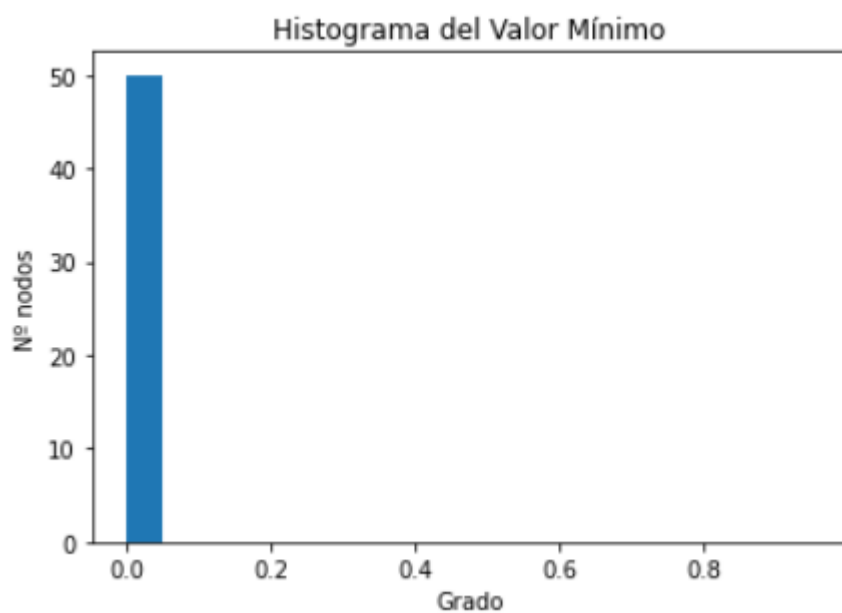
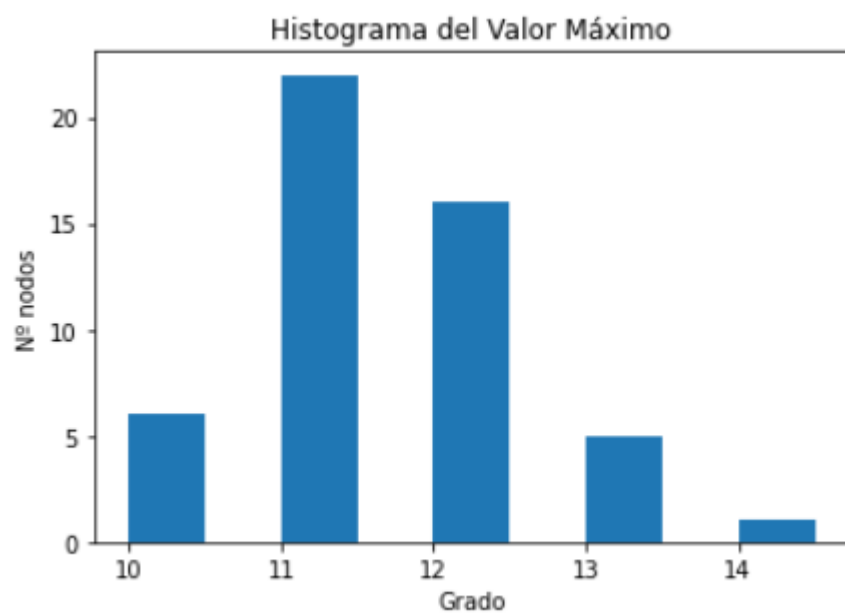


RED 1

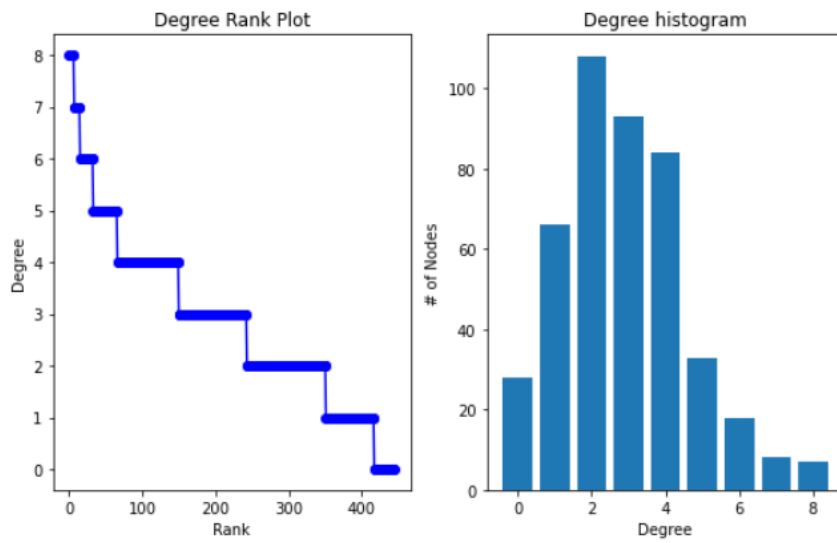
Número de nodos: 1500.0
Número de aristas: 2742.68
Camino mínimo medio: 1457.6
Grado medio: 3.66
Radio medio de la red 7.9
Diámetro medio de la red 12.3
Número de componentes conexas: 41
Media del coeficiente de clustering 0.00209
Coeficiente de clustering global 0.00237

Datos de las 50 redes





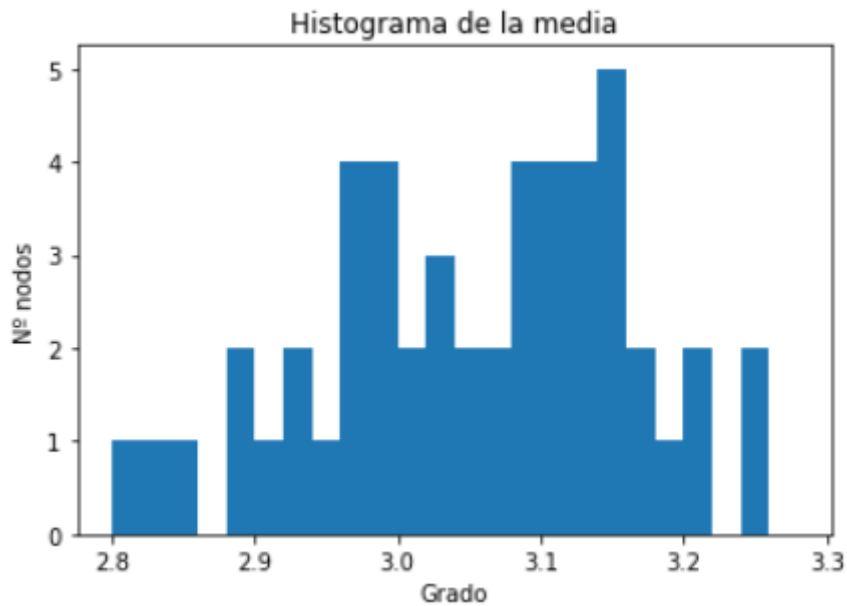
Distribución de grado de la Red2 Nº 26

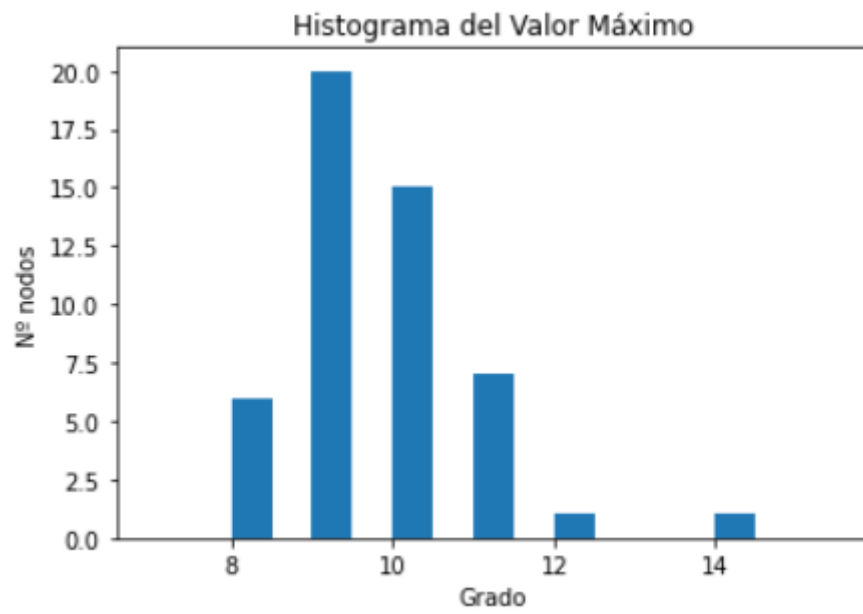
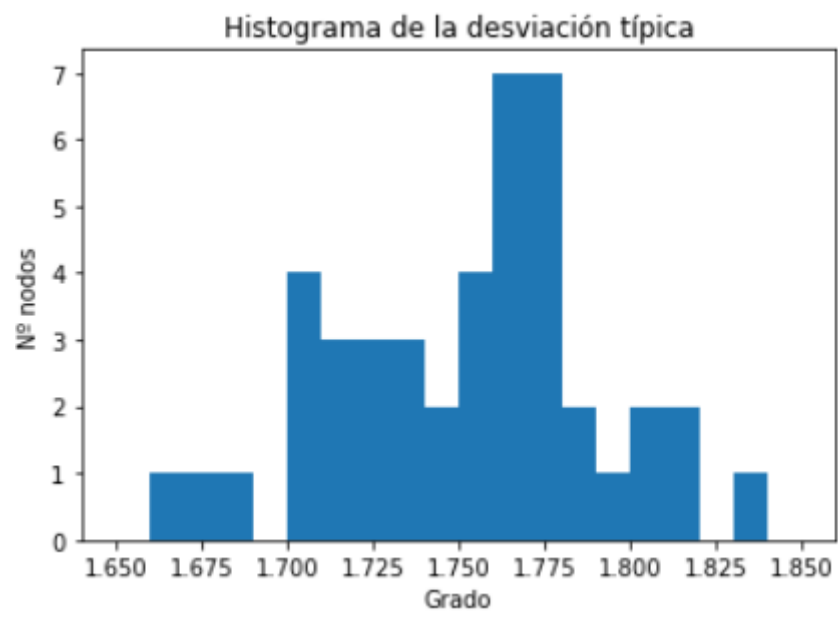


