# Pràctica 7: Introduction to igraph

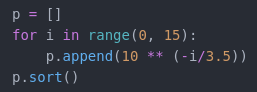
Introducció

En aquesta sessió ens hem introduït en el mòdul igraph per estudiar diferents networks i comunitats de nodes, en aquest cas de grafs no dirigits.

Com que cap de les dues domina el llenguatge R i a més no tenim instalat R studio als nostres ordinadors, hem decidit usar python i fixar-nos en el Notebook de Jupyter per a programar les funcions de la llibreria igraph.

Task 1

En aquesta primera tasca la nostra missió era reproduir el graf de l’enunciat. Per a fer-ho hem calculat quants punt contenia el graf i hem dividit els punts entre 0.0001 i 1 en aquesta quantitat, per aconseguir que el vector de probabilitats fos el més semblant possible. Ho hem fet de la següent forma:



Apart hem calculat el coefficient de clustering i el avarage shortest-path normalitzats amb el model de Watts-Strogatz amb el vector de probabilitats mostrat anteriorment.

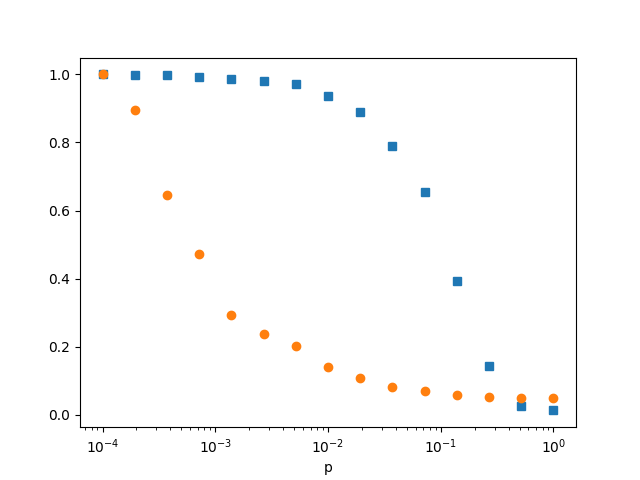
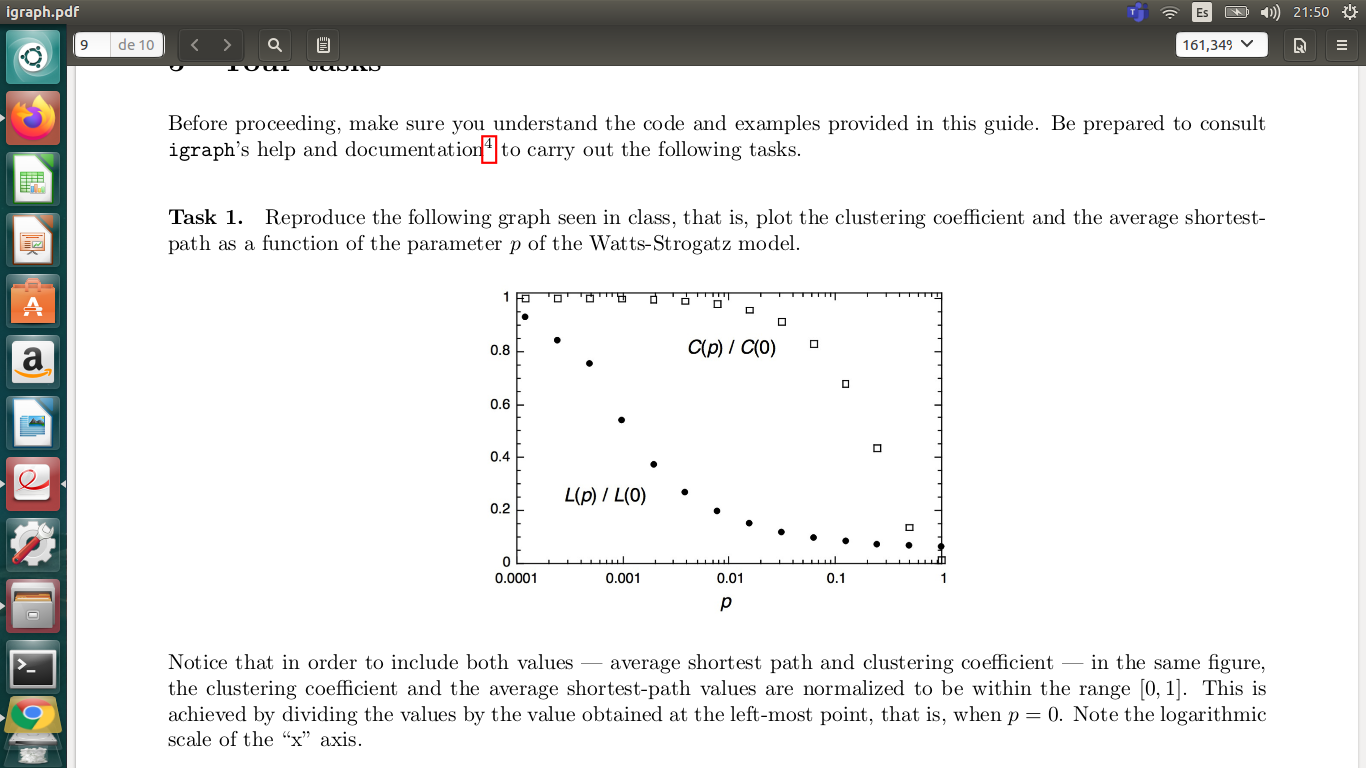
Per a implantar la funció primerament vam usar de mostra la del notebook:

