Práctica 2. Implementación de modelos de redes

Grupo 5 Marlon Campoverde, Rebeca Herranz Guillermo Ovejero, Carlos Sánchez Septiembre de 2020



Índice

		Página
1.	Introducción	1
2.	Desarrollo de modelos	1
3.	Barabasi-Albert Model	2
4.	Erdos-Renyi Model	27
5.	Comparación de modelos	36

1. Introducción

Este proyecto se ha focalizado en el desarrollo de un programa en Python que sea posible con él generar redes aleatorias y redes libre de escala.

Este documento se ha realizado mediante el estudio de dos modelos que son el de Erdos-Renyi y Barabasi-Albert.

El modelo Erdos-Renyi es un modelo de grafos en el cual según un número de nodos n, y con probabilidad p, cada nodo tiene una probabilidad p de conectarse con otro nodo.

En cambio el modelo de Barabasi-Albert no se usa una probabilidad igual para todos los nodos. Si no que se generan m nodos iniciales conectados entre si y en cada paso de simulación se añade un nodo el cual se conecta a m nodos mas siguiendo el *Preferential Attachment* que se basa en esta fórmula.

$$\Pi(k_i) = \frac{k_i}{\sum\limits_{j} k_j}$$

Una vez se hayan desarrollado los grafos de manera aleatoria se analizarán sus características y se visualizarán con la librería de análisis de grafos para python, *networkx*.

Después se compararán los modelos con la red de películas de la práctica 1 para analizar cual de los dos modelos es más semejante.

2. Desarrollo de modelos

Los modelos han sido desarrollados en un fichero llamado modelos.py donde se han programado los algoritmos para ambos modelos y se han realizado y guardado mediante dos formas, el notebook de jupyter donde luego se han analizado y mediante un CLI que genera el modelo que se quiere según los parámetros seleccionados, se han generado 100 grafos en total y se han seleccionado los que se han creído mas útiles para el estudio de ellos.

3. Barabasi-Albert Model

En el modelo de Barabási-Albert haremos lo mismo que con el anterior, ahora con dos valores de nodos diferentes, 500 y 5000, también hay que especificar el valor de m_0 , en la práctica hemos usado los valores 3 y 4, además de esto, hemos ido guardando en diferentes csv los pasos de simulación de este modelo, se han cogido solo unos pocos para realizar las visualizaciones y analizarlos, en el repositorio de están el resto de archivos con los restantes pasos de simulación. Hemos realizado la simulación del modelo BA^1 , de A formas diferentes:

- $n = 500, m_0 = 3$
- $n = 500, m_0 = 4$
- $n = 5000, m_0 = 3$
- $n = 5000, m_0 = 4$

Los pasos de simulación elegidos hasta llegar al 25, a partir de ahí las visualizaciones se hacían difíciles analizar, además hemos añadido un ranking de tamaños para los nodos y poder visualizar el *Preferential Attachment*.

En las siguientes figuras observamos los diferentes modelos generados, con diferentes valores para los pasos, desde 5 hasta 25, para cada uno de los modelos generados.

Aquí vemos unas metricas simples para comprobar las diferencias entre un modelo y otro.

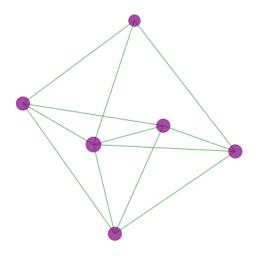
¹Modelo de libre escala de Barabasi Albert

 $../graphs/barabasi-albert \verb|\steps|ba_n_5000_m_3_steps_005.csv|$

Name:

Type: Graph

Number of nodes: 6 Number of edges: 12 Average degree: 4.0000

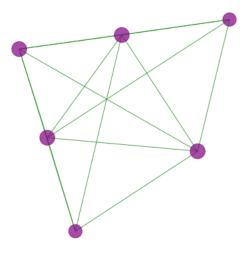


 $../graphs/barabasi-albert \verb|\steps|ba_n_5000_m_4_steps_005.csv|$

Name:

Type: Graph

Number of nodes: 6 Number of edges: 14 Average degree: 4.6667

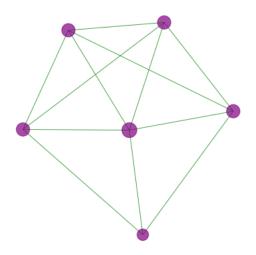


 $../graphs/barabasi-albert\steps\ba_n_500_m_3_steps_005.csv$

Name:

Type: Graph

Number of nodes: 6 Number of edges: 12 Average degree: 4.0000



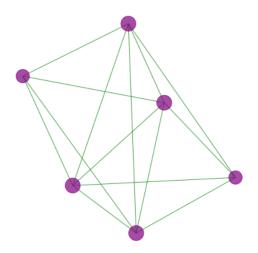
 $../graphs/barabasi-albert\steps\ba_n_500_m_4_steps_005.csv$