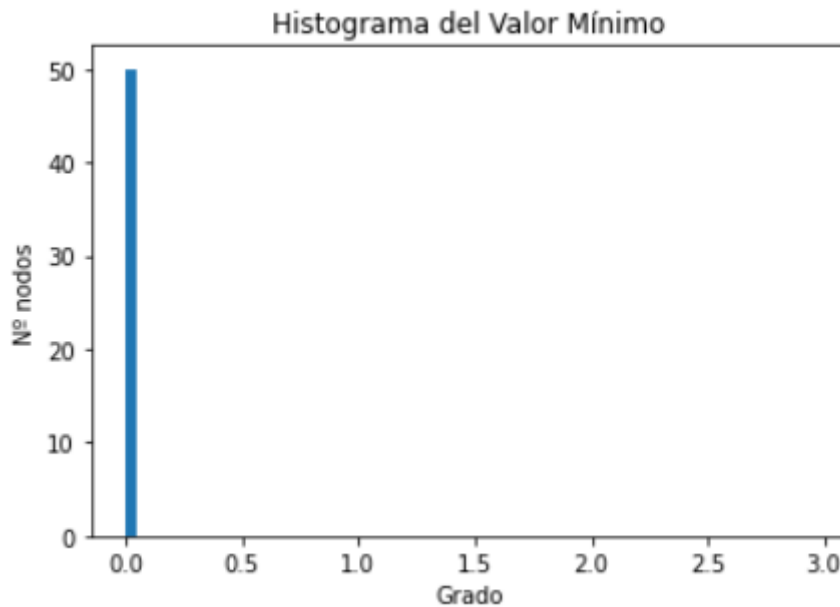
RED 2

Número de nodos: 445.0
 Número de aristas: 679.64
 Camino mínimo medio: 420.7
 Grado medio: 3.05
 Radio medio de la red 7.2
 Diámetro medio de la red 12.4
 Número de componentes conexas: 39
 Media del coeficiente de clustering 0.00532
 Coeficiente de clustering global 0.00684

Datos de las 50 redes





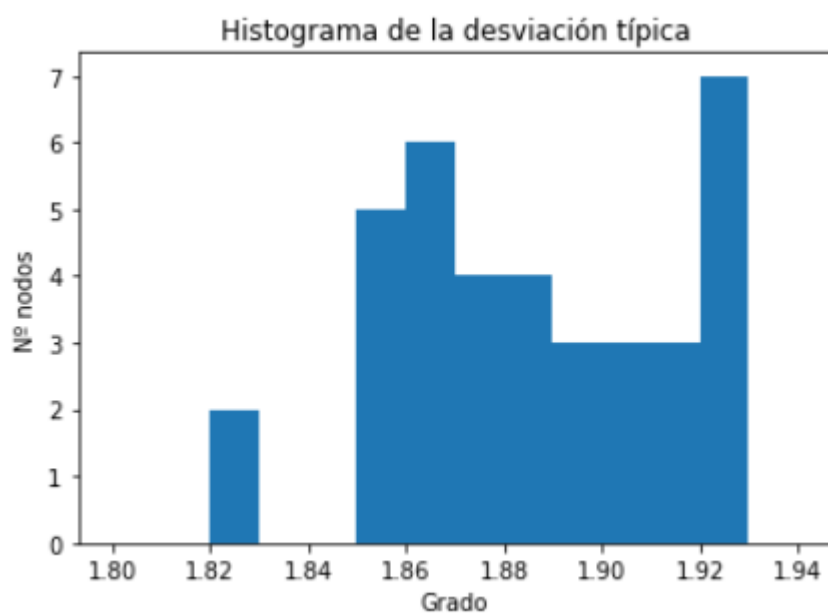
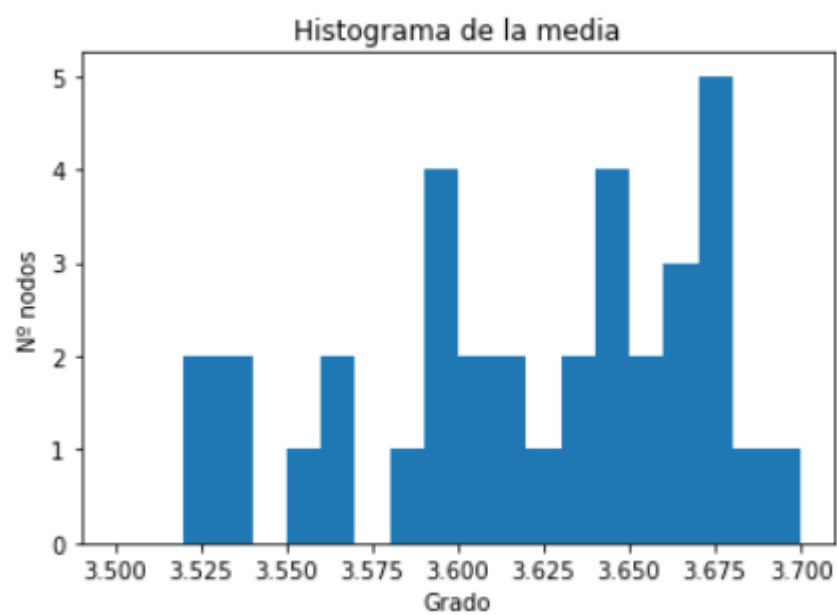


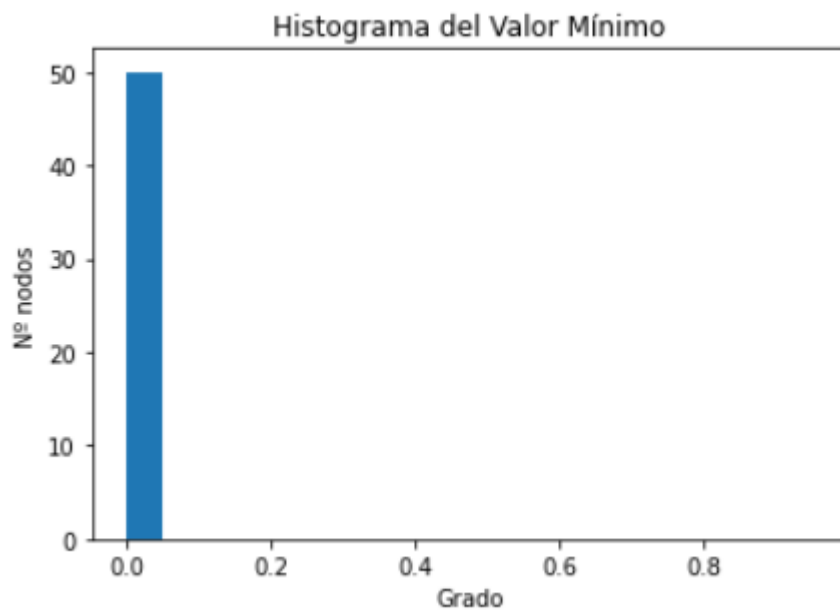
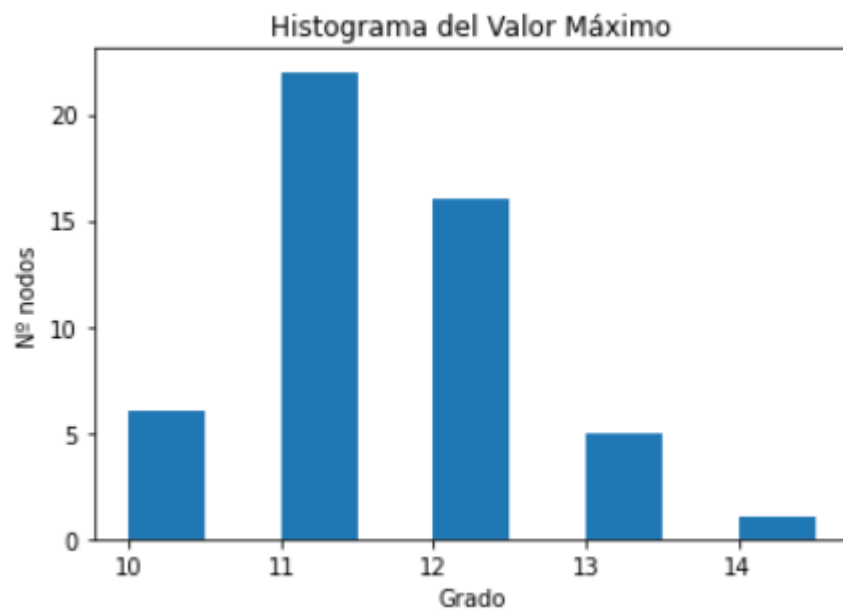
Aquí hemos usado una probabilidad muy baja para que los nodos se enlacen por eso que el grado medio sea tan bajo. Este modelo se ha comportado de la manera esperada, de forma parecida al modelo teórico. Tenemos el mismo número de nodos, pero el número de aristas es mucho mayor.

