Lecture 7: Asymmetric K-Center

February 5, 2007

Lecturer: Anupam Gupta *Scribe: Jeremiah Blocki*

In this lecture, we will consider the K-center problem, both in its symmetric and asymmetric variants. For this lecture, recall that a metric space (V, d) is consists of a set of Points V along with a function $d : V \times V \to \mathbb{R}^+$ which satisfies the following properties:

- $\forall x \in V, d(x, x) = 0$
- **(Triangle Inquality)** $\forall x, y, z \in V, d(x, y) + d(y, z) \geq d(x, z)$
- **(Symmetry)** $\forall x, y \in V, d(x, y) = d(y, x)$

If we drop the symmetry requirement, we get a potentially *asymmetric metric*. Such a metric may be obtained, e.g., by taking shortest paths distances in a *directed* graph.

1 Symmetric k-Center

Definition 1.1. The SYMMETRIC k-CENTER problem is a NP-Hard optimization problem whose input is a finite metric space (V, d). The desired output is a set of "centers" $F \subseteq V$ with $|F| = k$ minimizing

$$\Phi(F) = \min \max_{v \in V} d(F, v)$$

where

$$d(F, v) := \min_{f \in F} d(f, v)$$

In words, we want the distance *to* each vertex v *from* its closest center in F to be as small as possible. In the following, given an instance of k-center and a solution F, we will often refer to vertex v as being "assigned" to some center f_v — this just means f_v minimizes $d(f, v)$ over all $f \in F$, ties being broken in some fixed way.

1.1 2-Approximation Algorithm for Symmetric k-Center

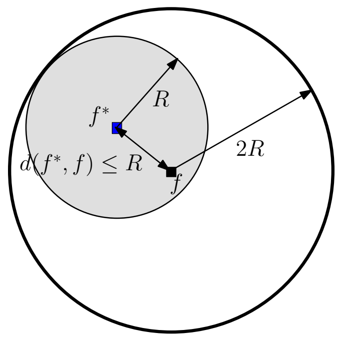
There are many (essentially identical) 2-approximations known for the k-center problem. Here is one of them: let R be our "guess" for the optimal value — we run the following procedure with our guess for R, and return the smallest value R for which this procedure returns at most k facilities.

Algorithm I (Symmetric k-center)

1. Label all vertices $v \in V$ as unmarked, and set $F = \emptyset$.
2. Pick any unmarked vertex f and add it to F.

3. Mark all vertices $v \in B(f, 2R) = \{v \in V \mid d(v, f) \leq 2R\}$.
4. if $\exists v \in V$ unmarked goto step 2, otherwise return F.

Fact 1.2. *If every vertex $v \in V$ is marked then $\Phi(F) \leq 2R$.*



Theorem 1.3. *If $R \geq$ Opt then the algorithm ends with at most k facilities (F).*

Proof. First, observe that all the points in F are at distance greater than $2R$ from each other, by construction. So, if the algorithm returned a set F with $|F| > k$, there exist $k + 1$ points in V at distance greater than $2R$ from each other. By the pegion-hole principle, two of them (say a and b) must be served by the same center f^*. But $d(f^*, a) \leq \mathsf{Opt} \leq R$ and $d(f^*, b) \leq \mathsf{Opt} \leq R$, so by symmetry and the triangle inequality, $d(a, b) \leq d(a, f^*) + d(b, f^*) \leq 2R$, which violates the fact that $d(a, b) > 2R$.

Alternative Proof. A different way to see this is the following. Let us fix some optimal solution F^*. Now consider the first center f opened by the algorithm. By the property of the optimum, there exists $f^* \in F^*$ with $d(f^*, f) \leq \mathsf{Opt} \leq R$. Moreover, any client v assigned to f^* in the optimal solution is at distance $d(f, v) \leq d(f, f^*) + d(f^*, v) = d(f^*, f) + d(f^*, v) \leq 2R$ from f, and hence will be marked upon the algorithm's opening f. (Informally speaking, center f "captures" some optimal center f^* and marks all the clients that f^* served in the optimum.) Hence, after the first round, the yet-unmarked certices have a solution with $k - 1$ centers (at distance at most R), and inductively our algorithm will find such a solution. □

In Homework 2, you showed that this problem is hard to approximate to a factor of $2 - \epsilon$ for any $\epsilon > 0$, and hence this result is optimal!

