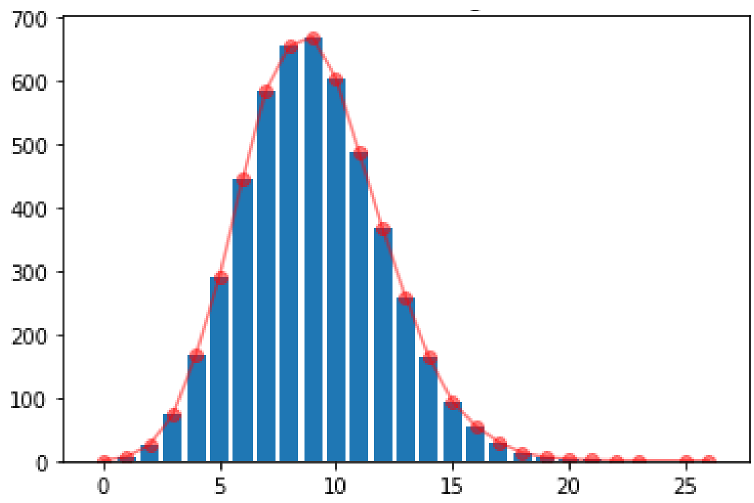
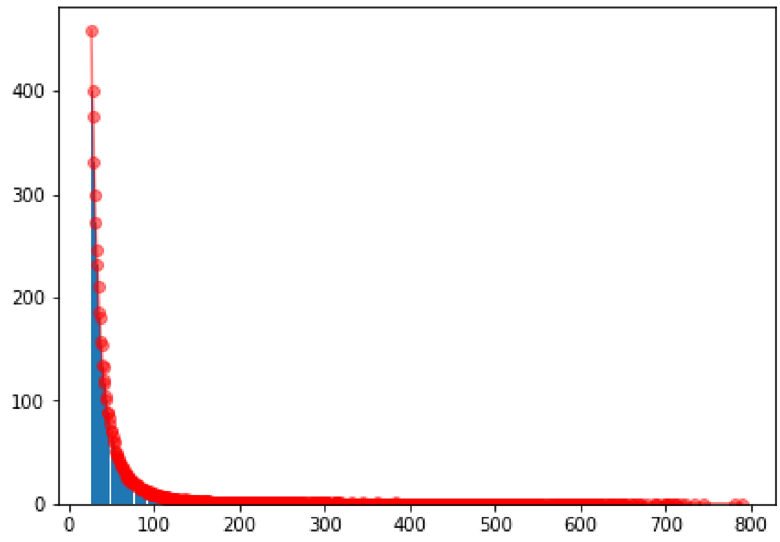
Práctica 2: Modelos de redes

Miguel Ferreras Chumillas / Pablo Daurell Marina
(Grupo 3)





1. Introducción

El objetivo de esta práctica es estudiar el comportamiento de dos modelos de redes: El modelo de Erdos-Renyi, para redes aleatorias, y el modelo de Barabasi-Albert, para redes de libre escala.

Se utilizará *networkx* para generar varias redes con cada modelo probando distintas configuraciones y se compararán los resultados de calcular las medidas principales de la red según el modelo que la haya generado con los resultados obtenidos con *networkx*.

Como networkx genera cada modelo con una cierta aleatoriedad, para calcular las características de cada red, generaremos 15 redes con la misma configuración y sacaremos los valores medios de cada medida. Las comparaciones del modelo teórico y la red real se realizarán en base a estas medias.

El código utilizado para generar las redes se puede ver en el jupyter notebook adjuntado a la práctica.

2. Modelo Erdos-Renyi

Este modelo se basa en la probabilidad (p) que tiene cada par de nodos de conectarse. Generaremos redes de 500 nodos y de 5000 nodos en cada una de las fases en las que puede encontrarse una red aleatoria, en función del aumento de p:

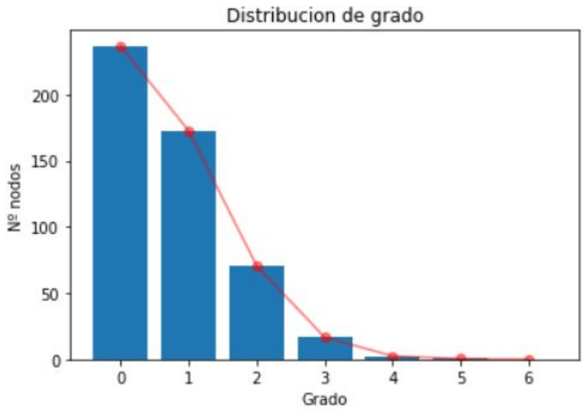
- Fase Subcrítica: $p < 1 / N$
- Fase Crítica: $p = 1 / N$
- Fase Supercrítica: $p > 1 / N$
- Fase Conectada: $p > \ln(N) / N$