En ambos casos vemos como el grado medio ronda el valor de 3, la mitad del grado medio en la P1 y el doble de componentes en la red. Podríamos pensar a partir de estos datos que ambas variables son inversamente proporcionales. En general todos los datos obtenidos son proporcionales entre las dos redes por lo que podemos asegurar que el número de nodos no desvía de forma notable el resultado del modelo. Llama la atención que ambas tengan prácticamente el mismo número de componentes ya que al haber más nodos en la primera la probabilidad de que dos componentes se enlacen es mayor aun sabiendo que no hay conexión preferencial. Si es verdad que el grado medio tampoco es muy elevado y el tamaño del componente gigante crece con $\langle k \rangle$. Esta puede ser la causa de que el coeficiente de clustering sea tan bajo y de que haya tantos componentes independientemente del número de nodos.

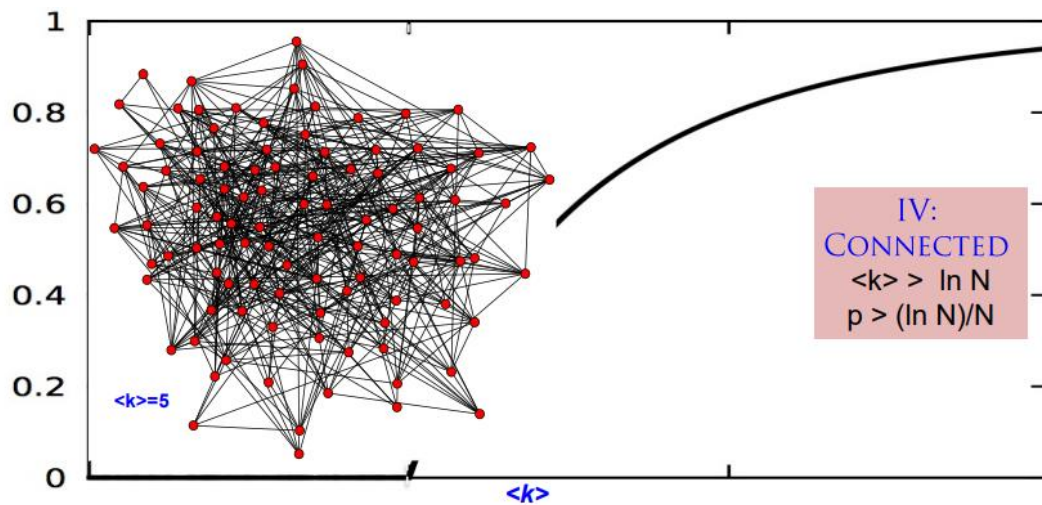
Este modelo dista mucho del grafo original.

Datos de las 50 redes



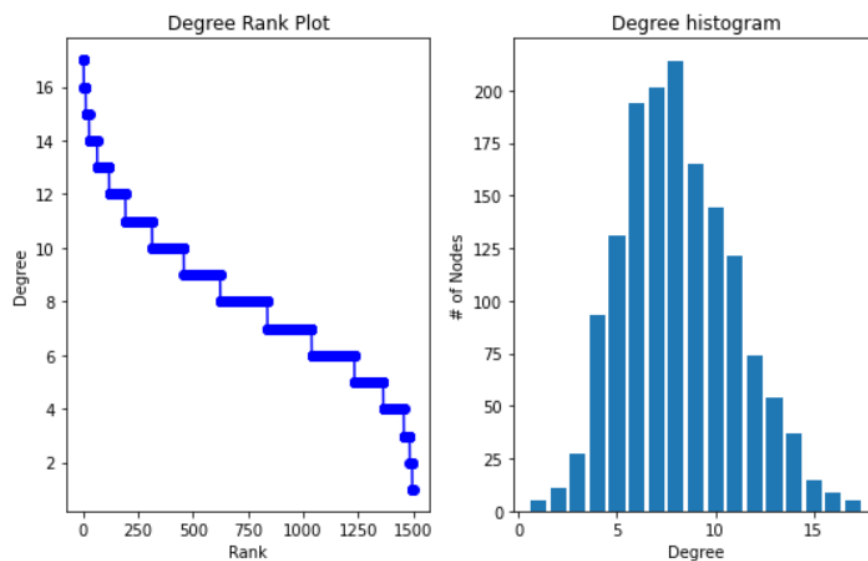


Comparación de distintas redes en etapa conectado



Only one cluster: $N_G = N$
 \rightarrow GC is dense.
 Cluster size distribution: None

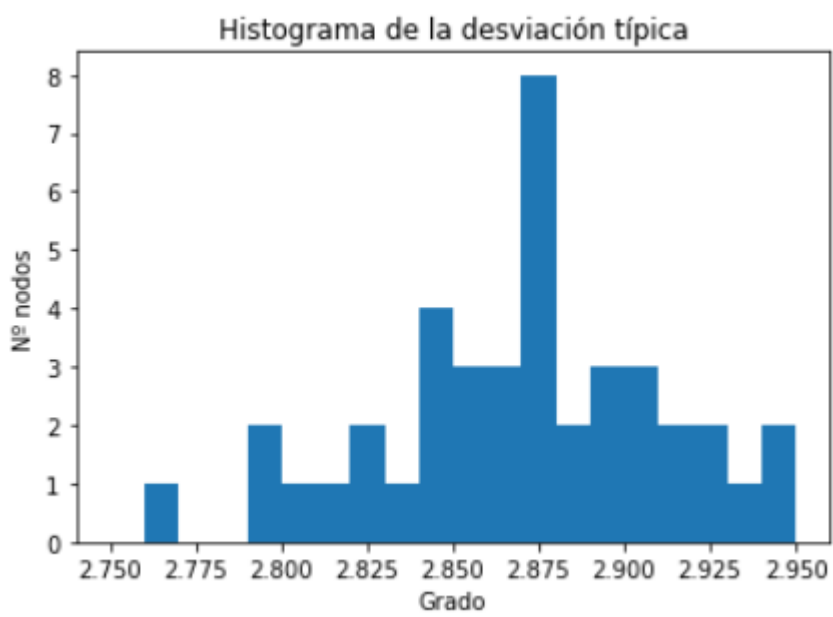
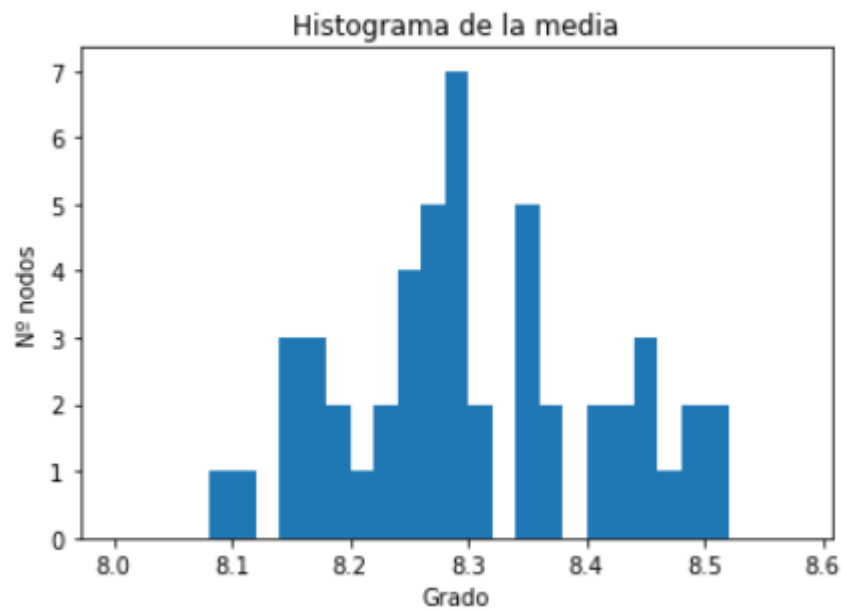
Distribución de grado de la Red1 Nº 28