2 Asymmetric k-Center

We now consider the question of relaxing the assumptions under which the above result was obtained: it is easy to see that if we do not have the triangle inequality, nothing better than an $O(\log n)$ approximation can be obtained. It is natural to ask the question: *what if we relax the assumption of symmetry?*

Definition 2.1. ASYMMETRIC k-CENTER is a NP- Hard Optimization problem whose input is an asymmetric metric (V, d) (i.e., it satisfies the properties of a metric, except for symmetry). We want to output a set of k centers $F \subseteq V, |F| = k$ minimizing

$$\Phi(F) = \min \max_{v \in V} d(F, v).$$

Remark 2.2. Note that the 2-approximation algorithm no longer works because we relied on symmetry to prove correctness: now if $f \in F$ is at distance R from $f^* \in F^*$, and $v \in V$ is also at distance R from f^*, this tells us nothing about $d(v, f)$ or $d(f, v)$.

In the rest of the argument, we again assume that we have guessed the optimal distance R, and just want to find a solution with $\Phi(F) \leq \alpha R$ for some small $\alpha \geq 1$. Clearly, we can run the algorithm for every possible value of R in the set $\{d(x, y) \mid x, y \in V\}$ and return the smallest value of R for which we successfully find a solution.

2.1 Some Notation

Definition 2.3. If $u, v \in V$ such that $d(u, v) \leq R$ we say that u *covers* v, or u 1-*covers* v. If u i-covers v and v j-covers w, then u $(i + j)$-covers w. Note that each node covers itself.

Using this notion, we can now simplify our representation for Asymmetric k-Center using a digraph. Construct $D = (V, A)$ with nodes V and arcs A, where $(v_1, v_2) \in A \leftrightarrow d(v_1, v_2) \leq R$ or equivalently if v_1 covers v_2. The goal is to find k nodes F such that all nodes in V are covered by some node in F.

Definition 2.4. If $D = (V, E)$ is a digraph and $v \in V$ then

1. $\Gamma_t^+(v) = \{x \in V : Dist(v, x) \leq t\}$
2. $\Gamma_t^-(v) = \{x \in V : Dist(x, v) \leq t\}$

where $Dist(x_1, x_2)$ is the length of the shortest path from x_1 to x_2 in D.

2.2 Why does the Symmetric Algorithm not work any more?

Consider the alternative view of the proof of Theorem 1.3. We said that since the center f was covered by some optimal center f^*, by symmetry f covered f^* as well. Hence it 2-covered everything that f^* covered. Note that while we do not have symmetry, if we could find a vertex f that covered everyone who covered f, this argument would still hold. Hence, consider the following definition:

Definition 2.5 (CCV's)**.** A vertex $v \in V$ is said to be a *Center Capturing Vertex* (CCV) if $\Gamma_t^-(v) \subseteq \Gamma_t^+(v)$. In words, v is a CCV if whenever some x covers v, v also covers x.

Note that in the symmetric case, every vertex was a CCV, then the 2-Approximation Algorithm for Symmetric k-Center would still work. In fact, if the following algorithm terminated with all vertices marked, we would have a 2-approximation.

Algorithm II (won't quite work)

1. Label all vertices $v \in V$ as unmarked, and set $C = \emptyset$.
2. Pick any unmarked CCV f and add it to C.
3. Mark all vertices v 2-covered by f.
4. if $\exists v \in V$ unmarked goto step 2, otherwise return C.

However, now there is no guarantee that Step 2 can be executed, since there may not be any unmarked CCVs in the remaining graph, even though there may be many unmarked vertices remaining. What then? We're stuck.

2.3 A Recursive Greedy Algorithm For Asymmetric k-Center

Let us try a different approach to the problem. Recall the theorem proved in Homework 1.

Theorem 2.6. *The greedy algorithm is a* $(1 + \lceil \ln(\frac{n}{\mathsf{Opt}}) \rceil)$*-approximation algorithm for unweighted set cover.*

We use this greedy set cover algorithm as the basis for the following algorithm, originally proposed in [PV98]. This algorithm seeks to cover vertices in some subset $X \subseteq V$: to cover all the vertices in V, we would invoke $\textsf{Recursive Cover}_R(V, k)$

Algorithm $\textsf{Recursive Cover}_R(X, k)$:

1. Set $F_0 = X$, $i = 0$, say *a covers b if* $d(a, b) \leq R$.
2. For each $v \in V$, let $S_v = \{z \in F_i : v \text{ covers } z\}$.
3. Run the greedy Set Cover Algorithm with sets $\{S_v\}_{v \in V}$ to get a set F_{i+1} of points that cover F_i.
4. Set $i = i + 1$.
5. If $|F_i| \leq 2k$, output F_i, else goto Step 2.

