2-10-2021

On Mentzer’s Hardness of the k-Center Problem on the Euclidean Plane.

Raymond Chen
Raymond.G.Chen.22@Dartmouth.edu

On Mentzer's Hardness of the k-Center Problem on the Euclidean Plane.

Raymond Chen

Dartmouth Computer Science Technical Report TR2021-1001
February 10, 2021

Abstract

An instance of the *k-center* problem consists of n points in a metric space along with a positive integer k. The goal is to find the smallest radius r such that there exists a subset of k centers picked among them such that every point is within distance r of at least one center.

Stuart Mentzer (Mentzer, 1988) wrote a paper showing that in the Euclidean plane, it is NP-Hard to approximate this problem up to a factor of $\sqrt{2+\sqrt{3}} \approx 1.93$. However, his report is missing some details. In this note, we present details of his full construction.

1 Introduction

In the k-center problem, we are given a set of points $P = \{p_1, p_2, ...p_n\}$ and a parameter k. We seek to choose a subset $F = \{f_1, f_2, ..., f_k\} \subseteq P$ that minimizes $\max_{p_i \in P} \min_{f_j \in F} d(p_i, f_j)$ where $d(p_i, f_j)$ is the metric distance between the points. The Euclidean k-center problem is when all points are on the plane and the distance metric is the Euclidean distance.

In general metrics, it is NP-hard to achieve a better than 2-approximation of the k-center problem (Gonzalez, 1985). But this fact is not known in the Euclidean Plane. The best result is due to Mentzer (Mentzer, 1988), which shows a hardness of factor $\sqrt{2+\sqrt{3}} \approx 1.93$. A better than 2-approximation is currently unknown for the Euclidean plane. The k-supplier problem is a generalization of the k-center problem, which has a set of facilities B that can be chosen as centers and a set of clients K that need to be covered. The k-center problem is an instance of the k-supplier problem where all points are facilities and clients. The best known result in the k-supplier problem on the Euclidean plane is an approximation factor of $1 + \sqrt{3} \approx 2.73$ (Nagarajan et al., 2013) and best known hardness is 2.645 (Feder and Greene, 1988).

Stuart Mentzer (Mentzer, 1988) wrote a paper showing that in the Euclidean plane, it is NP-Hard to approximate this problem up to a factor of $\sqrt{2+\sqrt{3}} \approx 1.93$. However, his report is missing some details. In this note, we present details of his full construction.

In the 3-SAT problem, we are given a set of variables $V = \{v_1, v_2, ..., v_n\}$ which take on True or False values, and a set of clauses $C = \{c_1, c_2, ..., c_m\}$. Each clause is a conjunction of three variables, either negated or unnegated. An example of such a clause is $c_1 = (v_1 \vee \neg v_2 \vee \neg v_3)$ which is true if v_1 is true, v_2 is false or v_3 is false. We seek to determine if there exists an assignment of True or False values to the variables such that all clauses are satisfied.

For every instance of 3-SAT, Φ, a graph version, G_Φ, can be constructed as follows. For every variable and clause in Φ, a vertex is introduced to G_Φ. And for each clause in Φ, three new edges are introduced, one to each variable it is associated with in G_Φ. If a variable is unnegated in the clause, we call the edge between the two *unnegated*. Otherwise, we call it *negated*. G_Φ is a P3SAT instance if and only if G_Φ is planar. Figure 1 shows an example of an instance of P3SAT. P3SAT is also NP-complete (Lichtenstein, 1982).

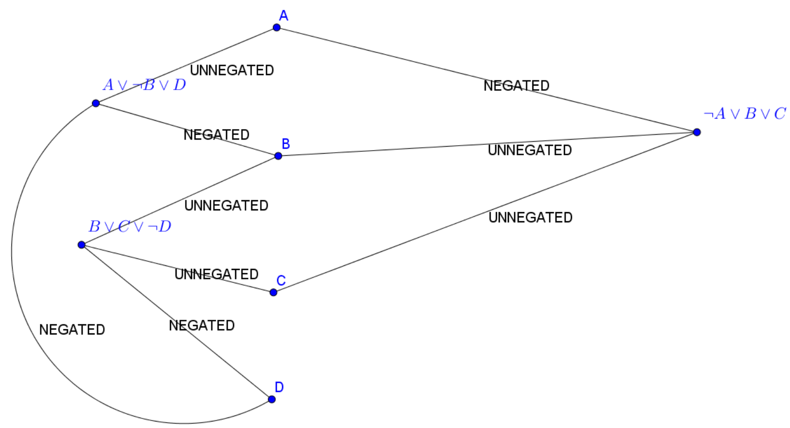

Figure 1: P3SAT Instance for $(\neg A \vee B \vee C) \wedge (A \vee \neg B \vee \neg D) \wedge (B \vee \neg C \vee \neg D)$