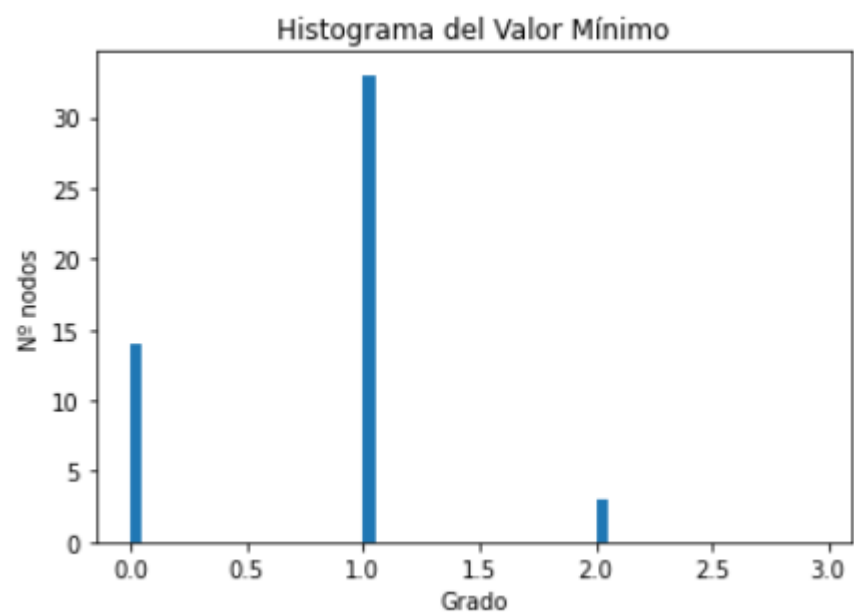
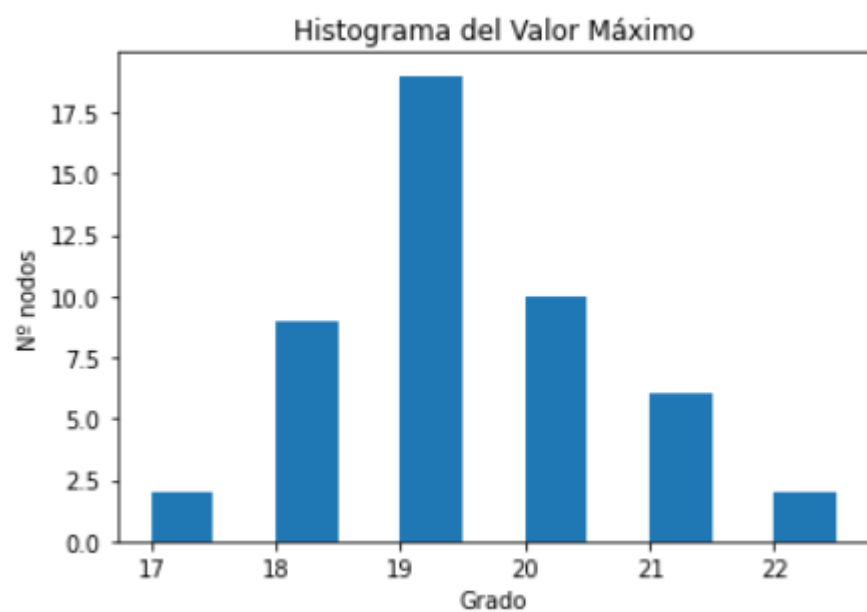


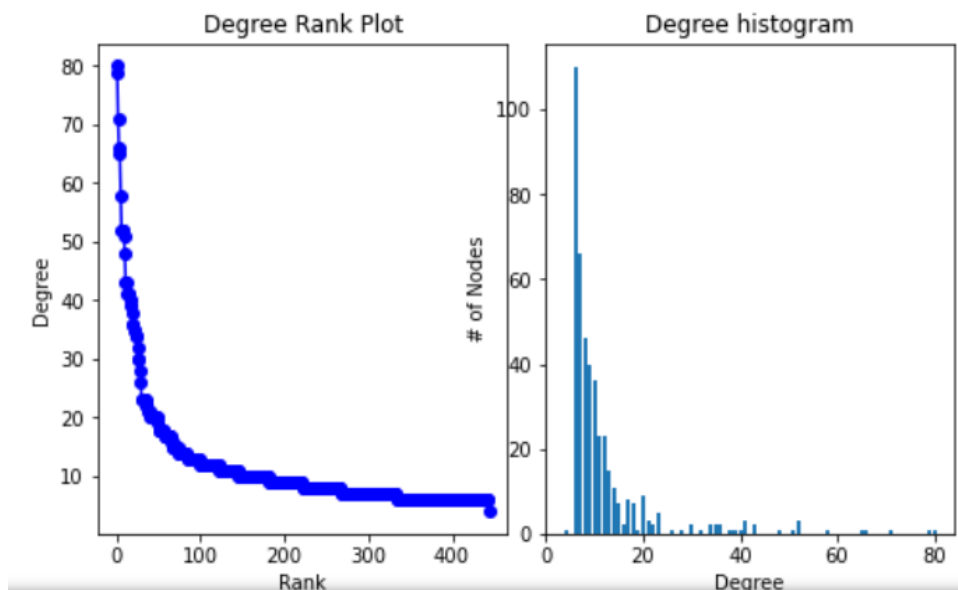
RED 1

Número de nodos: 1500.0
 Número de aristas: 6227.94
 Camino mínimo medio: 1499.7
 Grado medio: 8.3
 Radio medio de la red 5.0
 Diámetro medio de la red 6.5
 Número de componentes conexas: 1
 Media del coeficiente de clustering 0.00544
 Coeficiente de clustering global 0.00548

Datos de las 50 redes







RED 2

Número de nodos: 445.0
 Número de aristas: 2634.0
 Camino mínimo medio: 445.0
 Grado medio: 11.84
 Radio medio de la red 3.0
 Diámetro medio de la red 4.0
 Número de componentes conexas: 1
 Media del coeficiente de clustering 0.07945
 Coeficiente de clustering global 0.0632

Datos de las 50 redes

11.838 Es la media de las 50 redes

10.447 Es la desviación típica de las 50 redes

80 es el valor máximo y 4 el mínimo de las 50 redes generadas

Podemos observar que solo hay una componente como se suponía teóricamente. Al estar todos los nodos conectados el diámetro de la red es muy bajo y por tanto el camino corto entre distintos nodos también lo será. En este caso el camino corto devuelve cada uno de los nodos ya que todos están conectados y forman parte del componente máximo.

También vemos como el coeficiente de clustering es más bajo en la segunda red puede ser debido a que la relevancia está más repartida entre los nodos de la red al ser más. Si nos fijamos en el histograma de grado en la segunda red hay nodos con grado hasta 80 mientras que en la primera no supera el 20 y en la figura de rangos se ve como hay más nodos con grado por encima de la media mientras que en la primera red la mayoría de los nodos rondan un valor similar a la media y la cantidad de nodos con ese grado va disminuyendo según se aleja del valor de grado medio de la red.

En esta región $\langle k \rangle$ sigue una función logarítmica según el valor de N por tanto hemos obtenido un resultado que lo confirma ya que cuanto mayor es N menor es la probabilidad de generar un enlace $p = \langle k \rangle / (N-1)$.

Llama la atención que en este caso el grado medio no solo supere el de la segunda red si no que lo haga casi al 150% teniendo en cuenta que el valor de $\langle k \rangle$ es proporcional a N aunque al ser un logaritmo cuanto mayor sea el valor de N menor será la diferencia de $\langle k \rangle$ con $N-1$.

En esta representación de los modelos podemos darnos cuenta de que en esta fase tampoco nos acercamos al grafo original. En esta etapa de este modelo lo único que hemos podido acercar al valor deseado es el coeficiente de clustering que es el más característico de las redes reales y el más difícil de modelar con redes aleatorias, pero se ha necesitado duplicar prácticamente el grado medio de la red para llegar a un valor del coeficiente de clustering aun así diez veces menor del deseado

Modelo Barabasi-Albert

En teoría de redes se denomina Modelo de Barabási–Albert como un algoritmo empleado para generar redes aleatorias complejas libres de escala empleando una regla o mecanismo denominado conexión preferencial. Las redes generadas por este algoritmo poseen una distribución de grado de tipo potencial y se denominan: redes libres de escalas. Las redes de este tipo son muy frecuentes en los sistemas elaborados por el ser humano, así como en la naturaleza. Ejemplos de sistemas de este tipo son Internet, el world wide web, redes de citas, y algunas redes sociales, redes eléctricas.

Muchas de las redes observadas en la naturaleza caen dentro de la clase denominada como "redes libres de escala", esta afirmación viene a decir que sus distribuciones de grado siguen leyes de potencias (o libres de escala), mientras que otros modelos de grafos aleatorios tal y como el modelo Erdős–Rényi (ER) y el de Watts-Strogatz (WS) no exhiben tal característica de ley de potencias. El modelo de Barabási–Albert es uno de los propuestos para la generación de redes libres de escala. Incorpora dos conceptos generales: crecimiento y conexión preferencial (preferential attachment). Ambos conceptos pueden encontrarse extensivamente en las redes reales que nos rodean. La propiedad de crecimiento en teoría de redes significa que las redes poseen una cantidad de nodos creciente.

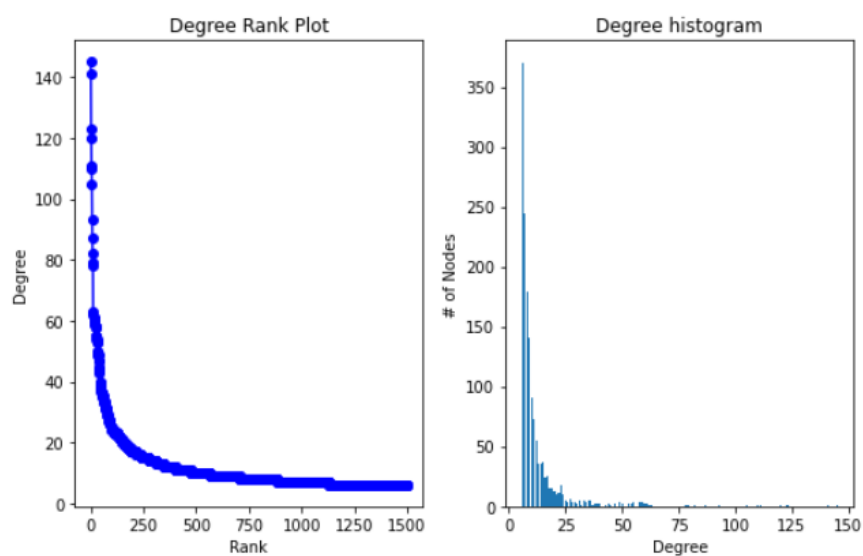
Crecimiento: en cada instante de tiempo añade un nuevo nodo con m enlaces que conectan con otros nodos de la red.

Conexión preferencial: la probabilidad de que un nodo tenga grado k es proporcional a sí misma.

Al cabo de un tiempo t este modelo genera $N = t + m_0$ y $L(t) = m_0 + mt$.