Name:

Type: Graph

Number of nodes: 6 Number of edges: 14 Average degree:

4.6667

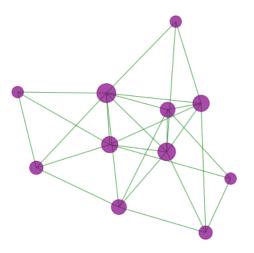


 $../graphs/barabasi-albert \verb|\steps|ba_n_5000_m_3_steps_010.csv|$

Name:

Type: Graph

Number of nodes: 11 Number of edges: 27 Average degree: 4.9091



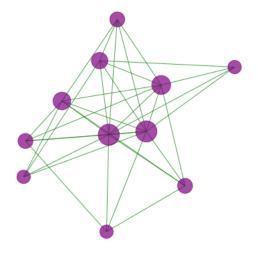
 $../graphs/barabasi-albert \verb|\steps|ba_n_5000_m_4_steps_010.csv|$

Name:

Type: Graph

Number of nodes: 11 Number of edges: 34

Average degree: 6.1818

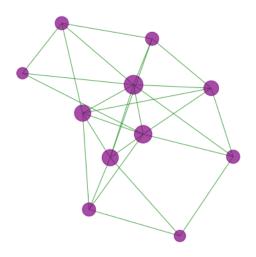


 $../graphs/barabasi-albert\steps\ba_n_500_m_3_steps_010.csv$

Name:

Type: Graph

Number of nodes: 11 Number of edges: 27 Average degree: 4.9091



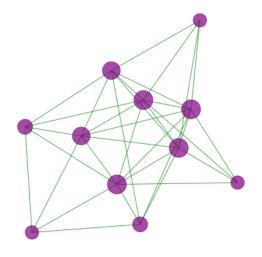
 $../graphs/barabasi-albert\steps\ba_n_500_m_4_steps_010.csv$

Name:

Type: Graph

Number of nodes: 11
Number of edges: 34

Average degree: 6.1818



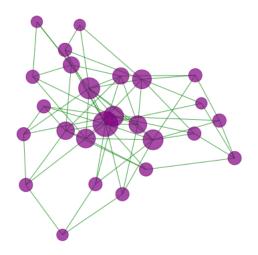
 $../graphs/barabasi-albert \verb|\steps|ba_n_5000_m_3_steps_025.csv|$

Name:

Type: Graph

Number of nodes: 26 Number of edges: 72

Average degree: 5.5385



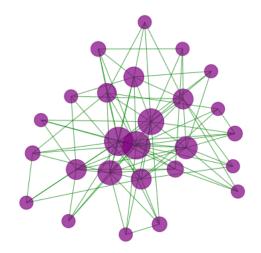
 $../ \texttt{graphs/barabasi-albert\steps\ba_n_5000_m_4_steps_025.csv}$

Name:

Type: Graph

Number of nodes: 26 Number of edges: 94

Average degree: 7.2308



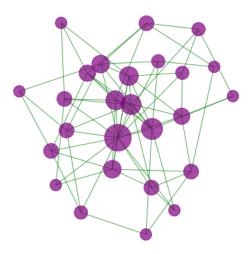
 $../graphs/barabasi-albert\steps\ba_n_500_m_3_steps_025.csv$

Name:

Type: Graph

Number of nodes: 26 Number of edges: 72

Average degree: 5.5385



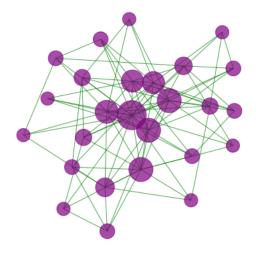
 $../graphs/barabasi-albert\steps\ba_n_500_m_4_steps_025.csv$

Name:

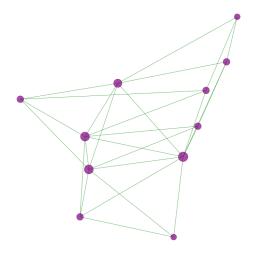
Type: Graph

Number of nodes: 26 Number of edges: 94

Average degree: 7.2308



Vemos diferencias entre los grafos con diferentes valores de m, cuando m es menor tiene menos aristas y por lo tanto menos grado medio



Degree Distribution

5
4
4
5
Degree

Degree Distribution

Figura 2: BA Degree Distribution m = 3 step = 10

Figura 1: Barabasi Albert m = 3 step = 10

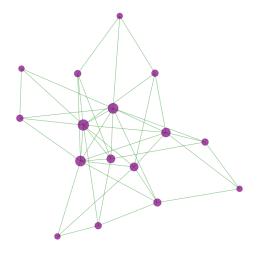
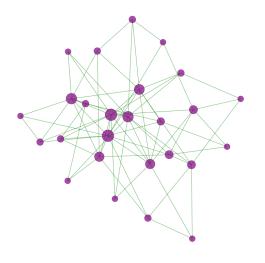