2 Reduction of Euclidean k-center from P3SAT

Given a P3SAT instance (Φ, G_Φ) with n variables, m clauses, we construct a set of points X_Φ on the Euclidean plane and a value k. Let d_i be the degree of variable i, that is the number of clauses that it is a part of either as itself or as its negation. Let $\text{Opt}(X_\Phi)$ be the minimum radius such that k centers can cover X_Φ. We will prove that if Φ is satisfiable then $\text{Opt}(X_\Phi) = 1$. Otherwise, if it is not satisfiable, then $\text{Opt}(X_\Phi) = \sqrt{2+\sqrt{3}}$. This will show that k-centers is NP-hard to approximate better than a factor of $\sqrt{2+\sqrt{3}}$ because $P3SAT$ is NP-hard.

In the reduction we will replace variables with cycles of points, edges with a chain of points, clauses with points, and will label points as N, T, F, C as a mnemonic. T, F labels will imply True and False assignments, C corresponds to clause points and N labels are points that should never be chosen as centers. The following **Truss** structure forms the backbone for the reduction. Starting with an equilateral triangle F, T, N, we seek to insert points N_F, N_T where $d(N_F, F) =$

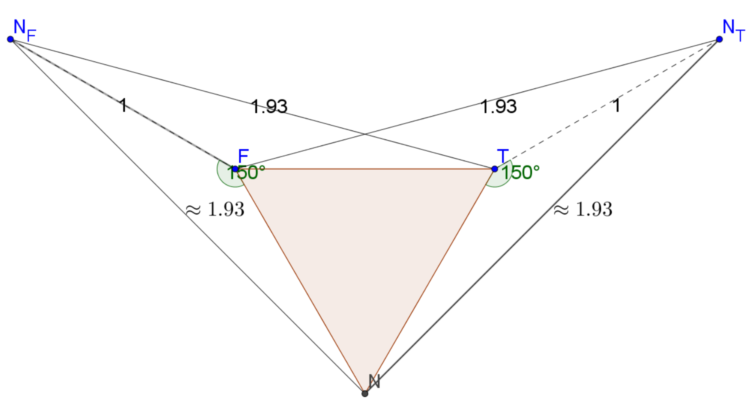

Figure 2: Truss

$1, d(N_T, T) = 1$ in such a way to maximize $\min(d(N_F, T), d(N_T, F), d(N_F, N), d(N_T, N))$. This builds the Truss in Figure 2, which has a distance of $\sqrt{2+\sqrt{3}}$ for any of the above pairs. In all figures, explicitly drawn lines are of length $\sqrt{2+\sqrt{3}}$. Trusses that connect variables to edges will be called **joints** and those that 'reinforce' variables and edges will be called **knees**. They will be formally defined below.

Definition 2.1. *For each* ***clause*** c_i*, we introduce a single point in the plane. These clause points will be connected to the variable cycles by an edge structure.*

Clause points are put far enough apart so that no two edges are within distance 2 units from one another at any pair of points. In order to ensure this condition, each clause should be at least 4 apart from one another.