Para $m=6$

Distribución de grado de la Red1 Nº 42

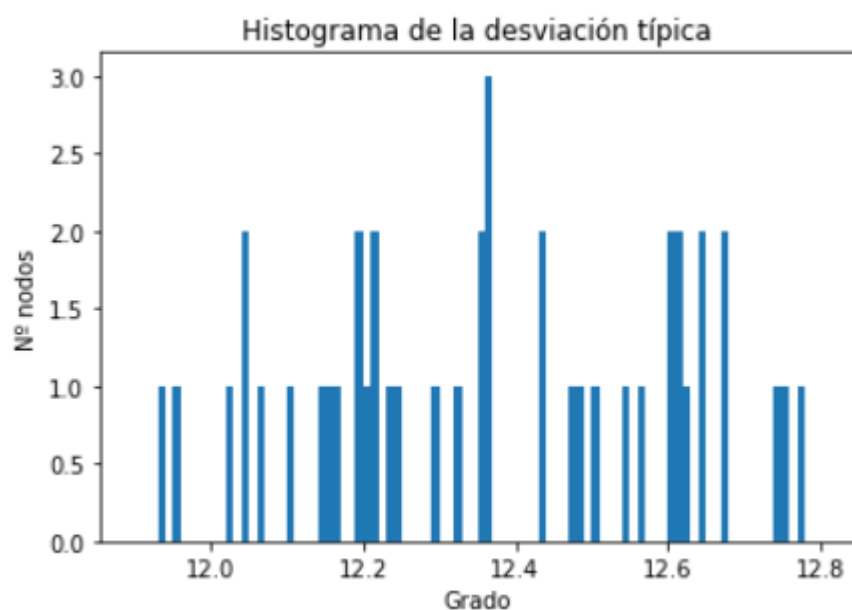


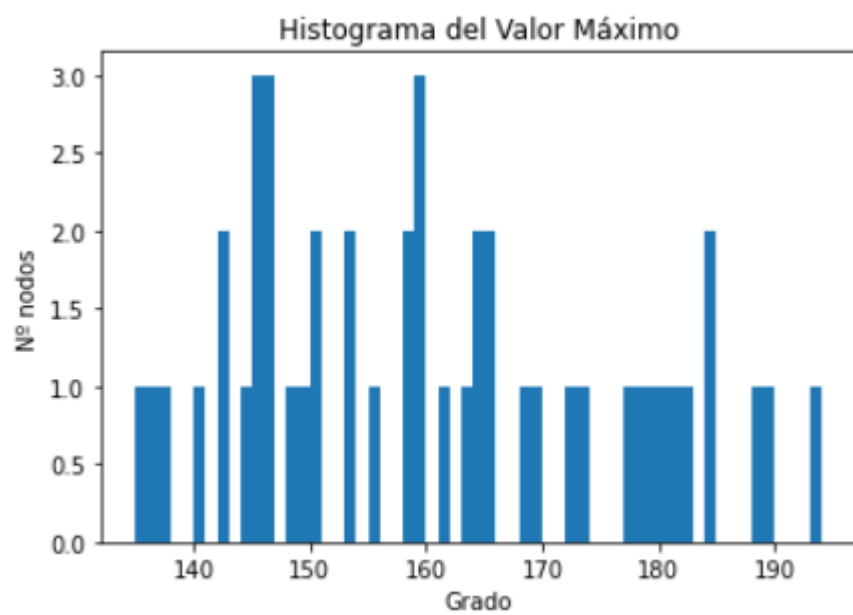
RED 1

Número de nodos: 1500.0
 Número de aristas: 8964.0
 Camino mínimo medio: 1500.0
 Grado medio: 11.95
 Radio medio de la red 3.0
 Diámetro medio de la red 5.0
 Número de componentes conexas: 1
 Media del coeficiente de clustering 0.03319
 Coeficiente de clustering global 0.02634

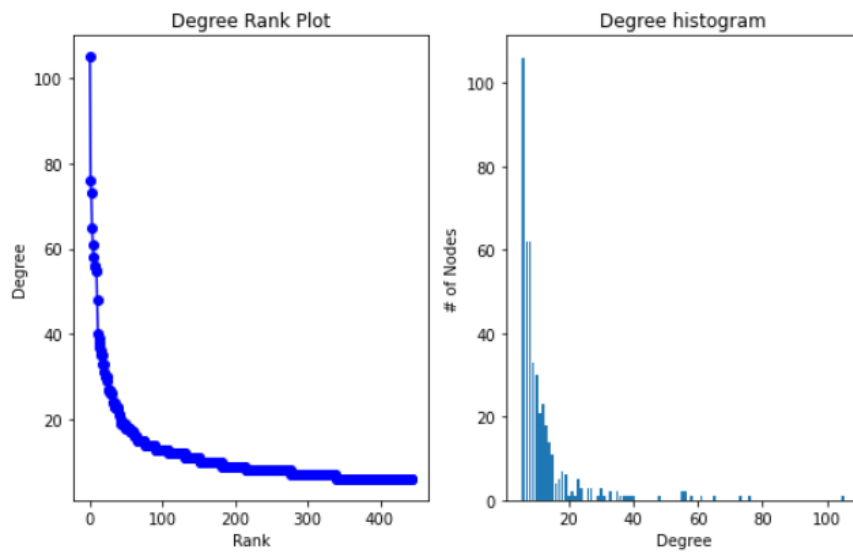
Datos de las 50 redes

11.952 es la media de las 50 redes





Distribución de grado de la Red2 N° 35

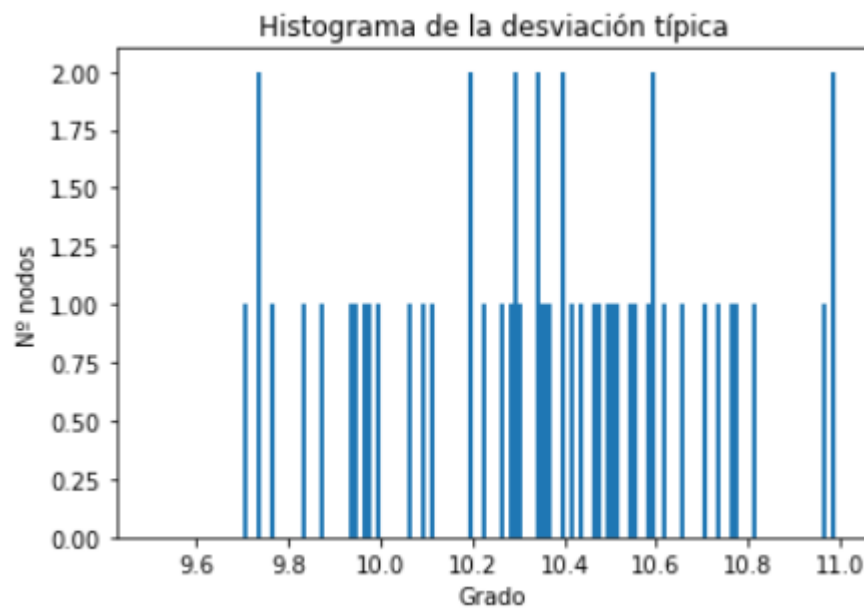


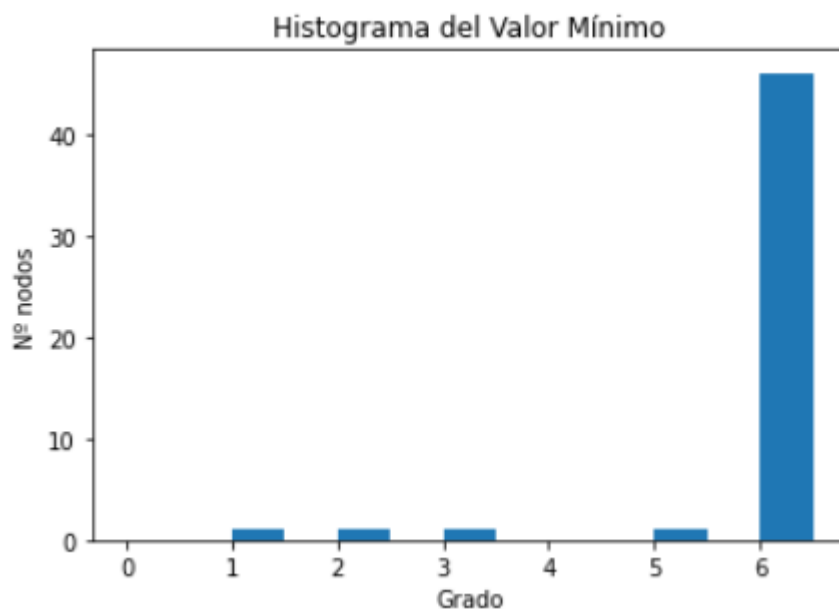
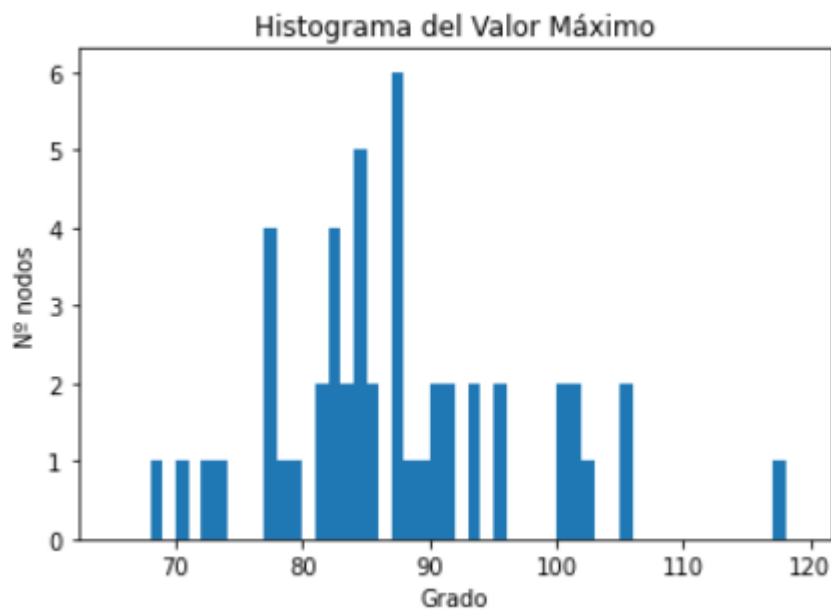
RED 2