Para cada fase se calcularán las siguientes medidas, calculadas como establece el modelo de Erdos-Renyi y comparándolas con los valores reales de la red:

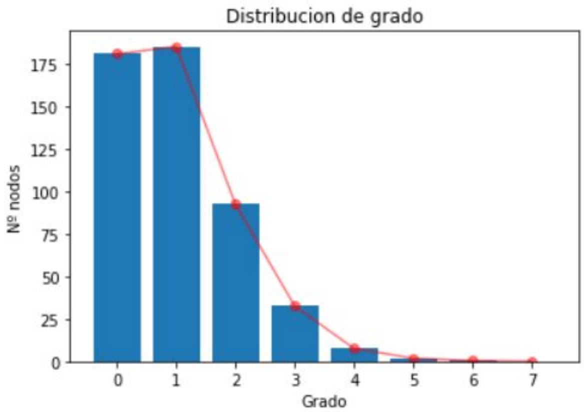
- Nodos: N
- Probabilidad: p
- Enlaces: $\langle L \rangle = p \cdot L_{max}$
- Grado medio: $\langle k \rangle = p \cdot (N - 1)$
- Coef. clustering: $\langle c \rangle = \langle k \rangle / N$
- Componentes conexas: Ncc

Redes de 500 nodos

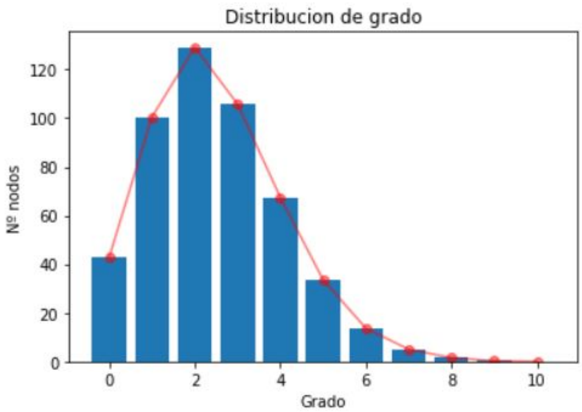
```
Red Aleatoria / Modelo Erdos-Renyi (N = 500, p = 0.0015)
---------------------------------------------------
Medidas TEORICAS
<k>:  0.7485
<L>:  187.125
<C>:  0.001497
-------------------------
Medidas REALES
<k>:  0.7538666666666667
<L>:  188.46666666666667
<C>:  0.0003111111111111111
Ncc:  312.0
```



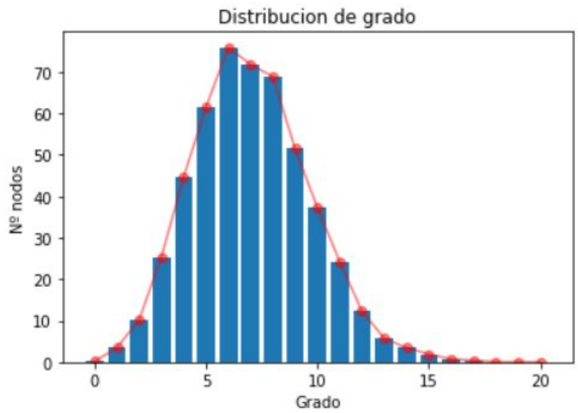

```
Red Aleatoria / Modelo Erdos-Renyi (N = 500, p = 0.002)
---------------------------------------------------
Medidas TEORICAS
<k>:  0.998
<L>:  249.5
<C>:  0.001996
-------------------------
Medidas REALES
<k>:  1.0192
<L>:  254.8
<C>:  0.0008533333333333333
Ncc:  246.0
```




```
Red Aleatoria / Modelo Erdos-Renyi (N = 500, p = 0.005)
---------------------------------------------------
Medidas TEORICAS
<k>:  2.495
<L>:  623.75
<C>:  0.004990000000000005
-------------------------
Medidas REALES
<k>:  2.4928000000000003
<L>:  623.2
<C>:  0.003648253968253968
Ncc:  49.0
```