Figura 4: Barabasi Albert m = 3 step = 20

Figura 3: Barabasi Albert m = 3 step = 15

Ya con 20 nodos con m=3 podemos observar como claramente unos pocos nodos centrales son los que acumulan mayor grado, mientras que la periferia tiene solamente un grado no mayor a 4. Como se ve en la Figura 2. son 3 nodos ($los\ hubs$) los que tienen grado 6 mientras que los 7 restantes tienen una media de grado de 3.4

Como se observa en la distribución del grado, ya podemos observar como se empieza a distribuir en forma de *Ley de Potencias*, vemos que sigue habiendo entre 2-4 nodos con el mayor grado de la red mientras que la mayoría se quedan en el grado 3 o 4, dependiendo de el valor de *m*, debido a que la conexión preferencial hace que los primeros nodos que llegan antes a la red y que mayor grado tienen son a los que más se quieren unir otros nodos.

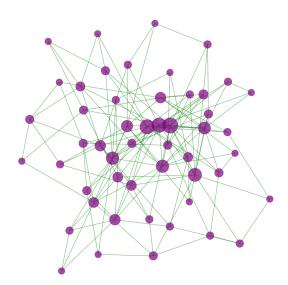


Degree Distribution

8
7
6
5
4
3
2
1
0
Degree Distribution

Figura 6: BA Degree Distribution m = 3 step = 25

Figura 5: Barabasi Albert m = 3 step = 25



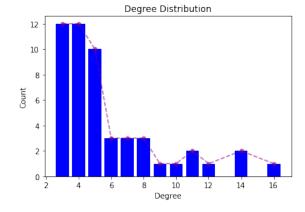


Figura 8: BA Degree Distribution m = 3 step = 50

Figura 7: Barabasi Albert m = 3 step = 50

Por último comparamos las medidas del modelo teórico con los resultados actuales obtenidos del computo de diferentes métricas, como pueden ser, numero de enlaces, grado medio, densidad, longitud media y el coeficiente de clustering.

Con estas formulas se realizan los cálculos para el modelo teórico, para el modelo generado usamos *networkx* para que nos calcule estas propiedades.

- Nodos de la red: *N*
- m0 = m + 1
- Numero de enlaces: L = (m0 * (m0 1)/2) + m * t
- Grado medio: $\langle k \rangle = 2 * m$
- Densidad: $D = \frac{L}{L_{max}}$
- Longitud media: $\langle d \rangle = \frac{\ln N}{(\ln (\ln N))}$
- Clustering: $C = \frac{(\ln N)^2}{N}$

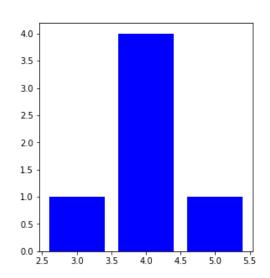
A continuación se mostrarán los datos más representativos de las redes que se quieren estudiar y comparar con el fin de llegar a las conclusiones correspondientes. Primero se mostraran los valores teóricos para dicha red, y a continuación los valores resultantes del análisis, junto a una distribución del grado y la representación del grafo.

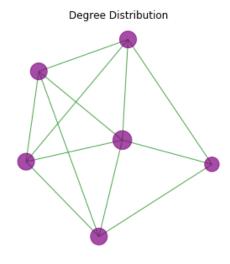
ba_n_500_m_3_steps_005.csv -----Teórico-----

N: 6 m: 3 <L>: 12.0 <k>: 6 D: 0.8

<d>: 3.072300009669941 C: 0.5350669992614002

N: 6 m: 3 <L>: 12 <k>: 4.0 D: 0.8 <d>: 1.2



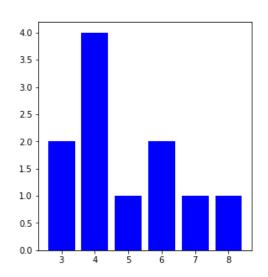


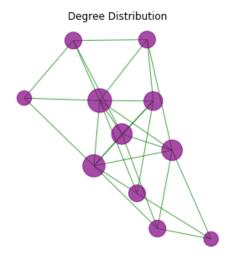
ba_n_500_m_3_steps_010.csv -----Teórico-----

