Denis Duval
denis.duval13@gmail.com

Lab session # 4
ALTEGRAD 2022

11/29/22

1 Question 1

To count the number of edges of the graph we just have to count the number of edges on each connected component and sum these numbers. For a complete graph (V,E), each subset of size 2 of V is an edge meaning all the $\binom{n}{2} = n(n-1)/2$ subsets of size 2 of V are represented in E. So the first connected component has 4950 edges. For a bipartite graph, we can list all the edges by indexing them on the vertex of one (fixed) of the two classes the edge is linking. Hence counting them is just summing the degrees of the vertices of one of the two classes. But if it is furthermore complete, then the degrees of the vertices will all be equal to the size of the other class. Hence the number of edges is the product of the sizes of the two classes. So the second connected component has 2500 edges. And the graph has 7450 edges.
Similarly, the number of triangles of a complete graph is the number of subsets of V of size 3 where each vertex in the subset is linked to the others. For a complete graph every subset of V of size 3 verifies this so we have $\binom{n}{3}$ of triangles. So the first component has 153450 triangles. Also, for a bipartite graph it is impossible to have a triangle. In fact this is incompatible with the bipartite property (if one exists then at least two vertices of a same class are the endpoints of an edge). The graph has thus 153450 triangles.

2 Question 2

The global clustering coefficient is defined as $\frac{3|triangles|}{|triads|}$, with a triad being two edges with a common vertex. The maximum value of is 1 and is reach for the graphs that have complete connected components.
In fact it is smaller than 1 because if we are given a triangle of a graph then we can assign it three triads (one for each vertex). So clearly 3 times the number of triangles is less or equal than the number of triads. Now for a complete graph this is an equality because each triad is in fact a triangle. It is the same for a graph consisting of multiple complete connected compenents.
Conversely, if this is an equality then we can show that every connected component is complete. Let's take a connected component and two nodes u and v within it. Because they belong to the same component there must exist a path $u_1 = u, u_2, ..., u_n = v$ with $n \geq 2$ and u_i, u_{i+1} share an edge for all i. Now we can prove by induction that u and u_i share an edge for $i \geq 2$. This is trivially true for $i = 2$. Suppose that u and u_i share an edge. Because u_i, u_{i+1} also share and edge it is a triad and hence a triangle (because the inequality was an equality). So u and u_{i+1} share an edge. So u and v share an edge.

3 Question 3

We can show that $D^{-1}A$ is a stochastic matrix. Hence all its eigenvalues have an absolute value smaller or equal to 1. Hence $I - D^{-1}A$ has non-negative eigenvalues. Also, because $D^{-1}A$ is stochastic the vector $(1, ..., 1)^T$ is an eigenvector of $I - D^{-1}A$ associated to the eigenvalue 0. It can be shown that it is the only one. If it is ignored it doesn't change the clusters of the k-means. In fact this vector add a same value in the first component for the m vectors that we try to cluster. But because the first component is the same, the distance between two of the $m - 1$ vectors depends only on the last $d - 1$ coordinates.

4 Question 4

The first three steps are deterministic (we could argue that the third step is stochastic if there is a choice to do if there are several eigenvectors associated to a same eigenvalue). The fourth step is stochastic, in fact it comes from that the initialisation of the clusters is random. So the algorithm has a stochastic output.

5 Question 5

For the first clustering result:
- the green community has 6 edges within it and has a sum of degrees $3 + 2 + 3 + 3 + 2 = 13$.
- the blue community has 6 edges within it and has a sum of degrees $3 + 3 + 3 + 4 = 13$.
The graph as $m = 13$ edges. So $Q = 6/13 - (13/26)^2 + 6/13 - (13/26)^2 \simeq 0.423$.

For the second clustering result:
- the green community has 2 edges within it and has a sum of degrees $3+2+3+3=11$.
- the blue community has 4 edges within it and has a sum of degrees $3+3+2+4+3=15$.
The graph as $m=13$ edges. So $Q=2/13-(11/26)^2+4/13-(15/26)^2 \simeq -0.05$.

6 Question 6

The feature vector associated with C_4 is $[4,4,0,...]$(the vertices lie with distance at most 2 from each other because of the cycle). The feature vector associated with P_4 is $[3,2,1,0,...]$.
The kernel evaluated at (C_4, C_4) is 32.
The kernel evaluated at (C_4, P_4) is 20.
The kernel evaluated at (P_4, P_4) is 14.