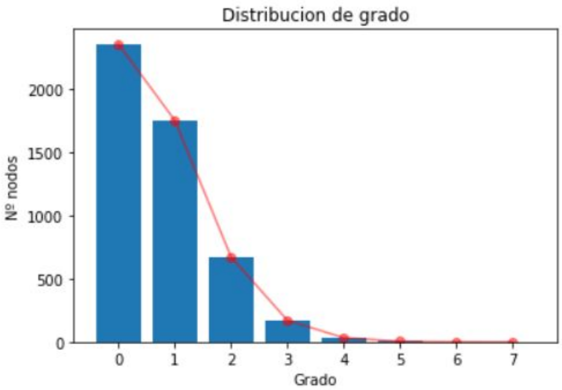



```
Red Aleatoria / Modelo Erdos-Renyi (N = 500, p = 0.014)
---------------------------------------------------
Medidas TEORICAS
<k>:  6.986
<L>:  1746.5
<C>:  0.013972
-------------------------
Medidas REALES
<k>:  7.042933333333334
<L>:  1760.7333333333333
<C>:  0.014483933392602113
Ncc:  1.0
```

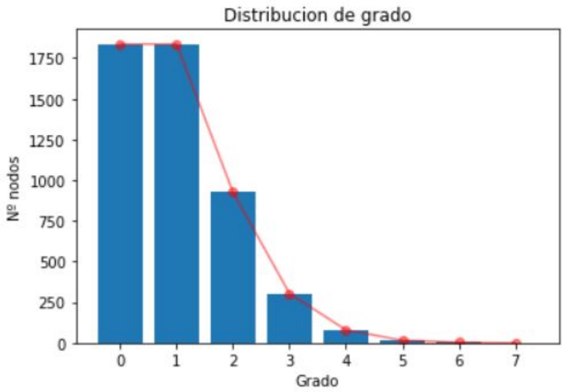


Redes de 5000 nodos

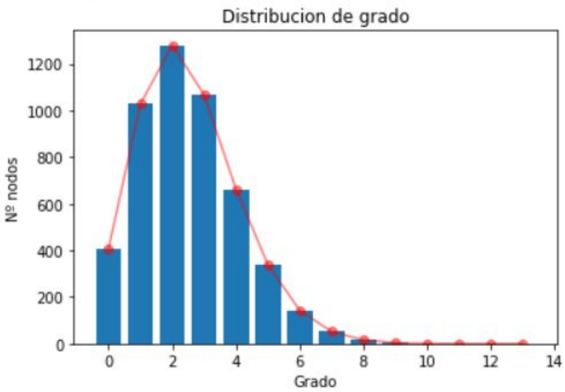
```
Red Aleatoria / Modelo Erdos-Renyi (N = 5000, p = 0.00015)
---------------------------------------------------
Medidas TEORICAS
<k>:  0.7498499999999999
<L>:  1874.6249999999998
<C>:  0.00014996999999999998
-------------------------
Medidas REALES
<k>:  0.7544533333333334
<L>:  1886.1333333333334
<C>:  0.0
Ncc:  3114.0
```



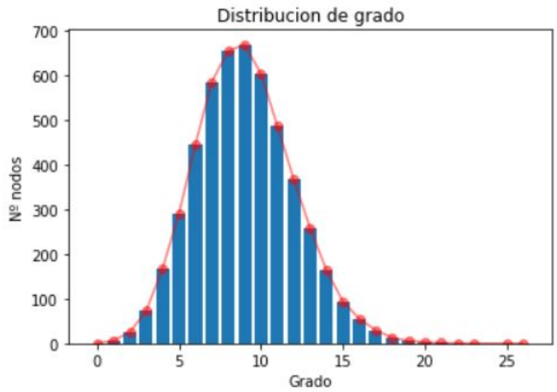

```
Red Aleatoria / Modelo Erdos-Renyi (N = 5000, p = 0.0002)
---------------------------------------------------
Medidas TEORICAS
<k>:  0.9998
<L>:  2499.5
<C>:  0.00019996
-------------------------
Medidas REALES
<k>:  1.0029066666666666
<L>:  2507.266666666667
<C>:  1.9999999999999998e-05
Ncc:  2494.0
```




```
Red Aleatoria / Modelo Erdos-Renyi (N = 5000, p = 0.0005)
---------------------------------------------------
Medidas TEORICAS
<k>:  2.4995
<L>:  6248.75
<C>:  0.0004999
-------------------------
Medidas REALES
<k>:  2.505333333333333
<L>:  6263.333333333333
<C>:  0.0004028465608465609
Ncc:  464.0
```




```
Red Aleatoria / Modelo Erdos-Renyi (N = 5000, p = 0.0018)
---------------------------------------------------
Medidas TEORICAS
<k>:  8.998199999999999
<L>:  22495.5
<C>:  0.0017996399999999997
-------------------------
Medidas REALES
<k>:  9.024053333333335
<L>:  22560.133333333335
<C>:  0.0017901061109562933
Ncc:  2.0
```