2.3.1 Analysis of the Recursive Cover Algorithm

Recall that $\log^{(0)} x = x$, and $\log^{(i)} x = \log \log^{(i-1)} x$. Moreover, $\log^* x = put formal def here$.

Some claim here that $\log^*_{3/2} x = \ln^* x + O(1)$.

Fact 2.7. $\log_{3/2} x \geq 1 + \ln x$ *whenever* $x \geq 2$.

Claim 2.8. *If* $|X| \geq 2k$*, the cardinality of any* F_i *found by the algorithm is at most* $k \cdot \log^{(i)}_{3/2} |X|$.

Proof. The proof is by induction. The base case is trivial, since $|F_0| = |X|$. Recall that, by the property of the optimal solution for the asymmetric k-center problem and our choice of R, there are k sets $\{S_x\}_{x \in \mathsf{Opt}}$ that cover each F_i.

For the inductive step, note that

$$\begin{aligned}
|F_i| &\le k(1+\ln(\frac{|F_{i-1}|}{k})) \\
&\le k\log_{3/2}(\frac{|F_{i-1}|}{k}) && \text{(by Fact 2.7)} \\
&\le k\log_{3/2}(\log_{3/2}^{(i-1)}|X|) && \text{(by I.H.)} \\
&\le k\log_{3/2}^{(i)}|X|.
\end{aligned}$$

This proves the claim. □

Proposition 2.9. *When $X = V$, the algorithm above stops after at most $t = \log^* n + O(1)$ rounds, we are left with $|F_t| = k_t \le 2k$ vertices. Moreover, this set F_t t-covers the vertex set V.*

Unfortunately, this algorithm is not an $O(\log^* n)$-approximation algorithm to the Asymmetric k-Center problem because of its failure to ensure that $k_t \le k$. It is only a *bicriteria* algorithm: it uses more centers than allowed ($2k$ instead of k) and returns a solution F with $\Phi(F)$ worse than $\mathsf{Opt} = \Phi(F^*)$.

2.3.2 Yet Another Failed Idea

Suppose we had an instance with the following property:

> *Not only did k centers cover all the vertices, but $k/2$ centers (say) 100-covered V.*

In this case, we could set $k' = k/2$, redefine coverage (now, a "covers" b if $d(a,b) \le 100 \cdot R$) and run the algorithm RECURSIVE-COVER$_{100\,R}(V, k')$. It would return a set of at most $2k' = k$ vertices F' that would $(\log^* + O(1))$-"cover" V, and hence have $d(F', v) \le 100 \cdot (\log^* + O(1)) \cdot R$ for all $v \in V$.

However, not all instances have this property. A moment's thought allows us to produce a counter-example: If $|V| = k$ and $d(x,y) = 1$ for all $x \ne y \in V$. Then k vertices cover V at distance 0, but no set of $k/2$ vertices can cover V at distance $c \cdot 0$ for any $c \in \mathbb{Z}$.

As an aside, note that in this counter example all vertices are center capturing. In fact, one can show that if there are no CCVs, then the property indeed holds: a variant of this property (given in the next section) allows us to get the final algorithm.

2.4 Bringing it all Together

Here's a new idea for an algorithm: While there are unmarked Center Capturing Vertices v, add v to F and mark all vertices 2-Covered by v. Stop when there are no more Center Capturing Vertices. <u>Now</u> *run the recursive greedy algorithm to cover the unmarked vertices!*

> Big questions: Does this algorithm work? Can we show that the remaining problem, after we find no CCVs, has the property desired in the previous section?
>
> Answers: yes (if we change the definition of marked slightly), and almost.

Let us begin by giving the algorithm, which combines ideas from all the previous sections.

Final AkC Algorithm

1. Label all vertices $v \in V$ as unmarked, and set $C = \emptyset$.
2. Pick any unmarked CCV f and add it to C; if none exist goto step 5.
3. Mark all vertices v 2-covered by f.
4. if $\exists v \in V$ unmarked goto step 2, otherwise return F.
5. Also mark all vertices 4-covered by the set of CCVs C. Let U be set of unmarked vertices.
6. Invoke RECURSIVE-COVER$_{5R}(U, \frac{1}{2}(k - |C|))$ to obtain $(k - |C|)$ new centers F covering U at distance $5R$.
7. Return $C \cup F$.

2.4.1 Analysis

To show that this algorithm is well-defined, we need to show that there *exists* a feasible set of centers $\widehat{F}$ that covers all of U at distance at most $5R$, and that $|\widehat{F}| \leq \frac{1}{2}(k - |C|)$. That will ensure that the recursive greedy algorithm will find a set of size at most $k - |C|$, as desired.