Número de nodos: 445.0
Número de aristas: 2634.0
Camino mínimo medio: 445.0
Grado medio: 11.84
Radio medio de la red 3.0
Diámetro medio de la red 4.0
Número de componentes conexas: 39
Media del coeficiente de clustering 0.07922
Coeficiente de clustering global 0.06421

Datos de las 50 redes

11.838 media de las 50 redes





En este caso vemos como la primera red ha llegado a la zona crítica que rige el régimen conectado al haber solo un componente gigante. En cambio, la segunda red parece estar en régimen supercrítico. De la misma manera que en el modelo anterior el número de aristas es proporcional al número de nodos ya que en la primera red tenemos más del triple de enlaces y ambas tienen un grado medio que dista bastante del de la red de la P1 sin obtener una aproximación notable en el resto de las características. No obstante, en este caso ambas redes tienen casi el mismo grado medio y su distribución de grado es bastante similar lo que hace que sus coeficientes de clustering sean bastante más próximos aun así podemos ver tanto en sus valores como en las gráficas que hay una ligera diferencia entre ambas. Como en el último caso del modelo anterior la segunda red tiene más cantidad de nodos con grado superior a la media.

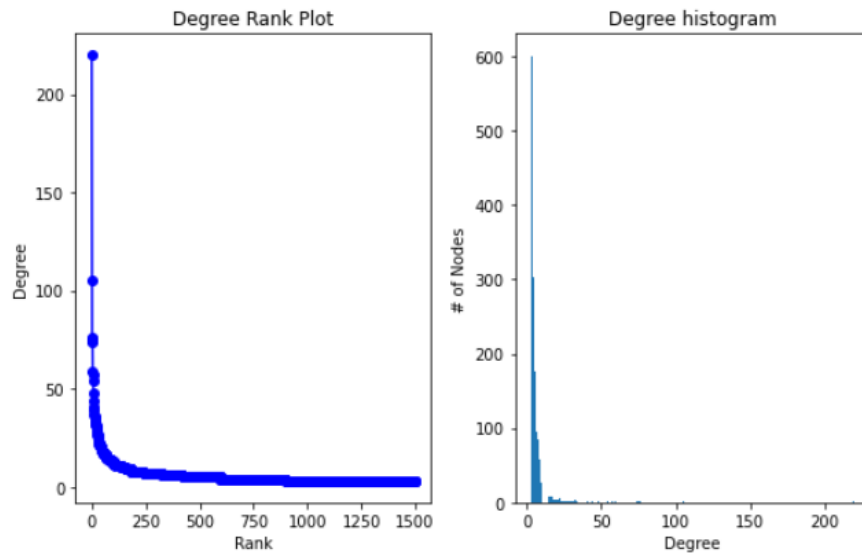
En conjunto, de la realización de la modelación con la primera parametrización podemos decir que se asemeja más a las redes reales sobretodo gracias al aspecto de conexión preferencial que intenta

incluir este modelo y al contrario que el de Erdos-Renyi este se adapta mejor a distintos valores de N.

Ahora probaremos para la mitad de m del apartado anterior a ver si obtenemos un grado medio más acorde al deseado y que resultados obtenemos para ambas redes.

Para $m=3$

Distribución de grado de la Red1 N° 1

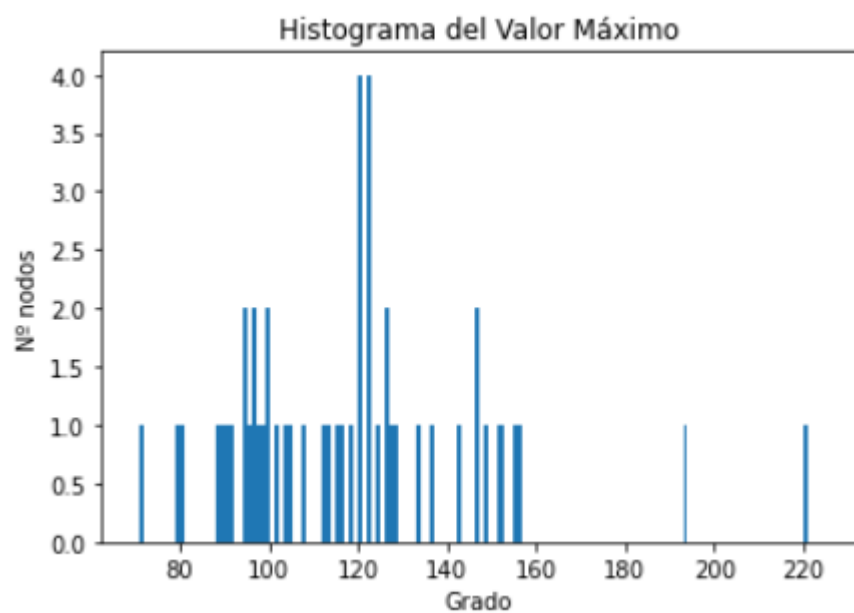
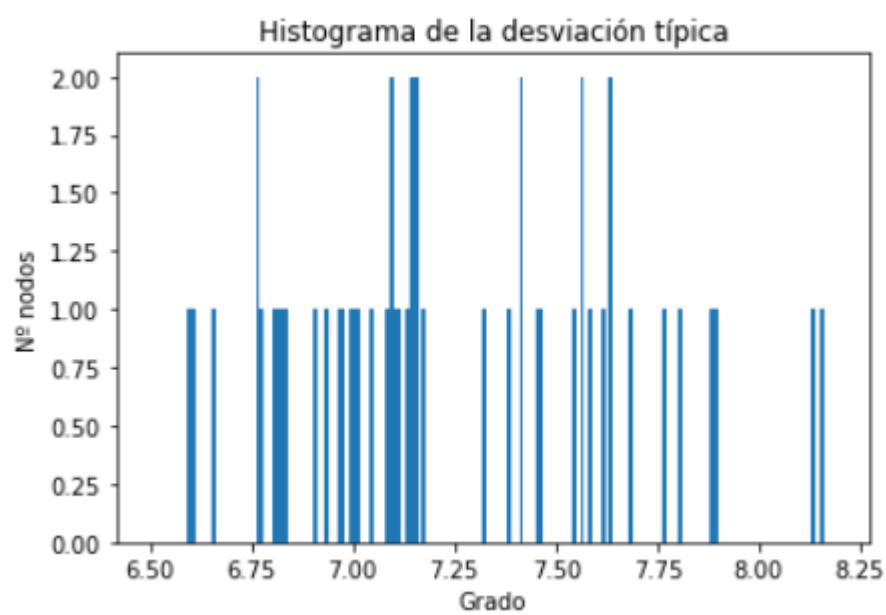


RED 1

Número de nodos: 1500.0
Número de aristas: 4491.0
Camino mínimo medio: 1500.0
Grado medio: 5.99
Radio medio de la red 4.0
Diámetro medio de la red 6.0
Número de componentes conexas: 1
Media del coeficiente de clustering 0.02257
Coeficiente de clustering global 0.01276