En las imágenes anteriores vemos las distintas redes generadas: subcrítica, crítica, supercrítica y conectada (de izquierda a derecha y de arriba a abajo), junto con sus medidas según el modelo teórico y sus medidas reales.

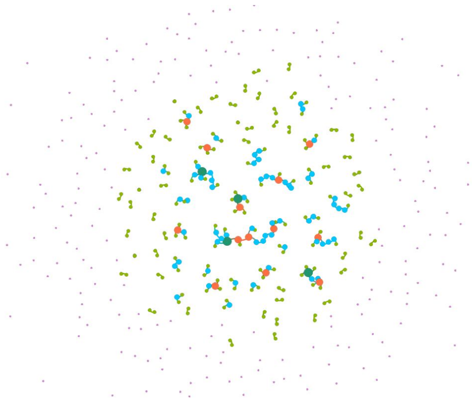
Vemos como, a medida que aumenta *p* (y por tanto aumenta el número de enlaces), la distribución de grados se asemeja cada vez más a una distribución de Poisson, tal y como se predice en el modelo teórico. La mayoría de nodos se concentran cerca del valor de grado medio.

Podemos ver que en todas las etapas excepto en la última existen una gran cantidad de componentes conexas y muchos nodos independientes. Es en la última etapa donde vemos que se genera una sola componente gigante con todos los nodos de la red.

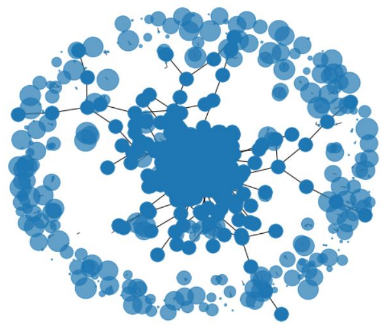
También vemos que a pesar de aumentar el número de nodos, no hay una gran variación en los datos y las medidas reales siguen siendo bastante cercanas al modelo teórico.

<u>Ejemplos del modelo en sus distintas etapas (N = 500)</u>

Subcrítica:

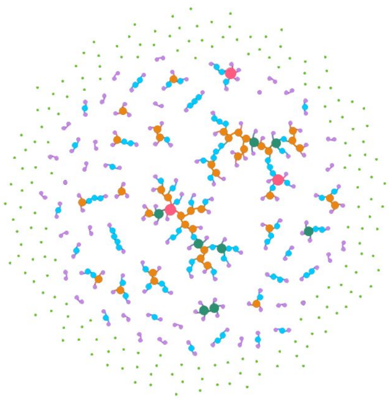


(Gephi)

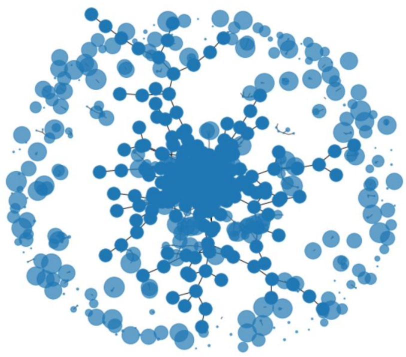


(Networkx)

Crítica:



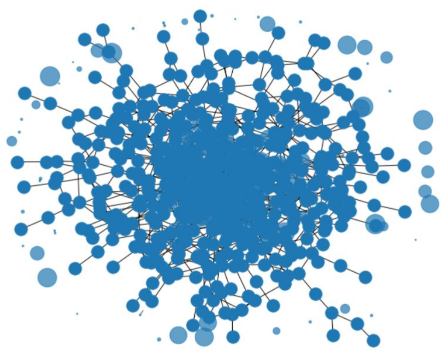
(Gephi)