*watts = Graph.Watts\_Strogatz(1,100,2,0.05).*

Però amb aquesta vam veure que no s’aconseguia l’evolució que té la funció del pd. Per tant, vam buscar el significat de cada paràmetre de la funció per a poder adaptarlos millor: (dim, size, nei, p). Per a adaptarlo doncs, hem pujat la size a 1000 i el nei (nombre de vèrtex veïns al que està connectat cada node en dues direccions) l’hem pujat a 3 però donava resultats similars en cas de 2, 3, i 4, i hem decidit quedar-nos amb l’intermig.

I per últim la probabilitat corresponent de cadascuna a la posició del vector de probabilitats calculat anteriorment.

Finalment hem obtingut un graf bastant semblant:



Task 2

Pregunta 1

En aquesta segona part hem analitzat el graf del fitxer edges.txt. En primer lloc, hem llegit els nodes i les arestes per saber les seves dimensions, així com el diàmetre i la transitivitat, obtingudes a partir de les funcions auxiliars facilitades per la llibreria igraph. Els resultats són els següents:

* Número de nodes: 61
* Número d’arestes: 602
* Diàmetre: 2
* Transitivitat: 0.5227690047741461

Amb aquestes dades ja ens fem una idea de les dimensions del graf que estem tractant i de la connectivitat dels seus vèrtexs. El diàmetre del cluster indica que el camí curt més llarg entre dos nodes del graf és 2, el qual és un diàmetre bastant petit per les dimensions del graf, qualitat característica d’un real network . La transitivitat (o coeficient de clustering) ens indica que hi ha al voltant d’un 50% de possibilitats de que dos vèrtexs connectats a un cert node, també ho estiguin entre ells.

També hem analitzat la distribució dels graus dels vèrtexs i hem analitzat la seva forma a partir d’un histograma, on es veu més clar aquesta distribució:

* Degree distribution:

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

[ 3, 4): \*\* (2)

[ 4, 5): (0)

[ 5, 6): (0)

[ 6, 7): (0)

[ 7, 8): \*\*\*\*\*\*\*\* (8)

[ 8, 9): (0)

[ 9, 10): \*\*\*\*\* (5)

[10, 11): (0)

[11, 12): \*\*\*\* (4)

[12, 13): (0)

[13, 14): \*\*\*\*\*\*\* (7)

[14, 15): (0)

[15, 16): \*\*\*\*\*\*\* (7)

[16, 17): (0)

[17, 18): \*\*\* (3)

[18, 19): (0)

[19, 20): \*\* (2)

[20, 21): (0)

[21, 22): \*\*\* (3)

[22, 23): (0)

[23, 24): \*\*\*\* (4)

[24, 25): (0)

[25, 26): \*\* (2)

[26, 27): (0)

[27, 28): \* (1)

[28, 29): (0)

[29, 30): \* (1)

[30, 31): (0)

[31, 32): \*\* (2)

[32, 33): (0)

[33, 34): \*\*\* (3)

[34, 35): (0)

[35, 36): \*\*\* (3)

[36, 37): (0)

[37, 38): \* (1)

[38, 39): (0)

[39, 40): \* (1)

[40, 41): (0)

[41, 42): (0)

[42, 43): (0)

[43, 44): \* (1)

[44, 45): (0)

[45, 46): (0)

[46, 47): (0)

[47, 48): (0)

[48, 49): (0)

[49, 50): (0)

[50, 51): (0)

[51, 52): (0)

[52, 53): (0)

[53, 54): (0)