N: 11 m: 3 <L>: 27.0 <k>: 6

D: 0.490909090909090909
<d>: 2.741732104405601
C: 0.5227183399371612

N: 11 m: 3 <L>: 27



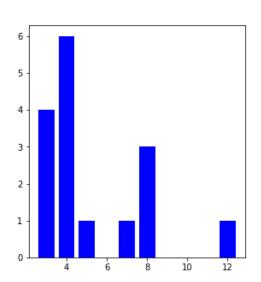


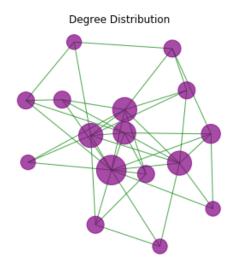
ba_n_500_m_3_steps_015.csv -----Teórico-----

N: 16 m: 3 <L>: 42.0 <k>: 6 D: 0.35

<d>: 2.7188068070511737 C: 0.4804530139182014

N: 16 m: 3 <L>: 42 <k>: 5.25 D: 0.35 <d>: 1.7 C: 0.328125





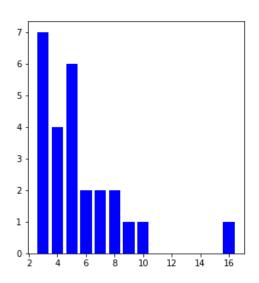
ba_n_500_m_3_steps_025.csv -----Teórico-----

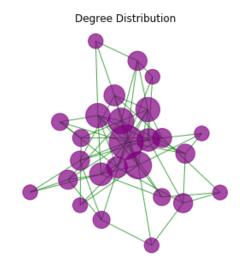
N: 26 m: 3 <L>: 72.0 <k>: 6

D: 0.22153846153846155 <d>: 2.758426498407852 C: 0.40827665581029104

N: 26 m: 3 <L>: 72

<k>: 5.538461538461538
D: 0.22153846153846155
<d>: 1.9692307692307693
C: 0.21301775147928995





ba_n_500_m_3_steps_050.csv -----Teórico-----

N: 51 m: 3

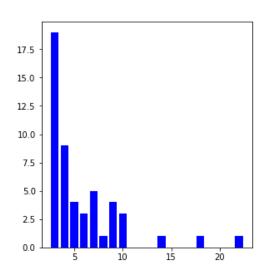
<L>: 147.0 <k>: 6

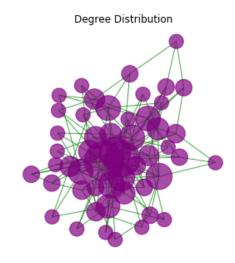
D: 0.11529411764705882 <d>: 2.871824234973293 C: 0.3031226040421185

N: 51 m: 3 <L>: 147

<k>: 5.764705882352941 D: 0.11529411764705882

<a><a>: 2.291764705882353<a>: 0.11303344867358708





ba_n_500_m_3_steps_075.csv -----Teórico-----

N: 76 m: 3

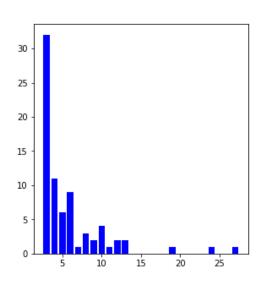
<L>: 222.0 <k>: 6

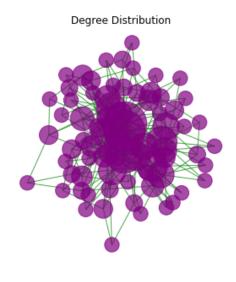
D: 0.07789473684210527 <d>: 2.9546457954813907 C: 0.2467796219035211

N: 76 m: 3 <L>: 222

<k>: 5.842105263157895
D: 0.07789473684210527

d: 0.07789473884210527
<d>: 2.460701754385965
C: 0.07686980609418283





ba_n_500_m_3_steps_100.csv -----Teórico-----

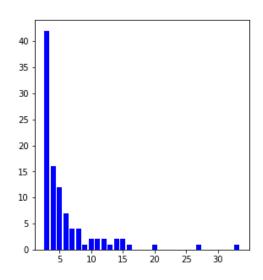
N: 101 m: 3

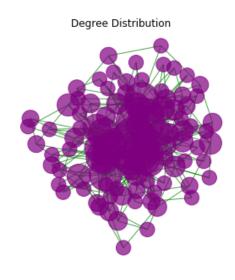
<L>: 297.0 <k>: 6

D: 0.0588118811881 <d>: 3.0177243826070073 C: 0.21088452856405085

N: 101 m: 3 <L>: 297

<a><a>k>: 5.881188118811881
D: 0.05881188118811881
<a><a>: 2.5752475247524753
C: 0.0582295853347711