(Networkx)

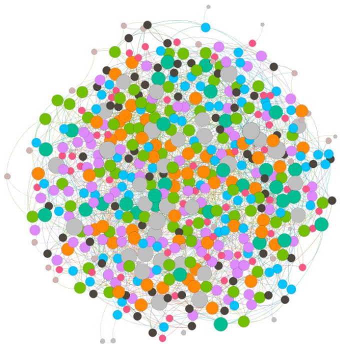
Supercrítica:

(Gephi)

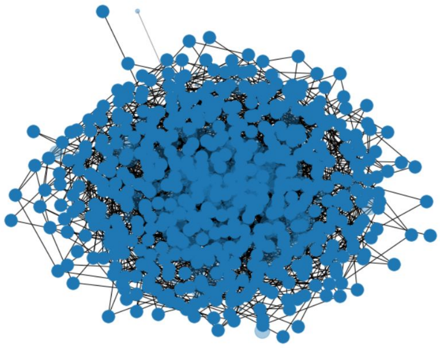


(Networkx)

Conectada:



(Gephi)



(Networkx)

En esta representación se puede ver de nuevo como en las distintas etapas va aumentando el tamaño de las componentes conexas y disminuyendo el número de nodos libres, a la par que como es lógico aumenta el grado de los nodos, sin embargo este es bastante regular, sin que haya nodos que destaquen especialmente. También es destacable la gran reducción en el número de componentes conexas que se aprecia entre la etapa de crítica y la supercrítica.

3. Modelo Barabási-Albert

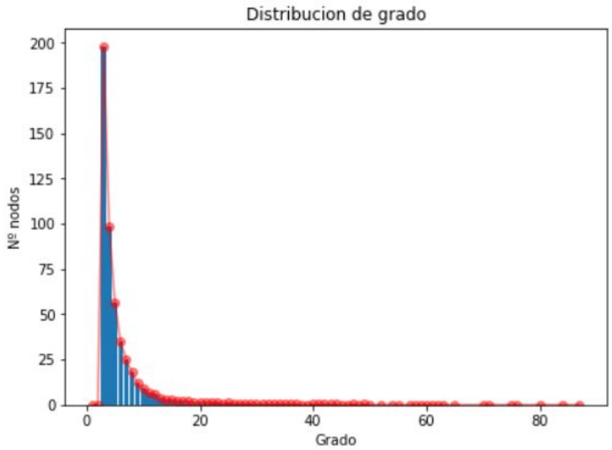
Para este modelo contamos con un valor 'm' que indicará con cuántos nodos se enlazará cada nuevo nodo que se añada a la red.
Generaremos redes de 500 nodos y de 5000 nodos, con valores para m de 3 y 4.

Para cada fase se calcularán las siguientes medidas, calculadas como establece el modelo de Barabasi-Albert y comparándolas con los valores reales de la red:

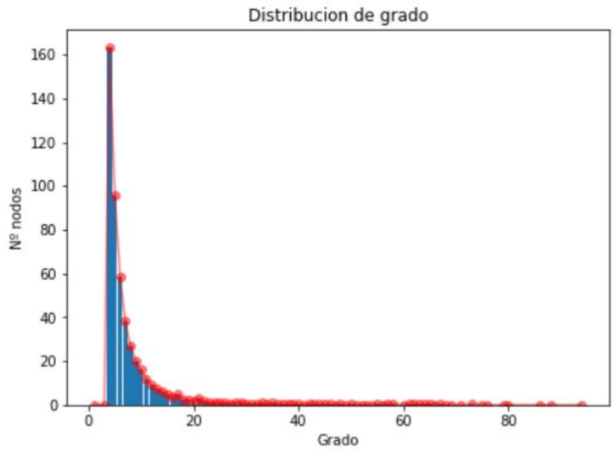
- Nodos: N
- Nodos iniciales: $m_0 = m$
- Nº conexiones de un nodo nuevo: $m <= m_0$
- Enlaces: $<L> = m_0 + m \cdot t$
- Grado medio: $<k> = 2 \cdot m$
- Coef. clustering: $<c> = \ln(N^2) / N$
- Distancia media: $<d> = \ln(N) / \ln(\ln(N))$
- Componentes conexas: Ncc