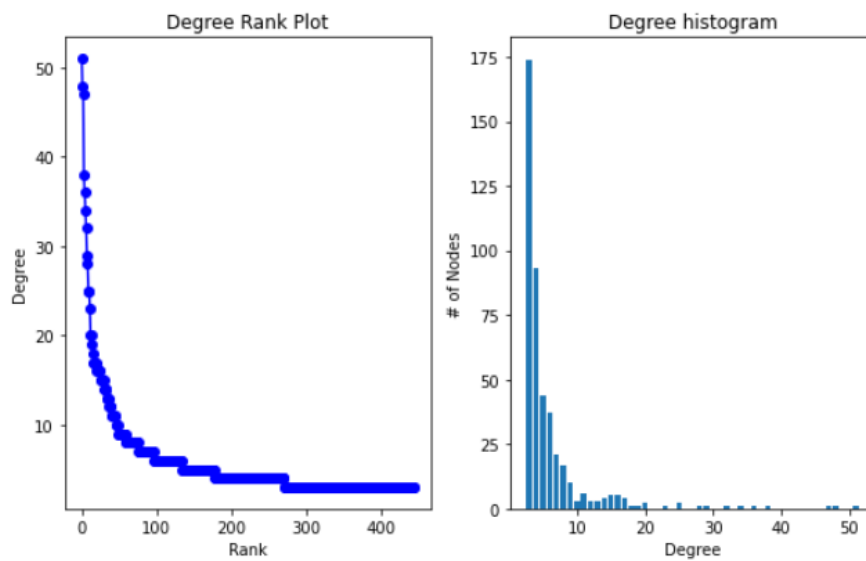
Datos de las 50 redes

5.988 es la media de las 50 redes





Distribución de grado de la Red2 N° 45

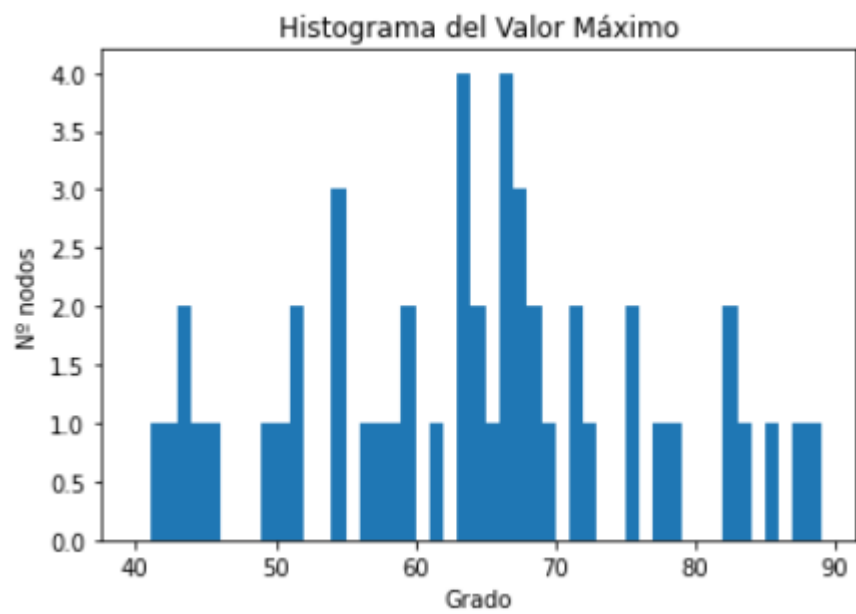
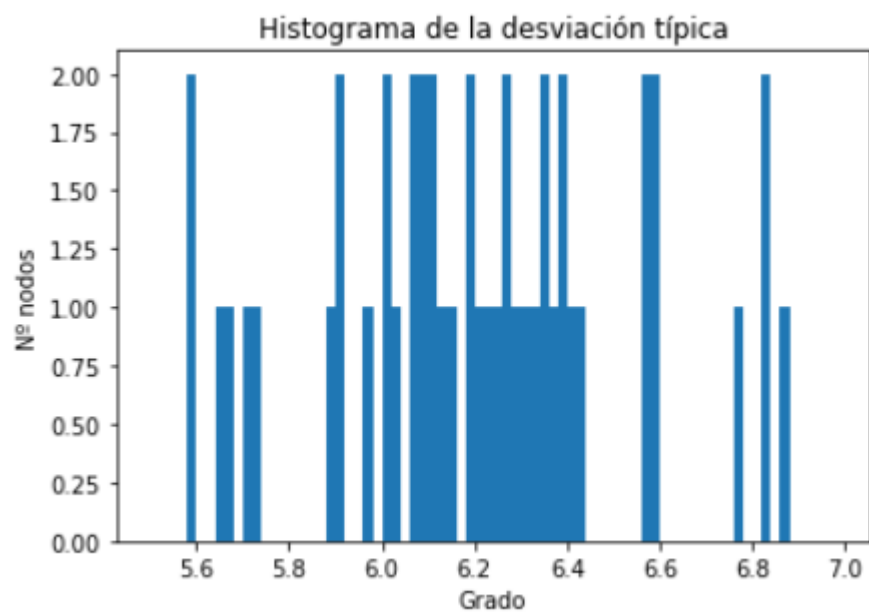


RED 2

Número de nodos: 445.0
 Número de aristas: 1326.0
 Camino mínimo medio: 445.0
 Grado medio: 5.96
 Radio medio de la red 3.3
 Diámetro medio de la red 5.3
 Número de componentes conexas: 1
 Media del coeficiente de clustering 0.05421
 Coeficiente de clustering global 0.03213

Datos de las 50 redes

5.959 es la media de las 50 redes





En esta realización podemos observar que el grado medio es casi idéntico al deseado obteniendo a su vez uno de los valores del coeficiente de agrupamiento más cercanos al deseado.

También podemos observar en las gráficas que, aunque algunos nodos lleguen casi a grado 200 o 50 la mayoría de ellos están por debajo de la media lo cual hace que no haya tantos nodos importantes en la red y siga sin poder alcanzar un coeficiente de agrupamiento próximo al que queremos.

El diámetro de la red es algo menor pero también es significativamente similar puesto que ambas redes tienen un diámetro relativamente pequeño según su tamaño. Sin embargo, con este valor de grado medio la red queda completamente conectada difiriendo así de la red original y sus 22 componentes distintas.

Estas redes se comportan de forma bastante parecida al modelo teórico ya que este nos dice que van a haber nodos preferentes que son los que más enlaces van a tener y los que tienen mayor probabilidad de que se formen nuevos enlaces como podemos observar en las gráficas de grados de los nodos.

Modelo Watts y Strogatz

El modelo Watts y Strogatz, en teoría de redes, se emplea para la construcción de algunas redes de mundo pequeño y con el objetivo de acercar los modelos aleatorios a las redes reales. Este modelo intenta satisfacer las propiedades de un mundo pequeño de la red y puede tener un coeficiente de clustering mucho más alto que los otros modelos. Genéricamente se trata de un modelo de generación de grafos aleatorios con distancias medias pequeñas y valores altos del coeficiente de agrupamiento (clustering coefficient).

La distribución de grado de los grafos generados mediante este algoritmo depende del valor de p . En el caso de $p=0$, en el que la red generada es una red regular, la distribución de grado es una Delta de Dirac centrada en $2k$. En el caso de $p=1$, en el que la red generada es una red aleatoria, la distribución de grado es una distribución de Poisson.

RED 1:

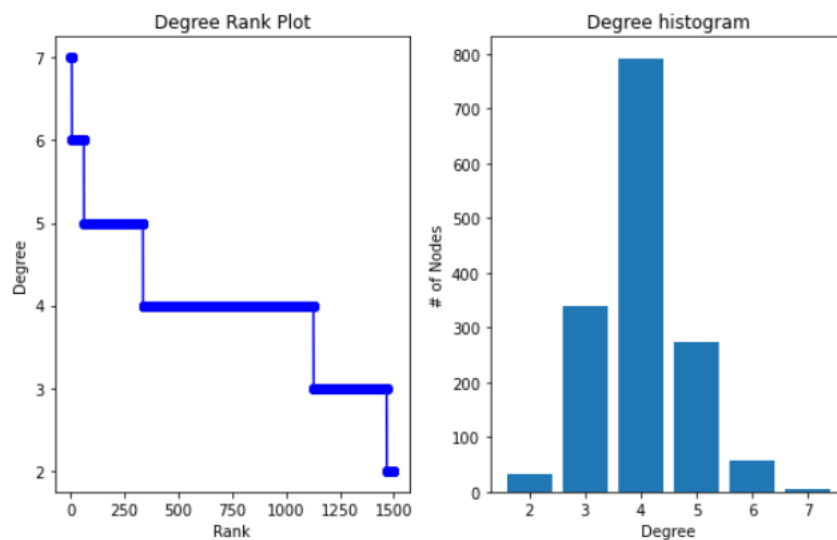
Num nodos = 1500

Vecinos adyacentes = 4 # Si es impar coge vecinos-1

probabilidad = 0.2

Distribución de grado de la Red1 N° 17

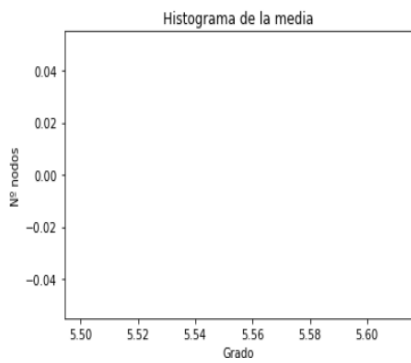
<Figure size 432x288 with 0 Axes>



RED 1

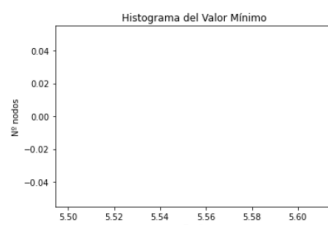
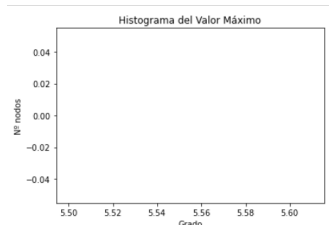
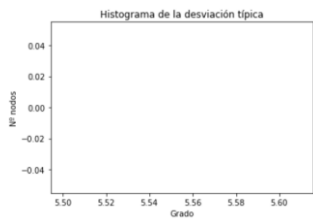
Número de nodos: 1500.0
 Número de aristas: 3000.0
 Camino mínimo medio: 1500.0
 Grado medio: 4.0
 Radio medio de la red 9.6
 Diámetro medio de la red 13.7
 Número de componentes conexas: 1
 Media del coeficiente de clustering 0.26534
 Coeficiente de clustering global 0.24057

Datos de las 50 redes

[illegible]

La media es la misma en todas las iteraciones por lo que todos los histogramas salen vacíos.

En este caso la media es 4.0 al igual que el número de nodos vecinos acordado para generar el grafo.



