CS 580 Algorithm Design, Analysis, And Implementation
Fall 2021 Vassilis Zikas HW 6

Problem 1

Collaborators: List students that you have discussed problem 1 with: Vishnu Teja Narapareddy, Tulika Sureka
Let OPT be the maximum distance of a vertex from its cluster center in the Optimal solution. We need to show that the cost of clustering obtained from the given Greedy algorithm in the problem is atmost 2*OPT.
Proof by contradiction:
a) Assume that the distance from the farthest point to all cluster centers is > 2*OPT.
b) This means that distances between all cluster centers is also > 2*OPT.
c) We have k + 1 points with distances > 2*OPT between every pair.
d) Each point has a center of the optimal solution with distance <= OPT to it.
e) There exists a pair of points with the same center X in the optimal solution (pigeonhole principle: k optimal centers, k+1 points)
f) The distance between them is at most 2*OPT (triangle inequality) which is a contradiction.
Hence, proved

Problem 2

Collaborators: List students that you have discussed problem 2 with: Vishnu Teja Narapareddy, Tulika Sureka
We will reduce the Dominating set problem to finding approximate k- center problem with approximation factor less than 2.

Construction:
From the given graph G of dominating set, we can create an instance of k-center by giving the distance between adjacent vertices as 1 and rest others as 2. We have chosen these values so as to satisfy the triangle inequality for any three vertices in the graph.

Proof:
Now, we can say that there is a dominating set if and only if the optimal radius for k-center problem is 1. If there is a dominating set D of size k, then we can choose the k-centers as the vertices in the set D, which will have a radius 1 as every vertex is at a distance of 1 from adjacent vertex and hence distance 1 away from the cluster center. Also, if there is a choice of k cluster centers S with radius 1, then by our construction as discussed above, each vertex is adjacent to $s \in S$, thus implying S is the dominating set.
The radius cluster that can be found from the given algorithm would be either 1 or 2. Now, let say algorithm A (defined in question 1) gives 1.5 approximation for k-center. If r* is the optimal cluster radius, then we can say that :

$$r* \leq r \leq 1.5r*$$

If there is a dominating set in graph G with k vertices, then r*=1. The algorithm A would guarantee:

$$1 \leq r \leq 1.5$$

Else, if the graph does not have a dominating set, then r*=2. The algorithm A would guarantee:

$$2 \leq r \leq 1.5 * 2$$

As these ranges are non-overlapping, we can say that algorithm A can correctly distinguish whether graph G has a dominating set or not, which implies there is a polynomial time algorithm for finding the dominating set. This is not possible unless P==NP, which means that every NP hard problem can be solved in polynomial time.
Hence, the existence of algorithm A would lead to polynomial time algorithm for 3-SAT.

Problem 3

Collaborators: List students that you have discussed problem 3 with: Vishnu Teja Narapareddy, Tulika Sureka
As described in the algorithm, we randomly place a vertex v in the set U with a probability of 1/2 and similarly in set W with a probability of 1/2.
For each edge e, let us define a random variable y_i which takes a value of 1 if that edge belongs to the cut and if it does not, then it's value becomes 0. Also, the probability of an edge belonging to the cut is 1/2 (as described in the algorithm given in the question). So, now we can say that size of the cut C is given by :

$$|C| = \sum_{e \in E} y_i$$

Let m be the total number of edges in the graph.
Now, we calculate the expected value of the size of the cut.

$$E[|C|] = \sum_{e \in E} E[y_i]$$
$$E[|C|] = \sum_{e \in E} Prob[e \in C]$$
$$E[|C|] = \frac{m}{2}$$

Let us say C* be the size of the maximum cut. The maximum value of C8 can be equal to number of edges in the graph i.e |E| = m
Hence, we can say that

$$|C| >= \frac{1}{2}|C*|$$

Problem 4

Collaborators: List students that you have discussed problem 4 with: None
Yes