ba_n_500_m_3_steps_125.csv -----Teórico-----

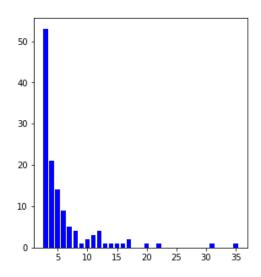
N: 126 m: 3

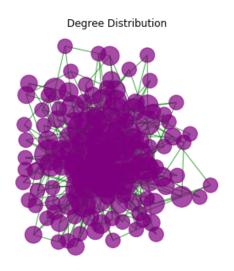
<L>: 372.0 <k>: 6

D: 0.04723809523809524 <d>: 3.0684220993419467 C: 0.1856319260595732

N: 126 m: 3 <L>: 372

C: 0.04686318972033258





ba_n_500_m_3_steps_150.csv -----Teórico-----

N: 151 m: 3 <I.>: 44

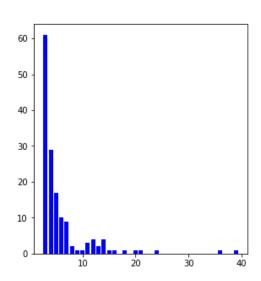
<L>: 447.0 <k>: 6

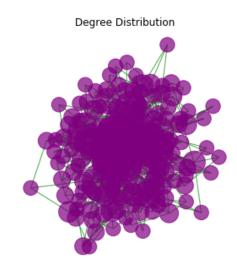
D: 0.03947019867549669 <d>: 3.1107430153202995 C: 0.16670925139675227

N: 151 m: 3 <L>: 447

<k>: 5.920529801324503
D: 0.03947019867549669
<d>: 2.735187637969095

C: 0.0392088066312881





$ba_3_500.csv$

-----Teórico-----

N: 500 m: 3

<L>: 1494.0

<k>: 6

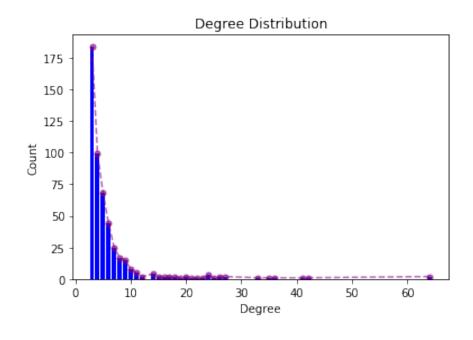
D: 0.011975951903807616 <d>: 3.401718228036341 C: 0.07724270763394936

N: 500 m: 3

<L>: 1494 <k>: 5.976

D: 0.011975951903807616 <d>: 3.2387014028056114

C: 0.011952



ba_3_5000.csv

-----Teórico-----

N: 5000 m: 3

<L>: 14994.0

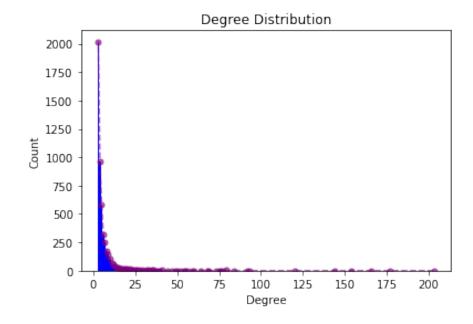
<k>: 6

D: 0.001199759951990398 <d>: 3.976119453625683 C: 0.014508515971981424

N: 5000 m: 3

<L>: 14994 <k>: 5.9976

D: 0.001199759951990398 <d>: 4.040210682136427 C: 0.001199520000000001



$ba_4_500.csv$

-----Teórico-----

N: 500 m: 4

<L>: 1990.0

<k>: 8

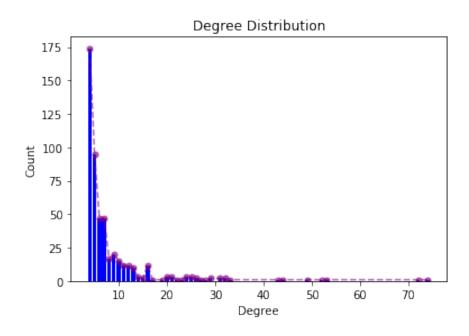
D: 0.01595190380761523 <d>: 3.401718228036341 C: 0.07724270763394936

N: 500 m: 4

<L>: 1990 <k>: 7.96

D: 0.01595190380761523 <d>: 2.946629258517034

C: 0.01592



ba_4_5000.csv

-----Teórico-----

N: 5000 m: 4

<L>: 19990.0

<k>: 8

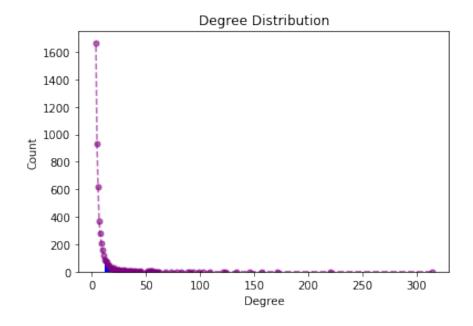
D: 0.0015995199039807962 <d>: 3.976119453625683 C: 0.014508515971981424

N: 5000 m: 4

<L>: 19990 <k>: 7.996

D: 0.0015995199039807962 <d>: 3.6969390678135627

C: 0.0015992



Debido a la comparación entre los datos teóricos y los obtenidos en estas redes se puede decir que la densidad vale prácticamente lo mismo en ambos casos debido a especificar la cantidad de aristas de nuestras redes; se trata de un parámetro que aumenta a medida que lo hace el número de aristas o disminuya el número de nodos.

