Matias Etcheverry
matias.etcheverry9@gmail.com

Lab session # 4
ALTEGRAD 2022

11/22/22

1 Question 1

A complete graph with n vertices is made of $\sum_{i=1}^{n-1} i = \frac{n(n-1)}{2}$ edges. A complete bipartite graph with the sets of points $\mathcal{S}$ and $\mathcal{T}$ is made of $|\mathcal{S}| \times |\mathcal{T}|$ edges: each vertex of in $\mathcal{S}$ is linked to each vertex in $\mathcal{T}$. In total, in our example, there are $\frac{100\times 99}{2} + 50 \times 50 = 7450$ edges.

To create a triangle in a complete graph, we may take any set of 3 points. As they are linked, they will form a triangle. Thus, there are $\binom{n}{3} = \frac{n!}{6(n-3)!} = \frac{n(n-1)(n-2)}{6}$ triangles. In a complete bipartite graph, there is no triangle. Indeed, a triangle in such graph would mean that there is an edge between 2 vertices in the same part of the graph (either in $\mathcal{S}$ or in $\mathcal{T}$). This is impossible by definition. In total, in our example, there are $161\,700$ triangles in the graph.

2 Question 2

Let's compute the Global Clustering Coefficient (GCC) on a graph $\mathcal{G}$. Let o the number of open triplets, and c the number of closed triplets. By definition, transitivity$(\mathcal{G}) = \frac{c}{c+o}$. Then,

$$0 \leq \text{transitivity}(\mathcal{G}) = \frac{c}{c+o} \leq 1$$

The highest possible transitivity is obtained when $o = 0$, that is when there is no open triplets in the graph. To do so, we need the graph to be complete.
Similarly, when the graph is complete, its transitivity is equal to 1.

3 Question 3

The Laplacian matrix L_{rw} is symmetric, positive and semi-definite. Thus, all its eigenvalues are non negative. Moreover, let $x \in \mathbb{R}^{|V|}$ the vector made of ones whose length is the number of vertex in the graph:

$$AV = DV$$

Indeed, each term equals the vector giving the degree of each vertex. And,

$$L_{\text{rw}}V = (I - D^{-1}A)V = V - V = 0$$

Thus V is the eigenvector of the eigenvalue 0.
This eigenvector can simply be ignored in the KNN. Indeed, it introduces a feature which equals to 1 for all node in the graph. It doesn't provide any structural information.

4 Question 4

The initialization of the K-Means algorithm being stochastic, the spectral clustering is stochastic. Indeed, the initialization of this algorithm `k-means++` has some random steps.

5 Question 5

Let's compute the modularity of the following graphs:

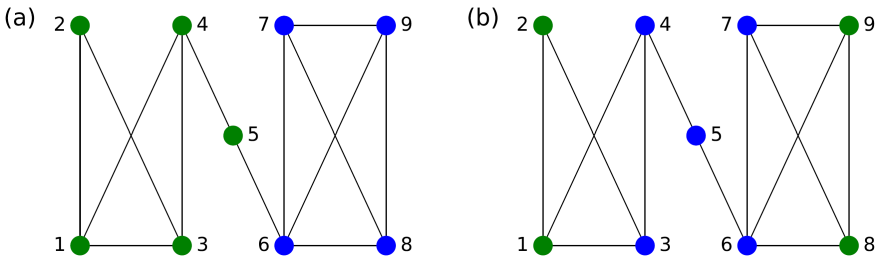

Figure 1: Graphs to compute modularity

In both cases, $m = |E| = 13$, $n_c = 2$

Parameter	Graph $\mathcal{A}$	Graph $\mathcal{B}$
l_{blue}	6	4
l_{green}	6	2
d_{blue}	13	15
d_{green}	13	11
$\frac{l_{\text{blue}}}{m} - (\frac{d_{\text{blue}}}{2m})^2$ $\frac{l_{\text{green}}}{m} - (\frac{d_{\text{green}}}{2m})^2$	$\frac{11}{52}$ $\frac{11}{52}$	$\frac{-17}{676}$ $\frac{-17}{676}$
Q	$\frac{11}{26} \approx 0.42$	$\frac{-17}{338} \approx -0.050$

In the end, we have $Q(\mathcal{A}) = 0.42$ and $Q(\mathcal{B}) = -0.050$. This result seems logical as the classification on the second graph is not good at all.

6 Question 6

Let's compute the shortest path kernels on the graphs below:

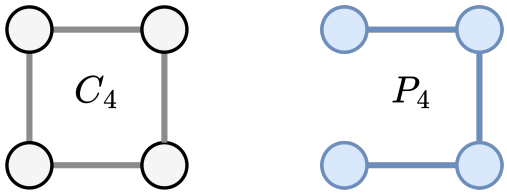

Figure 2: Graphs to compute kernels from

To do so, the easiest way is to compute the vector of shortest path distance frequencies:

Frequency	Graph C_4	Graph P_4
ϕ	$[4, 4, 0, 0]$	$[3, 2, 1, 0]$

In the end, we simply need to compute the scalar product to find the shortest path kernels on the different pairs:

Kernel value	Graph C_4	Graph P_4
Graph C_4	32	20
Graph P_4	20	14

Hopefully, the values are symmetric.

References