Definition 2.2. *A* ***cycle*** *is defined by two terminal points. Between these two points, the cycle will have length* $\approx J$*, with* $J-2$ *intermediate points. These intermediate points will have pairwise distance* 1*, while non-adjacent points will have distance* $\approx 2-\epsilon$ *for* ϵ *as small as you would like. A* ***closed cycle*** *is a cycle with terminal points that are the same. (Same as in Feder and Greene, 1988)*

Definition 2.3. *For every* ***variable*** v_i*, starting with a knee Truss, we create a closed cycle of* $3L$ *points, with one adjacent* ***knee*** *point, which comes from the Truss. The distance between any two*

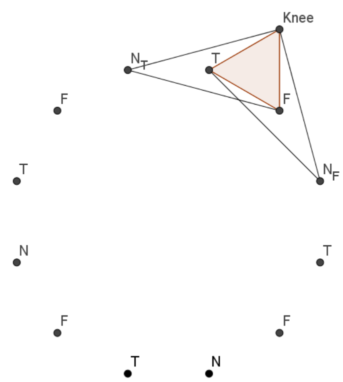

Figure 3: Example variable cycle with (12 + 1) points

adjacent points in this cycle will be exactly 1. And for any pair of points which have one intermediate point between them, the distance will be at least $\sqrt{2+\sqrt{3}}$*. There will be exactly* $2(d_i+1)$ *such pairs with distance exactly* $\sqrt{2+\sqrt{3}}$*. We can construct such a figure by increasing* L*, so that distances between all other non-adjacent vertices is close to 2. At a minimum,* L *needs to be* $> 3d_i+3$ *for every variable* i *so that each variable truss does not intersect. As such, where* m *is the number of clauses, a sufficient condition is* $L \geq 4m$ *(see Figure 3). Assign variables a label in the order* N, T, F *clockwise.*

We now highlight which non-adjacent vertices have distance exactly $\sqrt{2+\sqrt{3}}$*. For each edge that the variable is part of in P3SAT instance, we introduce a joint truss to the cycle that corresponds to the clause. For an unnegated edge, we create the truss with an* N, T *pair as base of equilateral triangle. For a negated edge, we create the truss with a* F, N *pair as base of equilateral triangle. These trusses must be made so that no points of separate trusses are within distance* 2*. See Figure 4 for an example variable with two outgoing edges.*

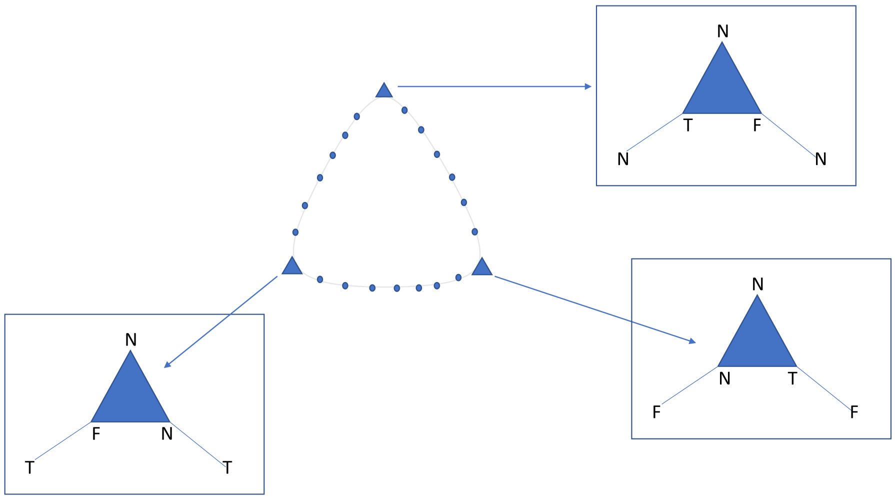

Figure 4: Example variable cycle with (12 + 3) points and part of an unnegated edge, negated edge

Observation 2.4. *No center covers more than three of the variable cycle points, and only centers inside the variable cycle cover exactly three. Further, the only points outside of the cycle that can cover variable points come from joints and knees.*

Observation 2.5. *Only the base T and F and the knee truss itself cover the knee point of a variable cycle with distance $< \sqrt{2+\sqrt{3}}$. And they cover with distance 1.*

Definition 2.6. *The following describes the **edge** between a clause and the "unnegated literal" variable cycle. We begin by inserting a truss between the clause point and variable cycle as in Figure 5. We connect this knee truss to the joint truss (with N_v, T_v as base and N_1 as the joint) that corresponds to the clause to as follows:*

We create points F_1, T_1 by extending the perpendicular bisector of N_v, T_v through N_1 at distances 1 apart. Then insert the point N_F of the truss such that it forms a $150°$ angle with F_1, T_v.