También se puede confirmar que el grado medio (<k>) depende en gran medida del número de aristas.

El resto de métricas son también muy parecidas entre las redes teóricas y las generadas, con excepción del coeficiente de cluster, el cual es notablemente mayor en los datos teóricos. Esto se puede deber a la probabilidad de unión entre nodos, la cual puede sufrir variaciones al generar

las redes mostradas en este apartado, lo que puede generar estas variaciones entre los datos.

A continuación se prestará atención a la evolución del diámetro de la red y del coeficiente de cluster.

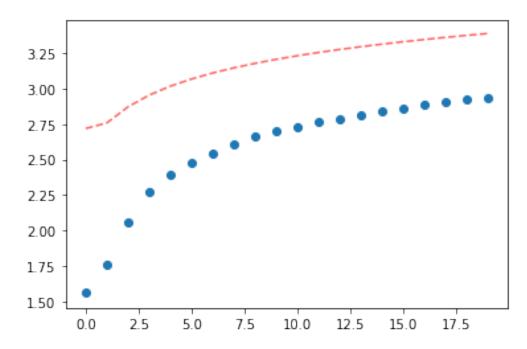
Aquí se puede ver como el diámetro y el coeficiente de cluster tienden a los resultados del modelo, cuantos mas nodos hay más se acerca a la función del modelo.

(en rojo la función teórica, y en azul según aumenta el numero de pasos)

../graphs/barabasi-albert\steps\ba_n_500_m_4_steps_475.csv

Actual Diameter Value: 2.934984520123839
Theoric Diameter Value: 3.3895367176856626
Difference: 0.4545521975618234

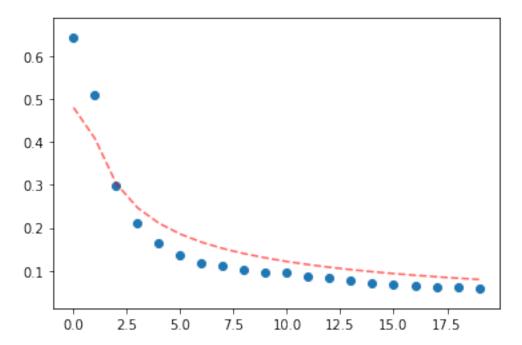
Mean absolute difference: 0.5861809926187067



../graphs/barabasi-albert\steps\ba_n_500_m_4_steps_475.csv

Actual Clustering Value: 0.05738586390277871
Theoric Clustering Value: 0.0798579355404949
Difference: 0.022472071637716196

Mean absolute difference: 0.04043597171375472



Como se puede observar en las gráficas de la distribución del grado, aumenta de forma potencial, siguiendo la *ley de potencias* del modelo teórico de Barabasi-Albert. Además, sobre todo en los primeros pasos de la simulación, podemos ver claramente otro dato característico, la *conexión preferencial*, donde vemos que los primeros nodos que llegan, finalmente son los que mas grado acaban teniendo. Por ultimo en las gráficas que muestran la evolución de dos valores diferentes, podemos observar como empiezan a converger el modelo teórico y el simulado.

4. Erdos-Renyi Model

Con el modelo de Erdos-Renyi además de incluir los nodos a generar, en este caso 500 para todos los grafos realizados con esto modelo, hay que añadir una probabilidad p=0< p<1 con la que cada nodo se conectara con probabilidad p con los otros n-1 nodos. Las etapas de evolución del grafo de erdos-renyi son:

- subcrítica (p < 1/N)
- crítica (p = 1/N)
- supercrítica (p > 1/N)
- conectada ($p > \frac{lnN}{N}$)

En estos ejemplos tendremos como número de nodos N=500

subcritica Name:

Type: Graph

Number of nodes: 500 Number of edges: 183 Average degree: 0.7320

Average degree. 0.7

critica Name:
Type: Graph

Number of nodes: 500 Number of edges: 244 Average degree: 0.9760

supercritica Name:

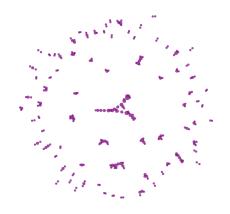
Type: Graph

Number of nodes: 500 Number of edges: 358 Average degree: 1.4320

conectada Name:
Type: Graph

Number of nodes: 500 Number of edges: 9989 Average degree: 39.9560

Como podemos observar, a mayor probabilidad, en los cambios de regiones obtenemos mas numero de enlaces y por lo tanto mayor grado medio



