KALLEL Mahdi
mahdi.kallel@telecom-paris.fr

Lab session # 4
ALTEGRAD 2020

12/21/20

1 Question 1

G contains two connected components. One fully connected with a 100 vertices and another which is a complete bipartile graph with 50 edges in each partition. For the first component we get C_{100}^2 edges ,for the second for each vertex in the first partition we have an edge going for the vertices of the second partition, this yields $50 * 50$ for From this we get that G has $\frac{100(99}{2} + 50 * 50 = 7450$.

By definition of the bipartile graph, it does not contain any triangles. For the fully connected graph choosing a triangle comes to choosing any triplet of vertices. Therefore the number of triangles is C_{100}^3

2 Question 2

The global clustering coefficient is $\frac{3|\text{Number of distinct triangles}|}{|\text{Number of distinct neighbours}|}$ for a vertex with degree d, the number of distinct neighbours is C_d^2 The coefficient is : $\frac{\sum_V N_t^V}{\sum_V C_{d_v}^2}$ where N_t^v is the number of distinct triangles in which the vertex V takes part. $\sum_V N_t^V = 3 *$ **Number of distinct triangles** because every triangle will figure 3 times in sum, once for each of its vertices.

For each pair of distinct neighbours that is connected, Edge V takes part in a triangle. Thus we get that $N_t^V \leq C_{d_v}^2$. Thus we get that $\frac{\sum_V N_t^V}{\sum_V C_{d_v}^2} \leq 1 \implies C \leq 1$

This equality if strong if we are dealing with a complete graph, or with a graph whose connected components are all complete and at least of size 3.

3 Question 3

In the adjacency matrix A, we have that for each line $\sum_j A_{i,j} = D_i$ where D_i is the degree of the i^{th} node. Normalizing with D^{-1} we get that $\sum_j A'_{i,j} = 1$. The matrix L_{rw} is symetric as the sum of two symetric matrices. Using property that " Any symmetric real matrix is positive definite if all of its diagonal entries are positive and greater than the sum of the absolute values of the corresponding off-diagonal entries". We can prove that $L_{\text{rw}} \in \boldsymbol{S}_n^+$.
For the vector $v = \mathbf{1}^T$ we have that $D^{-1}Av = \mathbf{1}^T$ as the sum of each line of D^{-1} is equal to one.
$\implies L_{\text{rw}}v = (I - D^{-1}A)v = 0 \implies$ v is an eigenvector of the minimal possible eigenvalue of L which is 0.

K means relies on the L2 norm for clustering points. In our case each line of U consists a point. In the case we keep v , the L2 norm of every couple of points is not affected as they have the "1" component in the same location. Therefore keeping v will not affect the k-means result in any way. This vector is an eigenvector whatever is the choice of our graph, so it does not bring any new information.

4 Question 4

The result of the k-means clustering depends on the initialization chosen for the clusters. This initialization is usually chosen at random. By changing the initialization we can change the final clusters and thus affect the spectral clustering. This is a stochastic algorithm.

5 Question 5

$$Q = \sum^{n_c} [\frac{l_c}{m} - (\frac{d_c}{2m})^2]$$

In the clustering (a) we have the following degree per node :
(1:3 , 2:2 , 3:3 , 4:3 , 5:2) / (6:4 , 7:3 , 8:3 , 9:3)

$$Q(A) = \frac{6}{13} - (\frac{13}{26})^2 + \frac{6}{13} - (\frac{14}{26})^2 = 0.42$$

In the clustering (b) we have the following degree per node:
(1:3, 2:2 , 8:3 , 9:3) / (3:3 , 4:2 , 5:2 , 6 :4 , 7 :3)

$$Q(B) = \frac{2}{13} - (\frac{11}{26})^2 + \frac{4}{13} - (\frac{14}{26})^2 = 0.146$$

6 Question 6

For c_4 we have that $\Phi(c_4) = (8, 4, 0, 0)$
For p_4 we have that $\Phi(p_4) = (6, 4, 2, 0)$

$$\Phi(c_4)^T\Phi(c_4) = 80,\ \Phi(p_4)^T\Phi(p_4) = 56,\ \Phi(c_4)^T\Phi(p_4) = 64$$

References