Lemma 2.10. *Suppose there are no CCVs in the set of vertices not 2-covered by C. Then U (the set of vertices not 4-covered by C) can be 5-covered with at most $\frac{k-|C|}{2}$ centers.*

Proof. Note that OPT covers U using at most $k - |C|$ centers (those that are not covered by the CCVs in C). Call this set of optimal centers C'. We now build a new digraph $D' = (C', A')$ on these centers, by adding the directed arc (u, v) for $u, v \in C'$ iff u 2-Covers v.

We claim each $x^* \in C'$ has in-degree at least 1. Consider a center x^*: since it is not center-capturing, there just be some $y \in V$ such that y covers x^* but x^* does not cover y. Now y must be covered by some $z^* \in F^*$: this z^* cannot be x^* (since x^* does not cover y). Moreover y is too far from C for some optimal center not in C' to cover it, and hence z^* is also in C'. Now (z^*, x^*) is an arc in A', and any vertices in U covered by x^* are 3-covered by z^*.

Finally, it remains to find a set of $\frac{k-|C|}{2}$ of these vertices in C' that 5-cover all of U. Since each vertex in this digraph D' has in-degree ≥ 1, our result follows from the next claim.

Claim 2.11. *Let $D = (X, A)$ be a digraph. Then $\exists W \subseteq X, |W| \leq |X|/2$ such that for every vertex $v \in X$ with in-degree ≥ 1, it is reachable in 2 hops from W.*

Proof. By induction on $|X|$. The base case $|X| = 2$ is trivial. For the inductive step, pick u with non-zero out-degree. Remove u and its neighbors to obtain X'. Note that $|X'| \leq |C| - 2$. By induction we can cover the vertices X' with non-zero in-degree using $|X'|/2$ vertices . The only vertices in X which might not be covered are u, u's neighbors and vertices at 2 hops from u (since their in-degree might have gone to zero). Hence, adding u to the set W' produces a set W of size at most $|X'|/2 + 1 \leq |X|/2$ such that every vertex in X is reachable in two hops from W. □

Applying the above claim with on the digraph $D' = (C', A')$ implies there are at most $|C'|/2 \leq (k - |C|)/2$ vertices that can reach all of C' in 2 hops, hence 4-cover all of C', and thus 5-cover all of U. □

3 LP formulations of Asymmetric k-Center

Consider the following LP formulation of the problem (similar to that used for k-median).

$$\begin{aligned} &\min \lambda \\ \text{s.t. } &\textstyle\sum_i y_i \leq k \\ &x_{i,j} \leq y_i \\ &\textstyle\sum_i x_{i,j} \geq 1 \\ &\textstyle\sum_i d_{i,j}\, x_{i,j} \leq \lambda \end{aligned} \tag{LP1}$$

Intuitively, y_i represents the facilities fractionally chosen. $x_{i,j}$ is how much the center v_j is fractionally covered by y_i. λ is the maximum average distance to any v_j fractionally covered by the v_i's.

Proposition 3.1. *The LP $(LP1)$ has an integrality gap of k even for the symmetric k-center problem.*

Proof. Consider an example with $k+1$ vertices and symmetric distances $d(i,j) = 1$. Clearly, any solution with k centers has $R = 1$. But we can fractionally set $y_i = \frac{k}{k+1}$ to obtain an LP solution with $\lambda = \frac{1}{k+1}$ □

3.1 A better LP: Using Thresholding

We can obtain a better LP solution using the idea of guessing the optimum value R, and setting $x_{ij} = 0$ for $d_{ij} = R$. In fact, we seek the smallest value R for which the following LP is feasible.

$$\begin{aligned} &\textstyle\sum_i y_i \leq k \\ &x_{i,j} \leq y_i \\ &\textstyle\sum_{i: d_{ij} \leq R} x_{i,j} \geq 1 \end{aligned} \tag{LP2}$$

This is equivalent to the following LP:

$$\begin{aligned} &\textstyle\sum_i y_i \leq k \\ &\textstyle\sum_{i \text{ covers } j} y_i \geq 1 \end{aligned} \tag{LP2'}$$