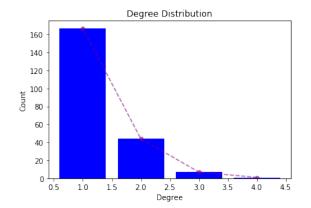
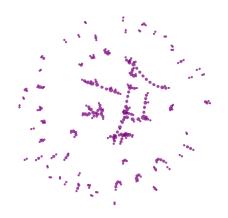
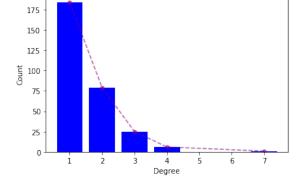


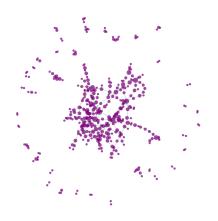
Figura 9: Subcrítica $p < \frac{1}{N} p = 0.0015$





Degree Distribution

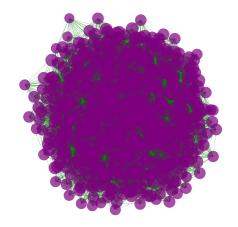
Figura 10: Crítica $p = \frac{1}{N} p = 0.002$



Degree Distribution

175
150
125
100
75
50
25
0
12 3 4 5 6

Figura 11: Supercrítica $p > \frac{1}{N} p = 0.003$



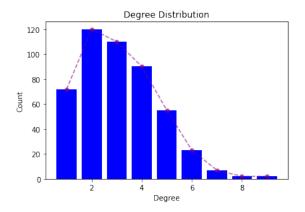


Figura 12: Conectada $p > \frac{\ln N}{N} \ p = 0.08$

Ahora pasamos a analizar diferentes métricas de las redes de erdos-renyi, de las cuales se han generado 5 de cada una y luego se hizo la media de todas las metrcias, con estas formulas se calcularan los valores teóricos para las diferentes propiedades, y de nuevo para calcular las de los modelos generados se usara *networkx*

- Nodos de la red: *N*
- Probabilidad: *P*
- Numero de enlaces: $L = p * L_{max}$
- Grado medio: < k >= p * (N 1)
- Densidad: $D = \frac{L}{L_{max}}$
- lacktriangle Numero de componentes conectadas: N_c
- Longitud media: $\langle d \rangle = \frac{\ln N}{\ln k}$
- Clustering: $C = \frac{k}{N}$

subcritica

-----Teórico-----

N: 500 P: 0.0015 <L>: 187.125 <k>: 0.7485 D: 0.0015 N_c: 316.4 <d>: -

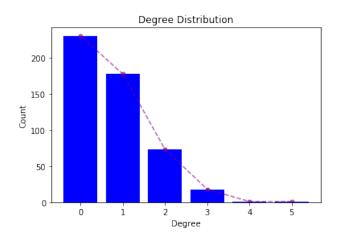
C: 0.001497

N: 500.0 P: 0.0015 <L>: 183.6

<k>: 0.734399999999999 D: 0.001471743486973948

N_c: 316.4 <d>: -

C: 0.00066096



critica

-----Teórico-----

N: 500 P: 0.002 <L>: 249.5 <k>: 0.998 D: 0.002 N_c: 255.6 <d>: -

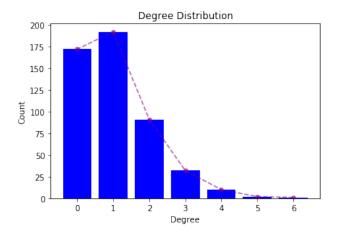
C: 0.001996

N: 500.0 P: 0.002 <L>: 245.4

<k>: 0.981599999999999 D: 0.0019671342685370742

N_c: 255.6 <d>: -

C: 0.00088842666666666



supercritica

-----Teórico-----

N: 500 P: 0.003 <L>: 374.25 <k>: 1.497 D: 0.003 N_c: 150.6 <d>: -C: 0.002994

0. 0.002001

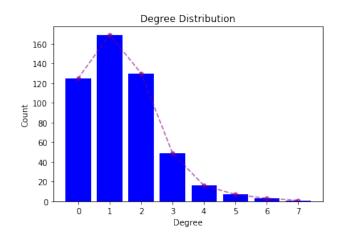
N: 500.0 P: 0.003 <L>: 368.4 <k>: 1.4736

D: 0.00295310621242485

N_c: 150.6

<d>: -

C: 0.001327066666666665



conectada

-----Teórico-----

N: 500 P: 0.08 <L>: 9980.0 <k>: 39.92 D: 0.08 N_c: 1.0

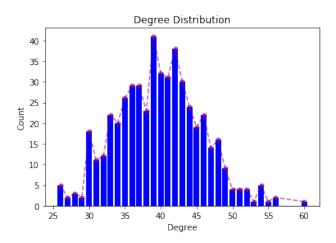
<d>: 1.685602025092921 C: 0.07984000000000001

