

CAIM – Sesión 7

Introduction to *igraph*



**UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH**

Santiago Arxé i Carbona
Bryan Leonardo Salto Salao

29/01/2022

Introducción

En esta sesión de laboratorio se repasarán y comprobarán conceptos explicados en teoría, se aprenderá a utilizar la librería *igraph* y cómo analizar, crear y representar una red.

Para realizar la sesión es importante saber que muchas redes en el mundo real tienden a seguir ciertas propiedades también conocido como fenómeno **Small-World**. Los grafos **tienden** a tener un diámetro pequeño y también **tienden** a ser transitivos. Gracias al estudio de los grafos, se pueden crear redes aleatorias que cumplan estas propiedades. Un ejemplo son los modelos de **Erdős-Rényi**, que siguen la primera propiedad, o el modelo **Watts-Strogatz**, que cumple las dos propiedades con un valor de p ($p \approx 0.01$).

La primera tarea será comprobar ese valor de p aproximado y comprobar que el *plot* resultante coincide con el de teoría. La segunda tarea será analizar una red y mirar si cumple las propiedades o es posible que se trate de un grafo random. Para ello se han creado dos scripts: *Tarea 1.ipynb* y *Tarea2.ipynb*.

Tarea 1

La principal misión del gráfico pedido en la práctica es representar la variación de diversas características del modelo de **Watts-Strogatz** en función del parámetro p (la probabilidad de conectar un nodo a otro nodo cualquiera).

Con tal de reproducir el gráfico en cuestión, era necesario calcular los coeficientes de *clustering* y el *camino más corto* en función del parámetro p , normalizar ambos vectores y representarlos conjuntamente.

Como puede observarse en la Figura 1, conforme aumenta la probabilidad de que los nodos se conecten entre sí, la distancia entre ellos disminuye de manera drástica. Pero por otra parte, el coeficiente de clustering también se ve reducido. Si queremos simular una red que cumpla el fenómeno Small-World (un diámetro bajo y un clustering alto) tenemos que escoger una $p = 0,01$. Si el valor de p es demasiado grande acabaríamos teniendo un grafo aleatorio (coef. clust. pequeño). Si es demasiado bajo el diámetro o la media del camino más corto serían muy grandes.

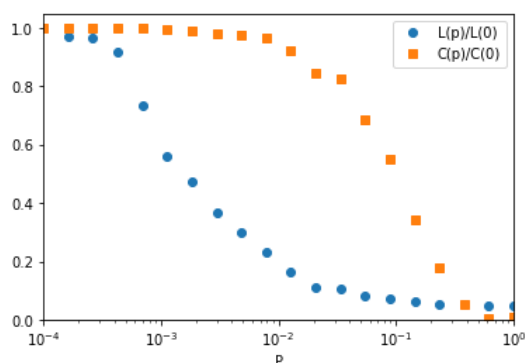


Figura 1: Representación del *average shortest path* (L) y del *clustering coefficient* (C)

Tarea 2

Descripción simplificada

Número de aristas	Número de nodos	Diámetro	Coeficiente de <i>clustering</i>
602	62	2	0,52