[54, 55): (0)

[55, 56): (0)

[56, 57): (0)

[57, 58): (0)

[58, 59): (0)

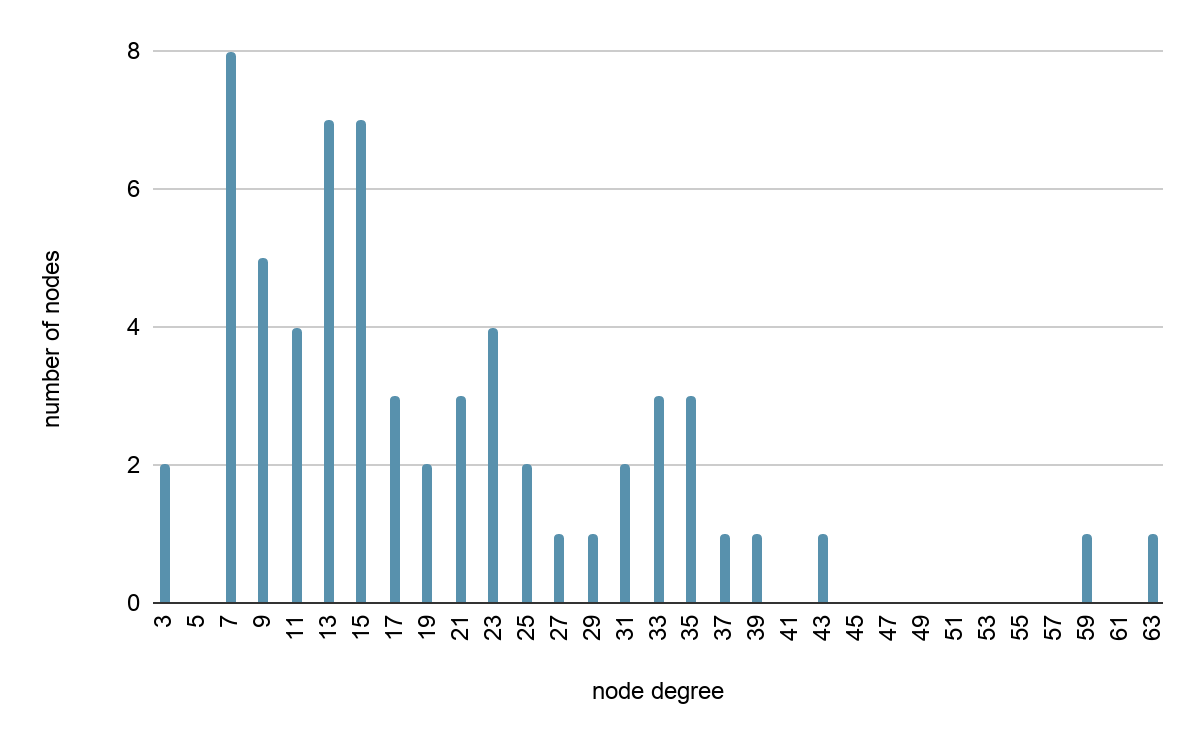
[59, 60): \* (1)

[60, 61): (0)

