

[Cerca i Anàlisi d'Informació Massiva]

Session 7: Introduction to igragh

Ana Mestre Borges

Carlos Roldán Montaner



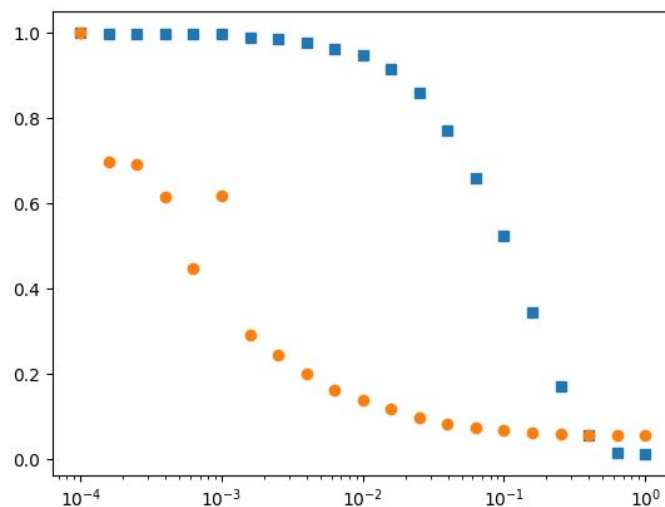
**UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH**

Introducción e implementación

En esta sesión de laboratorio vamos a estudiar el análisis de redes a partir del paquete `igraph`, en nuestro caso usaremos Python.

Task 1

Para reproducir el gráfico tal y como se pedía en la práctica, hemos tenido que calcular el coeficiente de clustering y el camino medio más corto como una función del parámetro p del modelo **Watts-Strogatz**. Para ello hemos hecho pruebas con diversos valores para el número de vértices, al principio habíamos probado con 100 pero no acababa de quedar una imagen muy clara así que hemos decidido aumentarlo a 1000. Si aumentamos más el valor se obtienen resultados muy parecidos pero es más lenta la ejecución, por lo que finalmente hemos decidido que este valor sería el adecuado. Los resultados se ven plasmados en el siguiente gráfico donde, además, se puede observar que los valores para el eje de las ordenadas está normalizado y que el eje de abscisas sigue una escala logarítmica.



Task 2

La siguiente parte de la práctica ha sido analizar una red concreta que nos ha sido dada y con la que hemos obtenido los siguientes resultados:

Number of edges	Number of nodes	Diameter	Clustering coefficient
602	61	2	0.522769

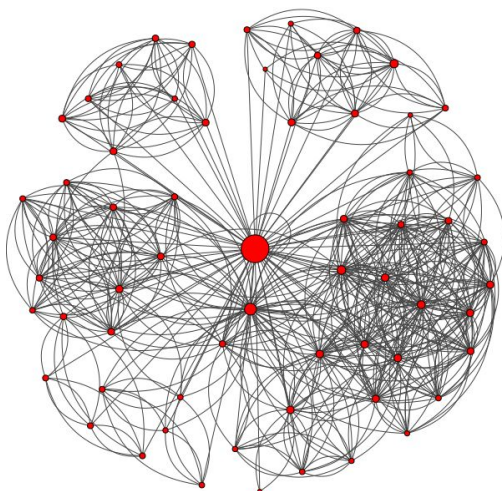
Degree distribution:

N = 62, mean +- sd:
19.4194 +- 12.5300

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

Degrees:

[63, 13, 9, 39, 35, 7, 9, 59, 13, 25, 13, 11, 33, 31, 19, 21, 27, 7, 23, 33, 7, 9, 13, 35, 37, 17, 13, 13, 17, 9, 35, 23, 25, 33, 31, 3, 21, 11, 11, 13, 21, 11, 23, 29, 43, 15, 15, 23, 15, 7, 15, 15, 7, 19, 7, 15, 7, 9, 15, 17, 3, 7]



Para analizar esta red observamos la distribución de los grados, que nos aporta información para ver cómo está repartida la red y nos puede ayudar a hacernos una idea de si se trata de una red aleatoria o no. En este caso podemos intuir que no lo es ya que vemos una organización un poco uniforme de los grados.

Para poder confirmar nuestras sospechas visualizamos la red a partir de un gráfico que representa el tamaño de cada nodo a partir del valor de su PageRank. Observamos que la mayoría de nodos tienen más o menos el mismo tamaño exceptuando unos pocos, es decir, se pueden apreciar unas comunidades bien definidas. También es fácil darse cuenta de que tiene una estructura bastante definida, constituida por un nodo central con mucha más importancia que el resto, al que están conectadas todo el resto de comunidades, cuyos nodos son mucho menos importantes. Por lo que podríamos concluir que se trata de una red no aleatoria.

Community detection:

En esta parte nos centraremos en estudiar las comunidades de la red, para ello hemos creado un grafo mediante el generador Erdos-Renyi y hemos usado el algoritmo `community_edge_betweenness()` y hemos obtenido que el tamaño de la mayor comunidad es 10. En las dos siguientes figuras se pueden observar las comunidades obtenidas, en la izquierda el histograma de tamaños de las comunidades y a la derecha el grafo con dichas comunidades. Con ambos se obtiene la misma información, el histograma nos dice que hay 5 comunidades de tamaño 1, otra de tamaño 3 y finalmente, la mayor, de tamaño 12. Lo mismo podemos observar si nos dedicamos a contar los vértices del mismo color en el grafo de la derecha, que en nuestro caso no es ningún problema, pero si se tratara de uno más grande sería más claro extraer los datos del histograma.