Redes de 500 nodos

```
Red Aleatoria / Modelo Barabasi-Albert (N = 500, m = 3)
-------------------------------------------------------
Medidas TEORICAS
<k>:  6
<L>:  1494
<C>:  0.07724270763394936
<d>:  3.401718228036341
-------------------------
Medidas REALES
<k>:  5.963999999999995
<L>:  1491.0
<C>:  0.05087629258490914
<d>:  3.2405974615898456
Ncc:  1.0
```

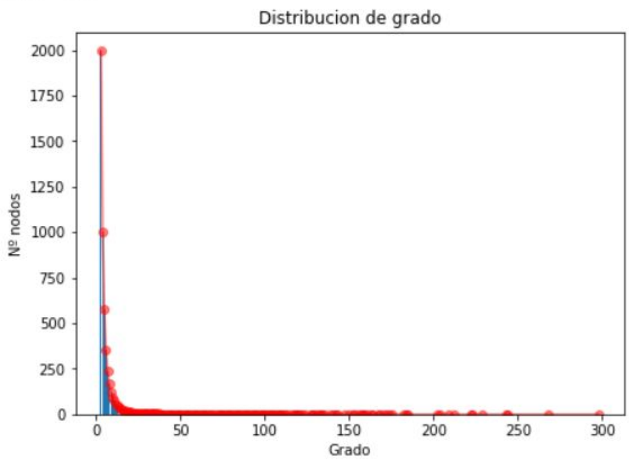



```
Red Aleatoria / Modelo Barabasi-Albert (N = 500, m = 4)
-------------------------------------------------------
Medidas TEORICAS
<k>:  8
<L>:  1988
<C>:  0.07724270763394936
<d>:  3.401718228036341
-------------------------
Medidas REALES
<k>:  7.9360000000000035
<L>:  1984.0
<C>:  0.05830046240522277
<d>:  2.94958663994656
Ncc:  1.0
```

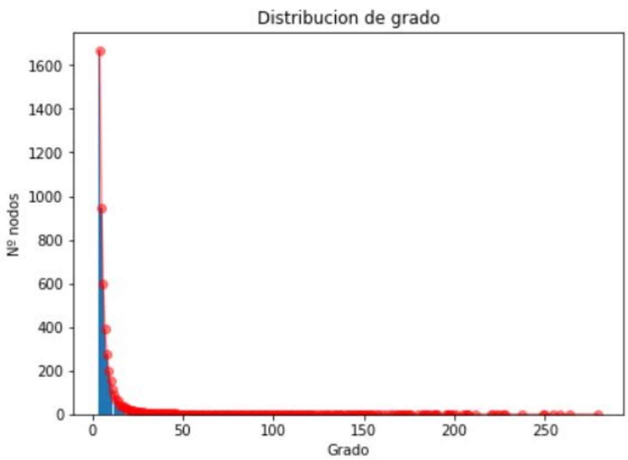


Redes de 5000 nodos

```
Red Aleatoria / Modelo Barabasi-Albert (N = 5000, m = 3)
-------------------------------------------------------
Medidas TEORICAS
<k>:  6
<L>:  14994
<C>:  0.014508515971981424
<d>:  3.976119453625683
-------------------------
Medidas REALES
<k>:  5.996399999999999
<L>:  14991.0
<C>:  0.008994752039619814
<d>:  4.055851986397278
Ncc:  1.0
```




```
Red Aleatoria / Modelo Barabasi-Albert (N = 5000, m = 4)
-------------------------------------------------------
Medidas TEORICAS
<k>:  8
<L>:  19988
<C>:  0.014508515971981424
<d>:  3.976119453625683
-------------------------
Medidas REALES
<k>:  7.9936
<L>:  19984.0
<C>:  0.010493291814187079
<d>:  3.6933536040541437
Ncc:  1.0
```



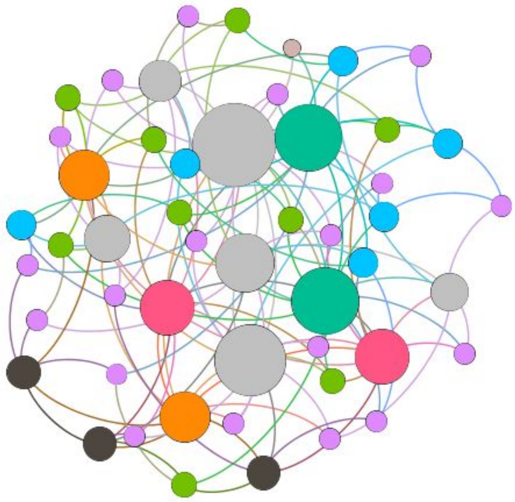
En las imágenes anteriores vemos las redes de Barabasi-Albert generadas con m = 3 y m = 4 junto con sus medidas según el modelo teórico y sus medidas reales.