Tabla 1: Descripción simplificada de la red

Distribución de los grados

N = 62, mean +- sd: 19.4194	[22, 23): (0)	[43, 44): * (1)
+ - 12.5300	[23, 24): **** (4)	[44, 45): (0)
[3, 4): ** (2)	[24, 25): (0)	[45, 46): (0)
[4, 5): (0)	[25, 26): ** (2)	[46, 47): (0)
[5, 6): (0)	[26, 27): (0)	[47, 48): (0)
[6, 7): (0)	[27, 28): * (1)	[48, 49): (0)
[7, 8): ***** (8)	[28, 29): (0)	[49, 50): (0)
[8, 9): (0)	[29, 30): * (1)	[50, 51): (0)
[9, 10): ***** (5)	[30, 31): (0)	[51, 52): (0)
[10, 11): (0)	[31, 32): ** (2)	[52, 53): (0)
[11, 12): **** (4)	[32, 33): (0)	[53, 54): (0)
[12, 13): (0)	[33, 34): *** (3)	[54, 55): (0)
[13, 14): ***** (7)	[34, 35): (0)	[55, 56): (0)
[14, 15): (0)	[35, 36): *** (3)	[56, 57): (0)
[15, 16): ***** (7)	[36, 37): (0)	[57, 58): (0)
[16, 17): (0)	[37, 38): * (1)	[58, 59): (0)
[17, 18): *** (3)	[38, 39): (0)	[59, 60): * (1)
[18, 19): (0)	[39, 40): * (1)	[60, 61): (0)
[19, 20): ** (2)	[40, 41): (0)	[61, 62): (0)
[20, 21): (0)	[41, 42): (0)	[62, 63): (0)
[21, 22): *** (3)	[42, 43): (0)	[63, 64): * (1)

Tabla 2: Distribución de los grados de la red

La Tabla 2 muestra la salida obtenida al mostrar la distribución de los grados de la red proporcionada. A simple vista, no puede observarse ninguna distribución demasiado marcada, a la vez que puede destacarse una gran desviación estándar en cuanto a la distribución de la frecuencia de los grados (19,4). Como puede observarse en la Figura 2, es necesario adentrarse algo más en el estudio de la red para intentar identificar algún patrón.

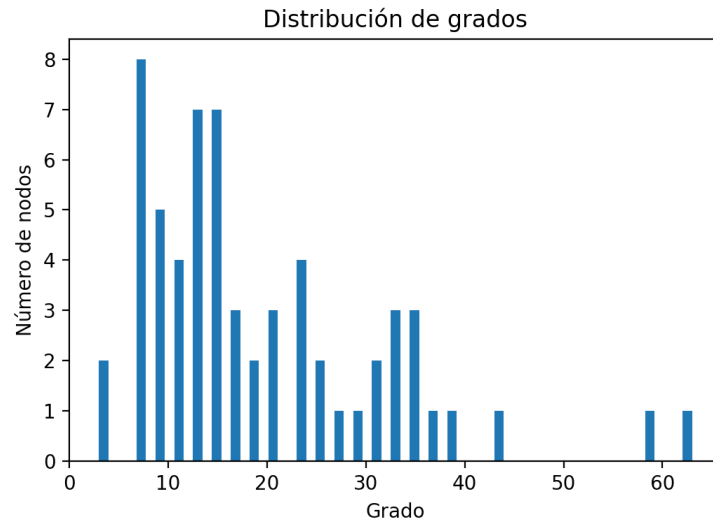


Figura 1: Representación de la distribución de las frecuencias de los grados

Representación de la red

Con tal de entender algo más la red a estudiar, es importante representarla de una manera correcta. Como ya se ha comentado en prácticas anteriores (y en clase), una de las características más básicas y representativas para entender las redes de nodos es la importancia (**PageRank**) de los nodos dentro de dicha red.

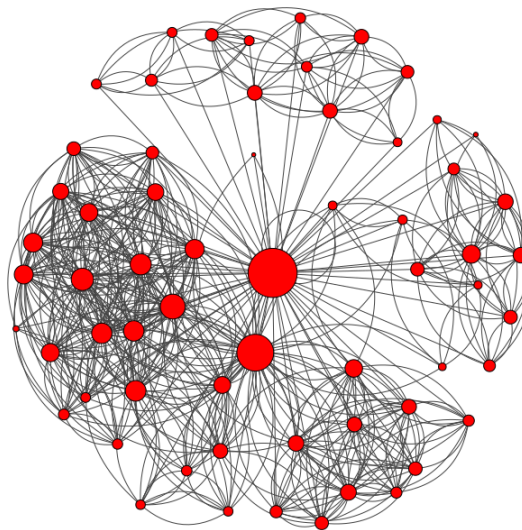


Figura 3: Representación de la red con el **PageRank** representado en el tamaño de los nodos

Por ello, se representó en la Figura 3 la red con las conexiones entre nodos y los nodos representados con un tamaño proporcional a su **PageRank**, lo cual permite comenzar a vislumbrar algunos patrones en dicha red. Vemos como hay dos nodos que son muy referenciados, cosa que podría explicar el diámetro pequeño.