N: 500.0 P: 0.08 <L>: 9945.8 <k>: 39.7832

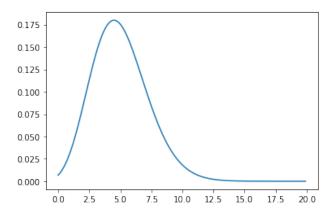
D: 0.07972585170340682

N_c: 1.0

<d>: 1.9587655310621244 C: 0.036362693333333335



Comparando los datos teóricos con lo resultados obtenidos en la práctica obtenemos datos muy similares y que según aumenta la conectividad de la red, su distribución de grado se va pareciendo más a una distribución de *Poisson*.



La etapa de red depende de la probabilidad de generar un nuevo enlace y del numero de nodos, por lo tanto para valores mayores de N, la probabilidad *p* para generar una región conectada deberá ser menor que para una red con menor numero de nodos.

El grado medio también irá cambiando dependiendo de la probabilidad que se elija y de los nodos que haya en el grafo. El grado medio influirá en algunas métricas, como en el clustering y en la longitud media.

La densidad depende del número de enlaces, que a su vez estará relacionada con la probabilidad. A su vez se puede observar que en la etapa conectada se ha obtenido un número menor de componentes conectadas, debido a que se forma una componente gigante completamente conectada, por eso en las regiones anteriores el numero de componentes conectadas aumenta.

Respecto al coeficiente de clustering que se ha obtenido comparado con el teórico, el resultante es más bajo ya que no existen nodos acumulados, y se puede observar el decrecimiento del dato debido a que el tamaño del grafo va aumentando y eso dificulta la formación de hubs.

5. Comparación de modelos

De los 2 modelos estudiados se puede observar que la red de la práctica 1 se parece mucho mas al modelo de Barabasi Albert.

Viendo la distribución de los grados del modelo y de la red real es posible ver como ambas siguen la ley de las potencias, haciendo que haya muchos nodos con poco grado y unos pocos con grado muy alto (los hubs). A diferencia de ellos dos, la Red Aleatoria (de erdos-renyi) tiene una distribución *binomial* o de *poisson*.

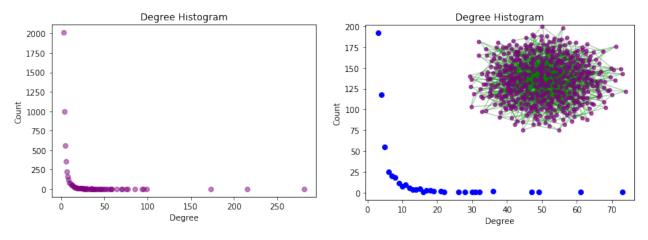


Figura 13: BA model n = 5000

Figura 14: BA model n = 500

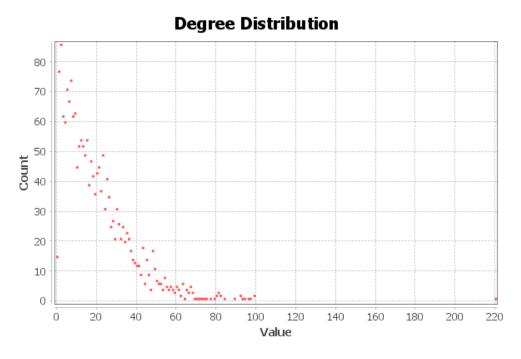


Figura 15: Degree Distribution from movie network

El modelo generado para que resultase lo mas similar a la red real ha sido con n=1895 y m=9 Se observa como la mayoría de medidas de la red de películas que se muestran no se desvían demasiado con el modelo generado.

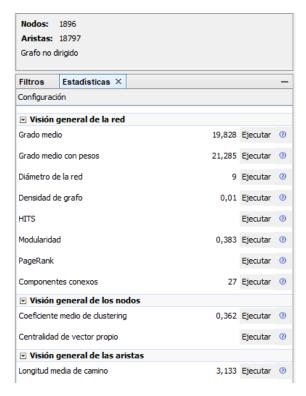


Figura 16: Medidas de la red de películas

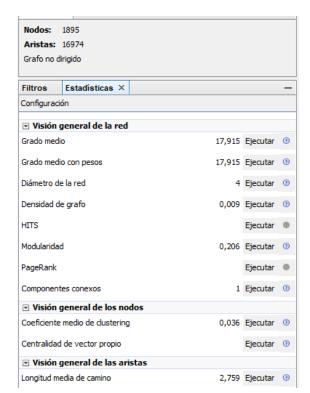


Figura 17: BA model n = 1895 m = 9