Tal y como predice el modelo, la distribución de grados sigue una ley potencial donde hay pocos nodos con grado alto y la mayoría de nodos se concentran en valores de grado bajos. Esto muestra la característica de este modelo por la cual los nodos tienden a unirse a los nodos con mayor peso *(preferential attachment)*.

En este caso vemos que siempre tenemos una componente conexa ya que el modelo se basa en añadir nuevos nodos a la red ya existente.

Ejemplos del modelo con distinta m (N = 50)

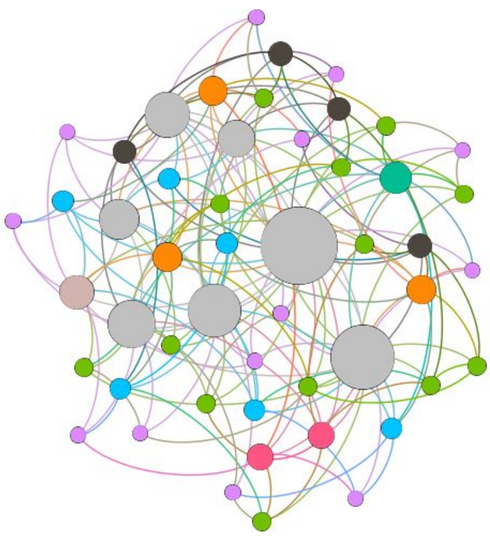
m = 3



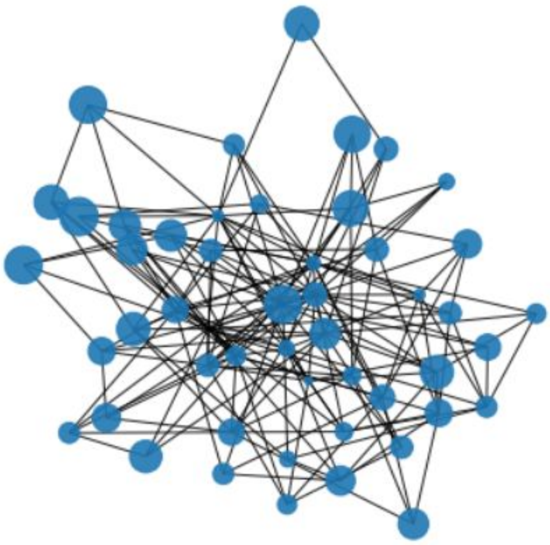
(Gephi)

(Networkx)

m = 4



(Gephi)



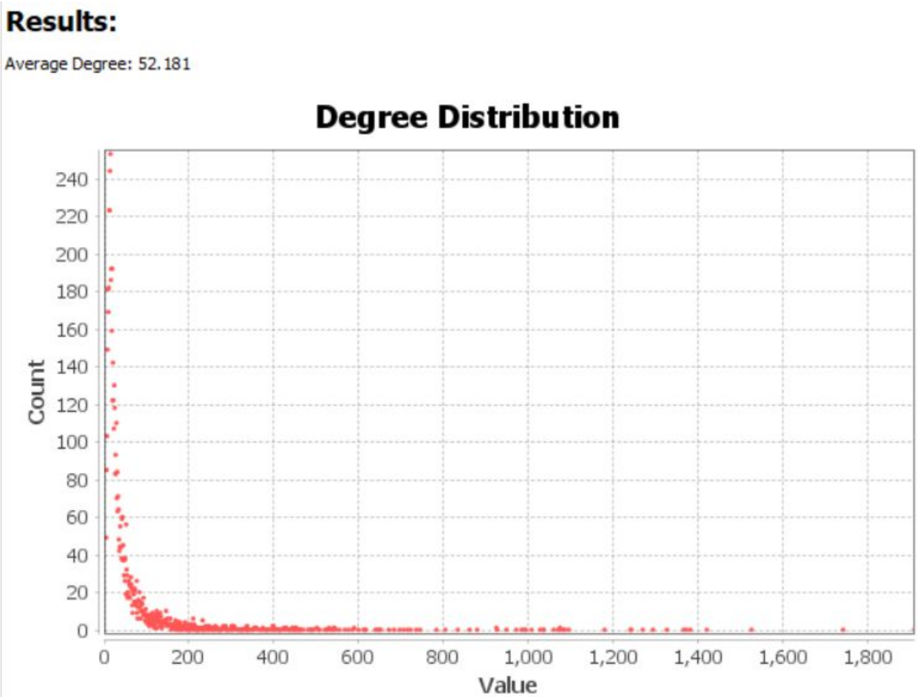
(Networkx)