Community detection (detección de comunidades)

La Figura 3 parece indicar la presencia de algunos grupos de nodos más interconectados entre sí. Estos grupos, también llamados comunidades, pueden detectarse y definirse a partir de su cercanía en cuanto a conexiones y la densidad de las mismas.

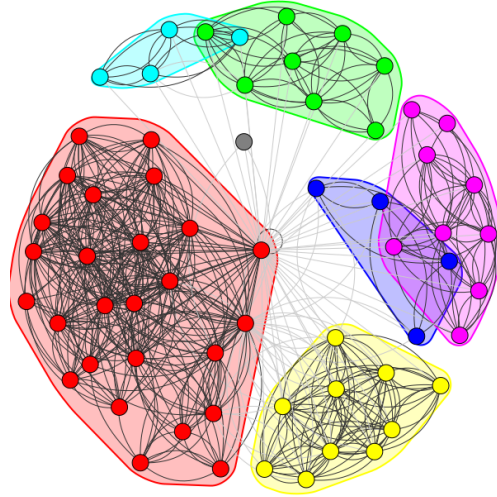


Figura 4: Comunidades detectadas en la red

Como ya podía intuirse, aparecen varios (7) *clusters* claramente diferenciados (algunos más que otros). Obviamente, los *clusters* siguen estando conectados entre ellos, pero la frecuencia de conexiones entre ciertos nodos permiten clasificarlos como un “grupo” con gran relación.

Los nodos, su cercanía y los *clusters* encontrados pueden representarse, también, como un árbol cuyas bifurcaciones van indicando las diferencias entre los diferentes grupos identificados. Como puede observarse en la Figura 5, este árbol permite ver, de manera no tan clara pero sí orientativa, los *clusters* ya mencionados anteriormente.

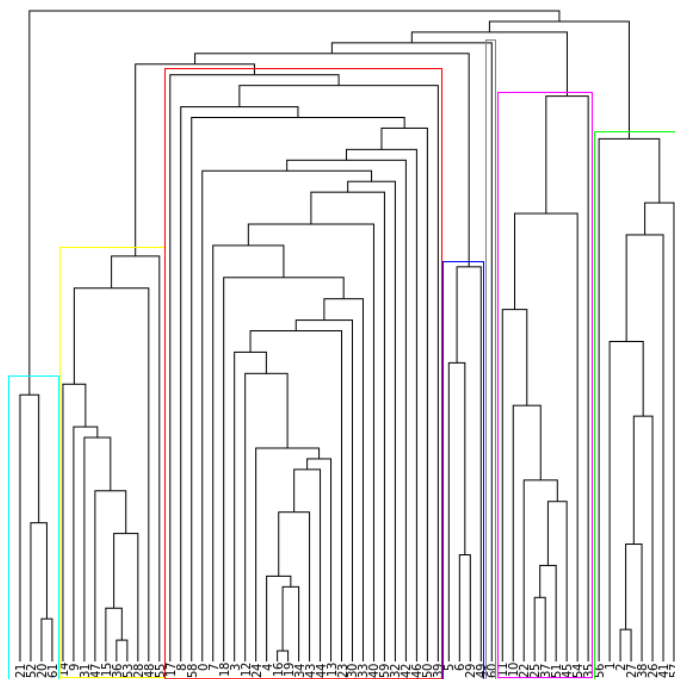


Figura 5: *Community edge betweenness* de la red con los *clusters* identificados