AA Problema 1

Grup 1.1

8 de juny de 2021

To prove that Clustering is NP-complete, we need to prove that it's NP and NP-hard. Trivially, it's in NP because we can just iterate over all pairs of of points and check that if the distance between them is more than B then they are in different clusters, which takes $\mathcal{O}(n^2)$ time. To prove that it's NP-hard, we will reduce a known NP-hard problem to it. In this case, we'll reduce Vertex Coloring to Clustering. The Vertex Coloring problem is defined as follows:

> Given a simple undirected graph G and an integer k, decide whether you can assign a color to each vertex in G such that no more than k unique colors are used and no two vertices with the same color are connected by an edge.

Given an instance of Vertex Coloring with $G(V, E)$ and k, we define an instance of Clustering as follows: use the same k, define p_i for each $v_i \in V$, and define

$$d(p_i, p_j) = \begin{cases} 0 & i = j \\ 1 & i \neq j, (v_i, v_j) \notin E \\ 2 & i \neq j, (v_i, v_j) \in E \end{cases}$$

This reduction can be done in polynomial time, and the new instance satisfies the restrictions of the Clustering problem, as $d(p_i, p_i) = 0$, $d(p_i, p_j) > 0$ for $i \neq j$, and $d(p_i, p_j) = d(p_j, p_i)$ because G is undirected.

Choose $B = \frac{3}{2}$. The answer to the Vertex Coloring problem is yes iff the answer to the Clustering problem is yes, because we can color all the vertices in the same cluster with the same color, and since $B < 2$, no two vertices in that cluster will have an edge between them. We'll use at most k colors since there are at most k clusters. Therefore, we have reduced Vertex Coloring to Clustering, concluding our proof that Clustering is NP-complete.