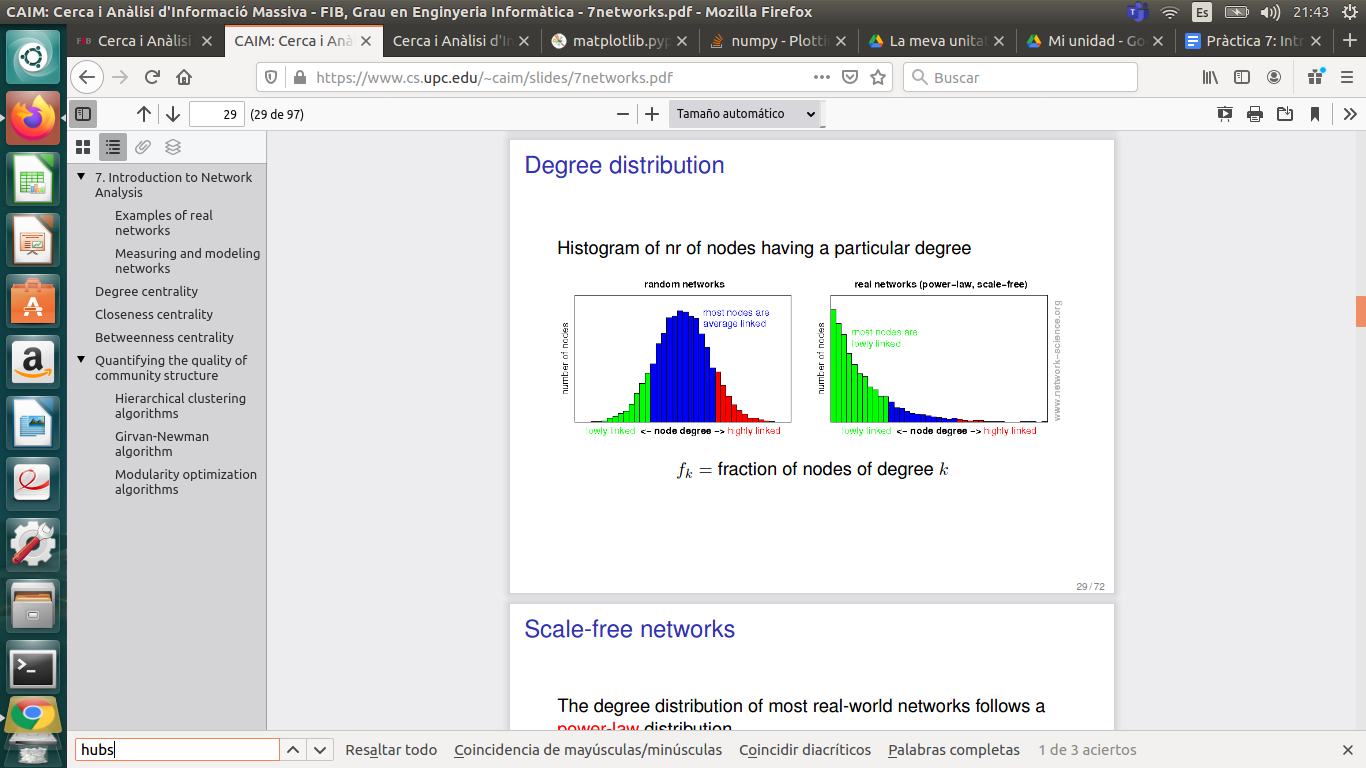
[61, 62): (0)

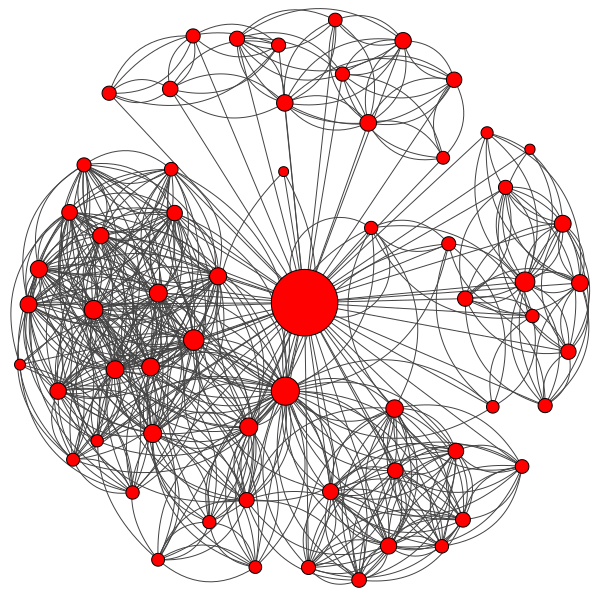
[62, 63): (0)

[63, 64): \* (1)



Com es pot observar al diagrama, la distribució dels graus dels nodes segueix una distribució *power-law,* pròpia dels real network, el que ens fa pensar que no es tracta d’un graf aleatori. En aquestes imatges de teoria podem veure que la nostra figura s’assembla molt més a la distribució de la dreta (real network).





Per últim en aquesta primera pregunta, hem visualitzat una representació d’aquest graf, modificant la mida dels vèrtexs proporcional al seu pagerank. Observem que hi ha un node central que connecta amb tots els altres nodes del graf. Després de la nostra deducció de que es tracta d’un real network, podem dir que es tracta d’un hub (característics en aquests tipus de graf), és a dir un node amb un grau molt elevat.

Pregunta 2

En aquesta segona pregunta, hem decidit usar l’algorisme community multilevel per a detectar les comunitats de la nostra xarxa. Primer hem obtingut la mida de la comunitat amb més nodes, 19 vèrtexs. A continuació hem generat dues imatges, un histograma sobre les mides de les comunitats, i una representació del graf amb cada comunitat marcada amb una regió de diferent color. Com indiquen les imatges trobem 6 comunitats distingibles de diferents mides, sent la més petita de 4 vèrtexs i la més gran de 19, com ja hem dit.