Podemos ver, a diferencia de en el anterior modelo que siempre hay una componente conexa, y además una diferencia entre el grado de los nodos muy superior a la que podíamos apreciar en el Erdos-Renyi

4. Comparativa con la red estudiada en la Práctica 1

Medidas de la red de la práctica 1 (Red de héroes de Marvel):

- **Nodos**: 6408
- **Aristas**: 167163
- **Grado medio**: 52.181
- **Coef. global de clustering**: 0.7744001182184556
- **Distancia media**: 2.638390763311266



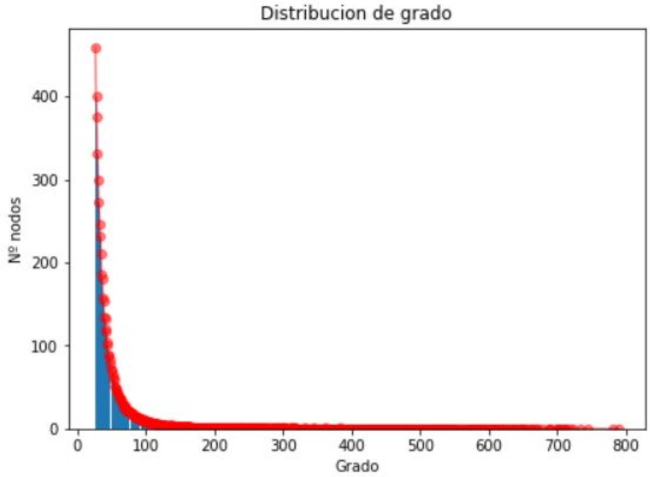
Distribución de grado de la red de la práctica 1 (Gephi)

Podemos ver en la distribución de grado que esta red es bastante similar al modelo de Barabási-Albert. Alejándose bastante de la distribución de Poisson a la que tienden las redes de Erdos-Renyi, se acerca mucho a una distribución exponencial con gran cantidad de nodos con grado bajo y muy pocos nodos con grados altos.

Esto tiene bastante sentido ya que en la red de la práctica 1 (héroes de Marvel) existen varios nodos principales que corresponden con los personajes principales de Marvel y es muy probable que el resto de nodos estén conectados a ellos y que no estén conectados con otros nodos menos relevantes de la red, lo cual encaja con la premisa del enlace preferencial del modelo de Barabasi-Albert.

Comparando con las medidas calculadas de las redes de Barabási (punto 3) podemos ver que coinciden en su gráfica de distribución de grado, pero se quedan cortas en el resto de estadísticas. Para encontrar una red más similar, creamos un conjunto de redes con los mismos nodos que la red de la práctica 1 y distintas m. Con eso hallamos que con m = 26 se crean redes mucho más parecidas.

```
Red Aleatoria / Modelo Barabasi-Albert (N = 6408, m = 26)
-------------------------------------------------------
Medidas TEORICAS
<k>:  52
<L>:  165958
<C>:  0.011989782727723988
<d>:  4.037819391379427
-------------------------
Medidas REALES
<k>:  51.78901373283396
<L>:  165932.0
<C>:  0.030172176447186217
<d>:  2.5423606398042717
Ncc:  1.0
```



Medidas de las redes BA generadas con m = 26 (networkx)

Podemos ver que las medidas generadas son bastante similares, acercándose bastante bien al valor de grado medio y el número de aristas. Sin embargo, el modelo no genera nodos que despunten tanto como lo hacían en la red anterior, con una diferencia de grado máximo de 1000. Y el clustering es tremendamente menor al no seguir la estructura de comunidades que tanto caracterizaba a la red de la anterior práctica.