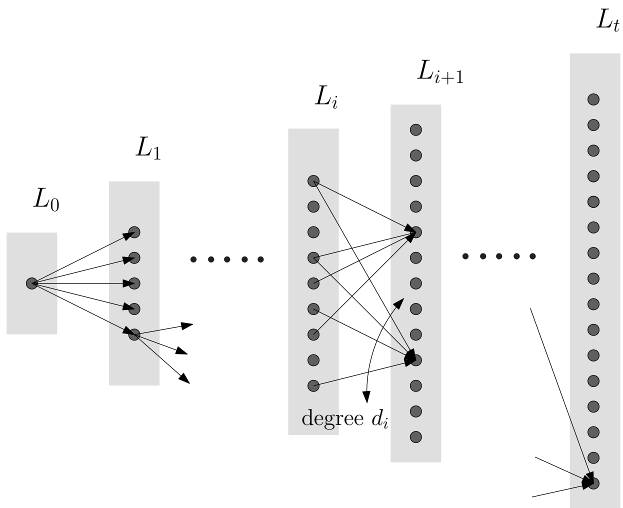
In the symmetric case, it can be shown that this LP has a constant integrality gap. Aarcher [Arc01] showed that the in the asymmetric case, the integrality gap was at most $O(\log^* k)$, which is an improvement on the $O(\log^* n)$ result in the previous section. The example showing this integrality gap was tight was given in [CGH$^+$05]. (They also showed a matching hardness result, which we will do in the coming lectures.)

3.2 Integrality Gap

Theorem 3.2. *The integrality gap of this LP is* $\Omega(\log^* n)$.

The gap instance is a layered asymmetric k-center instance as in the picture, with $t+1$ layers. The layer L_0 has a single vertex, whereas L_i has n_i vertices. For every vertex $v \in L_i$, we independently and uniformly at random choose a set of d_{i-1} vertices $u_1, \ldots, u_{d_{i-1}}$ from L_{i-1} and say that these are the (only) vertices that cover v. The parameters are set as follows:

- The degrees: $d_1 = t^2$, and $d_i = 2^{(d_{i-1})^3}$.
- The sizes of layers: $n_i = d_i d_t$
- The number of centers: $k = t d_t$



Proposition 3.3. *Based on the definition of the degrees* d_i*, it can be shown that*

$$t \geq \log^* n - \log^*(\log^* n) - \theta(1).$$

Proposition 3.4. *There is a feasible fractional solution to* $(LP2')$.

Proof. For $v \in L_0$, set $y_v = 1$. For $1 \leq 1 \leq t-1$ and $v \in L_i$, set $y_v = \frac{1}{d_i}$. Since the in-degree of each node in L_{i+1} is d_i, this will ensure that the solution satisfies the fractional coverage constraints. Also

$$1 + \sum_{i=1}^{t-1} n_i \times \frac{1}{d_i} = 1 + \sum_i \frac{d_i d_t}{d_i} = 1 + (t-1)d_t \leq k.$$

□

Lemma 3.5. *For any fixed set* $C \subseteq V \backslash L_0$ *of k-centers, with high probability there exists a vertex* $v \in L_t$ *which is not* t*-covered by* C.

The proof considers the set Y_i be the set of vertices which cannot be reached from C. By definition, $|Y_1| = n_1 - k \geq 2d_t$. Now using the random construction of the in-arcs into each node, one shows that the expected size of each Y_i is large: moreover, the independence allows us to use a Chernoff bound to show that this each $|Y_i| \geq 2d_t$ with very high probability. Now using the above lemma, and taking a union bound over all $\binom{n_1}{k}$ choices of C (since C can be assumed to all lie in L_1) gives us the following lemma, and the large integrality gap.

Lemma 3.6. *The probability that there exists a set $C \subseteq L_1$ of k centers that t cover L_t is very small. Hence,* $\mathsf{Opt} = t + 1$ *with high probability.*

The details of this integrality gap proof can be found, e.g., in [HKK03]. In the upcoming two lectures, we will show the $\Omega(\log^* n)$ hardness for this problem.

References

[Arc01] A. Archer. Two O (log* k)-Approximation Algorithms for the Asymmetric kCenter Problem. *Integer Programming and Combinatorial Optimization: 8th International IPCO Conference, Utrecht, the Netherlands, June 13-15, 2001: Proceedings*, 2001.

[CGH+05] J. Chuzhoy, S. Guha, E. Halperin, S. Khanna, G. Kortsarz, R. Krauthgamer, and J.S. Naor. Asymmetric k-center is log* n-hard to approximate. *Journal of the ACM (JACM)*, 52(4):538–551, 2005.

[HKK03] Eran Halperin, Guy Kortsarz, and Robert Krauthgamer. Tight lower bounds for the asymmetric k-center problem. *Electronic Colloquium on Computational Complexity (ECCC)*, 10(035), 2003.

[PV98] R. Panigrahy and S. Vishwanathan. An O(log* n) Approximation Algorithm for the Asymmetric p-Center Problem. *J. Algorithms*, 27(2):259–268, 1998.