

6.854 Advanced Algorithms

Problem Set 8

John Wang

Collaborators:

Problem 1-a: Suppose the optimum diameter d is known. Devise a greedy 2-approximation algorithm (an algorithm which gets k clusters each of diameter at most $2d$).

Solution: We start and pick an arbitrary center point x_1 . This will be center of cluster c_1 . We will place all points within a distance of d from x_1 into cluster c_1 . In other words, for all points y such that $d(x_1, y) \leq d$, we will set $y \in c_1$. Now, we will continue this procedure k times. At the end, we will have formed k clusters, each of diameter at most $2d$.

I will now show the correctness of this 2-approximation. First, we know that each point x will be inside a cluster c of diameter d . Therefore, a cluster of diameter $2d$ centered at x will completely contain cluster c as a subset. This follows because any point in cluster c must be the center of the larger cluster, which implies that in any orientation of the point, the smaller cluster will be contained in the larger cluster. This means that OPT's cluster c will be contained in the approximation algorithm's cluster. It is clear by the triangle inequality that no point in each cluster is more than a distance of $2d$ away from another point in the same cluster.

From this, we see that each one of OPT's clusters k_i will be wholly contained by the approximation algorithm's clusters c_i . Thus, each point will be accounted for once the approximation algorithm terminates. Therefore, we see that the approximation algorithm will find k clusters each of diameter at most $2d$ which will cover the entire space of points. Moreover, since we know that the optimum diameter is d , we see that this is a 2-approximation. \square

Problem 1-b: Consider an algorithm which (k times) chooses as a center the point at maximum distance from all previously chosen centers, then assigns each point to the nearest center. By relating this algorithm to the previous algorithm, show that you get a 2-approximation.

Solution: If the algorithm picks points which are at least d apart from each other, then each center will belong to a different cluster k_i in OPT. This follows because clusters in OPT have a diameter of d , which means that if two points x and y have a distance $d(x, y) > d$, then x and y cannot belong to the same cluster. Now, if all k centers are d distance apart from each other, then we have created k clusters where no point in any of the clusters is further than d from each of the centers. This means the maximum diameter of the clusters will be $2d$.

Thus, all that needs to be shown for a 2-approximation is that the algorithm will pick center points which are at least d apart from each other. Suppose this is not true. Then after the algorithm has found p centers that are d apart from each other, there will no longer be any points which are d apart from all the previously chosen centers. This, however, implies that all of the remaining points can be allocated into some of the clusters which already exist. This means that we do not need all k of the centers for OPT's algorithm, which means we can decrease the maximum diameter (in the worst case, we can set the diameter to the second larger diameter, and make a cluster for the single point which is left out). This shows that OPT does not provide the optimal diameter, which is a contradiction. Therefore, we algorithm must pick center points which are at least d apart from each other and so the algorithm is a 2-approximation. \square