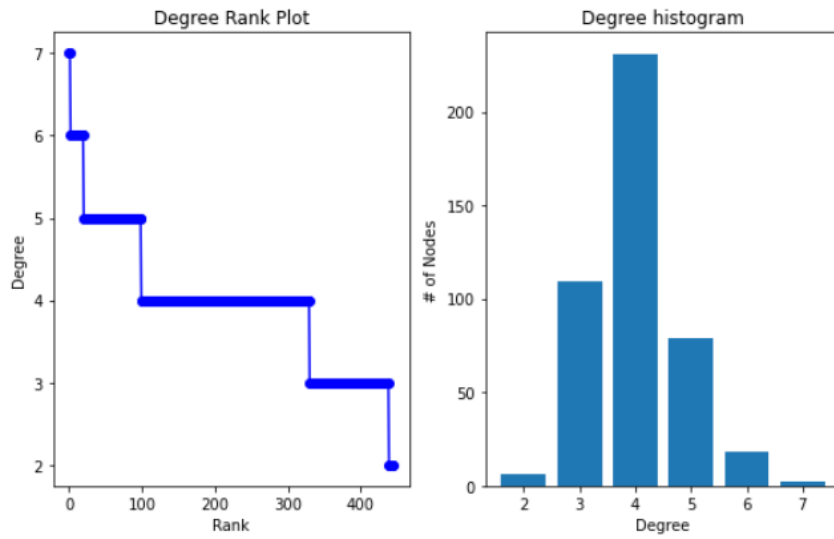
Red 2:

Num nodos = 445

Vecinos adyacentes = 4 # Si es impar coge vecinos-1

probabilidad = 0.2

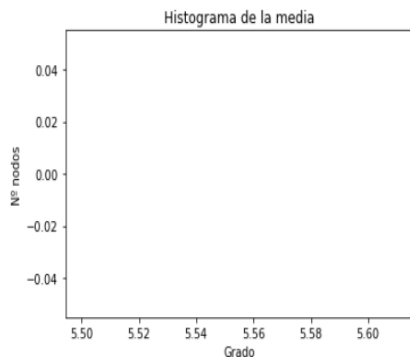
Distribución de grado de la Red2 N° 36



RED 2

Número de nodos: 445.0
 Número de aristas: 890.0
 Camino mínimo medio: 445.0
 Grado medio: 4.0
 Radio medio de la red 7.9
 Diámetro medio de la red 11.3
 Número de componentes conexas: 1
 Media del coeficiente de clustering 0.27162
 Coeficiente de clustering global 0.24686

Datos de las 50 redes

[illegible]

En este caso la media es 4.0 al igual que el número de nodos vecinos acordado para generar el grafo.

RED 1:

Num nodos = 1500

Vecinos adyacentes = 2 # Si es impar coge vecinos-1

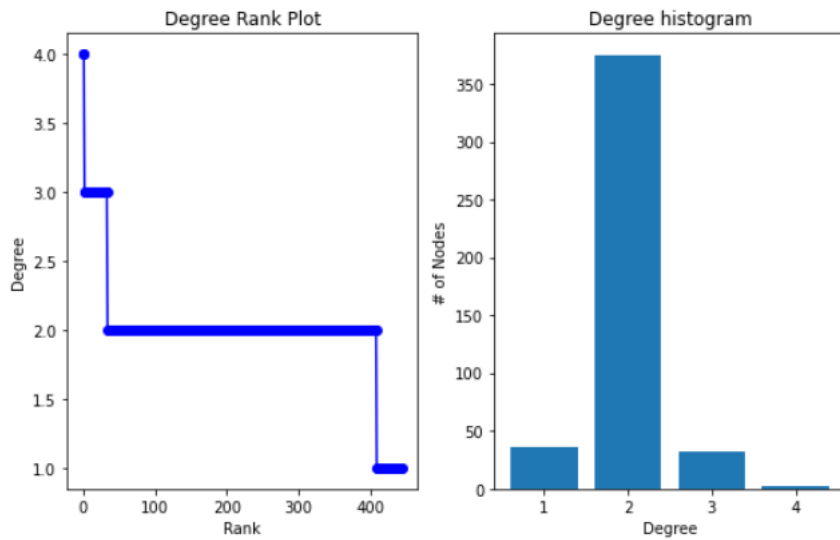
probabilidad = 0.1



Datos de las 50 redes

probabilidad = 0.1

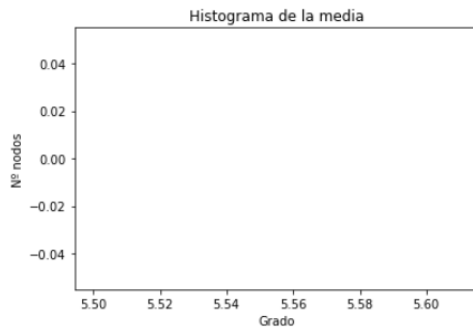
Distribución de grado de la Red2 N° 35



RED 2

Número de nodos: 445.0
 Número de aristas: 445.0
 Camino mínimo medio: 354.1
 Grado medio: 2.0
 Radio medio de la red 58.8
 Diámetro medio de la red 116.6
 Número de componentes conexas: 2
 Media del coeficiente de clustering 0.00063
 Coeficiente de clustering global 0.00074

Datos de las 50 redes

[illegible]

La media es la misma en todas las iteraciones por lo que todos los histogramas salen vacíos.

En este caso la media es 2.0 al igual que el número de nodos vecinos acordado para generar el grafo.. .

Como el modelo teórico pronosticaba podemos darnos cuenta del alto coeficiente de agrupamiento (clustering) del que consta estas redes en comparación con los grafos analizados anteriormente. Esto quiere decir que los nodos vecinos de cada nodo están conectados también, a cuanto mayor interconectividad entre nodos vecinos más cercano es el clustering de un nodo.

Esto ocurre por definición del propio modelo ya que este busca crear distancias medias pequeñas y coeficientes de agrupación altos.

Pongamos como ejemplo un nodo A conectado a 3 nodos (B, C, D); si B está conectado con C y D, y, a su vez C y D están conectados, podríamos decir que A tiene un coeficiente de agrupamiento de 1. Si en su lugar B estuviera conectado con C y D tendría $2/3$ de clustering y si por ejemplo sólo C y D estuvieran conectados A tendría un clustering de $1/3$.

Como podemos observar en los dos primeros casos tenemos que hay el doble de aristas que de nodos esto debido a que todos los nodos tienen grado 4 lo que quiere decir que hay el doble de aristas que de nodo ya que por cada arista hay siempre dos grados de los nodos conectados. Así que podríamos decir que Número total de aristas = $(\text{grado} * \text{nodo}) / 2 \rightarrow 4 * 1500 / 2 \rightarrow 2 * 1500 = 3000$. Esto se debe al número de vecinos adyacentes del que consta cada nodo que en las primeras dos redes son 4 por eso cada nodo consta de grado 4, en las otras dos redes que tienen el mismo número de nodos que las dos anteriormente analizadas consta de grado 2 y por ende de la mitad de las aristas 1500 en el caso de la red 1 y 445 en el caso de la red 2 ya que cada nodo está conectado únicamente con dos nodos. Aquí también entra en juego la probabilidad de recablear las aristas que en las dos primeras redes teníamos un 0,2 y en la segunda un 0,1.

Podemos notar una notable disminución del clustering entre los dos primeros y los dos segundos, esto debido principalmente a que el grado medio ha disminuido por lo que hay menos interconectividad en el grafo y por ende un menor coeficiente de agrupamiento. Sospecho también que una menor probabilidad de recableado hace que haya más componentes conexas ya que no se conectan todos los nodos del grafo entre sí y se crean pequeños circuitos dentro de este.

En general este modelo se ajusta bastante al modelo teórico, pero se aleja de la misma manera a nuestro grafo modelo de la práctica 1, por lo que no es un buen candidato.

Conclusión:

Después de crear diferentes redes con diferentes parámetros y modelizándolas de diferentes maneras, me he dado cuenta de cuan útil son estos tipos de generaciones aleatorias, pudiendo rápidamente hacernos una idea de cómo sería una red imaginaria. La idea más común de red aleatoria es que son grafos cuyas aristas están conectadas aleatoriamente con una determinada probabilidad. En el terreno analítico son muy útiles para experimentar los escenarios de una red desconectada a una red totalmente conectada, siendo importantes las fases intermedias que puede dar lugar a las famosas redes de mundo pequeño o redes complejas también llamadas redes libres de escala (Watts y Strogatz, 1998).

En nuestro caso al compararla a nuestra red de la P1 hemos tenido bastantes dificultades en encontrarles similitud ya que nuestra red de la práctica 1 era basada en un estudio sobre tejedores republicanos donde se habían tomado diferentes observaciones a lo largo de 10 meses de estudio. Esto era un problema muy real que al intentar abstraerlo a la generación de grafos aleatorios es complicado que se acerquen a los parámetros que buscamos, por ejemplo, nadie se ha acercado un ápice al número de componentes o al coeficiente de agrupamiento. Pero teniendo en cuenta las métricas obtenidas de cada generación hemos llegado a la conclusión que el “modelo” que más se asemeja a lo que buscábamos es “configuration model” y el modelo evolutivo de Barabasi-Albert con $m=3$ ya que estos sí que tienen algunos rasgos en común con nuestro grafo para nada despreciables como pueden ser el grado medio o el diámetro manteniendo un buen valor de coeficiente de agrupamiento en relación con los demás modelos.

Aunque por último me gustaría añadir que las redes generadas por los diferentes modelos distan mucho de la red modelo (PV1) esto debido a que es una red muy específica con atributos y valores difícilmente replicables para en la generación aleatorias de grafo ya que estas generaciones son muy “artificiales” y para nada orgánicas al contrario del estudio en el que se basa nuestra PV1. Por lo que era realmente complicado si quiera acercarnos a crear una red aleatoria de las mismas características.