CS 4820, Spring 2019	Homework 10, Problem 1
Name: Kevin Klaben NetID: kek228	

(1) *(15 points)* Consider the problem of MAX-SPREAD, where we are given as input a list of locations $\{1, \ldots, n\}$, and a distance function d (supplied as a $n \times n$ matrix), where $d(i, j)$ denotes the distance between location i and location j, and an integer C. The distance function is restricted to be a metric (see below for a definition). For a set $S \subseteq \{1, \ldots, n\}$, define $spread(S) = \min_{i,j \in S, i \neq j} d(i, j)$. Your task is to output a subset $S \subseteq \{1, \ldots, n\}$ of cardinality C that **maximizes** $spread(S)$.

(1a) *(5 points)* Consider the decision version of MAX-SPREAD, denoted by D-MAX-SPREAD, where the input, in addition to the input of MAX-SPREAD, contains an integer α, and it is a YES instance if and only if there exists a subset $S \subseteq \{1, \ldots, n\}$ of cardinality C such that $spread(S)$ **is at least** α. Prove that D-MAX-SPREAD is NP-complete.

(1b) *(10 points)* Design an efficient 2-approximation algorithm for MAX-SPREAD.

[*Definition of a metric:* The distance function $d : \{1, \ldots n\} \times \{1, \ldots, n\} \to \mathbb{R}$ is a metric if it satisfies the following constraints: for all $i, j, k \in \{1, \ldots, n\}$,

- $d(i, j) \geq 0$
- $d(i, j) = 0$ if and only if $i = j$,
- $d(i, j) = d(j, i)$,
- $d(i, j) + d(j, k) \geq d(i, k)$.]

(1a) Step 1: Poly-time verifier
Given a solution subset S which is of cardinality C we can certainly verify it as correct in polynomial time. We can do this through two nested for loops. We first loop over each element of S and then loop over each other element of S. We simply check in constant time if the value in the matrix between these two items in S is at least alpha. If it is then we continue looping if not then we return false. If the end of the outer loop terminates without returning false then we return true meaning the input is a correct solution.

Reduce from Independent set to D-MAX-SPREAD

Take in instance of Independent set, we have an undirected graph of n nodes. If there are n nodes in the undirected graph we can initialize a matrix d that is n x n with all the values initialized to 1 and then each value d[i][i] is set to zero. Next we can iterate over all the nodes in the undirected graph thus for each node i in the graph and for each other node in the graph j if i is connected to j then d[i][j]=2 and d[j][i]=2. If i and j are not connected then the values remain untouched. Value of alpha will be set at 2. In addition, an instance of IS requires that a subset of at least C nodes be found which are independent, this same C is used as the C input to D-MAX-SPREAD which is the cardinality that S must be. It is clear that the constructed d is a metric as it satisfies all condition of a metric.

It clearly satisfies $d(i, j) \geq 0$ as every distance value is initialized to 1 or 0 (on the diagonal) and can only be changed to 2.

The second condition $d(i,j) = 0$ if and only if $i = j$ is satisfied as the diagonal is initialized to zero and those values cannot be changed, while all other values are initialized to 1 and can only be changed to 2.

The third condition $d(i,j) = d(j,i)$ is satisfied as all values as such are initialized to 1 and when they are changed to 2 they must be changed together.

The final condition $d(i,j) + d(j,k) \geq d(i,k)$ must be satisfied as the only values present which might cause the inequality to be false require either d(i,j) or d(j,k) to be zero. If either is zero, then either i=j or j=k. If i=j then d(j,k)=d(i,k) and the inequality holds, similar when j=k, then d(i,j)=d(i,k) and again the inequality holds. As this is true in all cases, the d constructed is a metric.

The reduction is constructed in polynomial time. This can be seen as to construct d we iterate over each node in the graph and each node it could potentially be connected to thus taking O(n^2) time.

Claim: Given a solution to independent Set we also have a solution to D-MAX-SPREAD.

Proof: Given the nodes, in the solution are independent, we know none of them are connected directly. This means that these same node in the D-MAX-SPREAD will all be connected via distances of 2. This means that all spreads will be at least 2 and we will output YES.

Claim: Given a solution to D-MAX-SPREAD we also have a solution to IS.

Proof: Given the nodes of D-MAX-SPREAD all must be connected via distances of 2. This indicates that these nodes were not connected in the original graph in IS and thus these same nodes are an independent set solution.

(1b) IDK