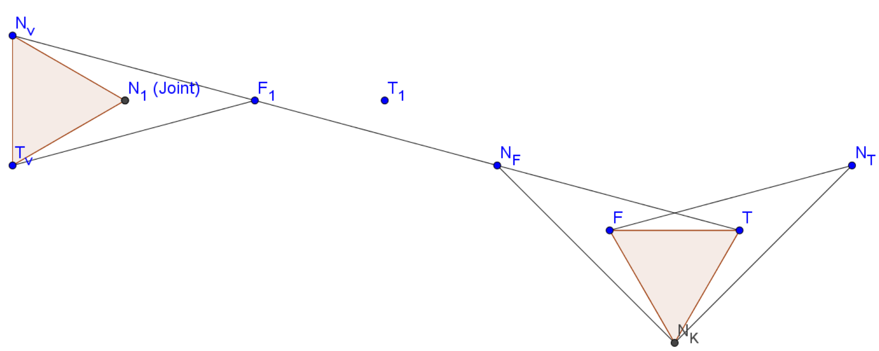

Figure 5: Truss connected to an unnegated variable

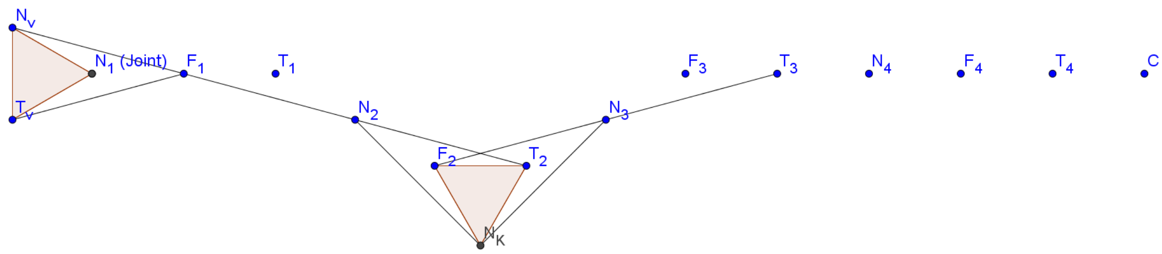

Figure 6: Example of a full unnegated edge

We can then connect the Variable + Truss to a clause as follows: We create points F_3, T_3 such they are parallel to N_1, F_1, T_1 and form a $150°$ degree angle with the truss point N_T. Then add triplets in the order N, F, T along this line until you reach C.

A negated literal edge is created in the same manner, but instead starting with an F, N, pair in the variable cycle. Just swap all instances of F and T afterwards.

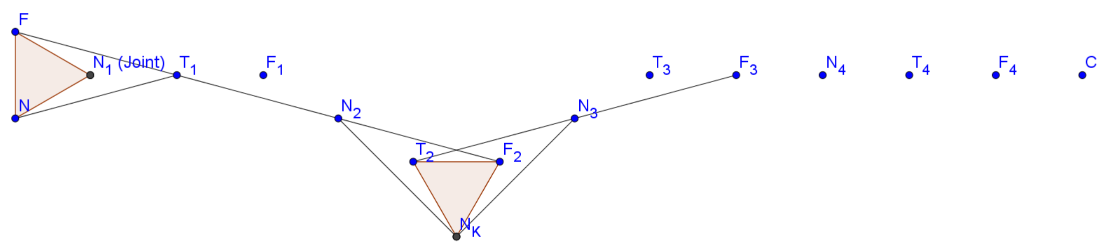

Figure 7: Example of a full negated edge

Given an edge $e \in P3SAT$, let us call the set of edge points in the reduction associated with it E_e. It consists of all points that were introduced from connecting the variable to the clause, except for the joint N_1.

Observation 2.7. *For every edge $e \in P3SAT$, $|E_e| = 3e_j$ for some integer $e_j > 2$, and all centers cover at most three points in $E_e \backslash N_K$, with only centers in $E_e \backslash N_K$ covering exactly three. Moreover, the only points that can be covered by external centers are the two points on the ends. For an unnegated edge, F_1, T_e. For a negated edge, T_1, F_e.*

An example of the reduction for the clause $A \vee \neg B \vee \neg C$ is shown in Figure 8.

Now that we've constructed the reduction, let us count how many points are in each figure.

Let $3L + 1$ is the number of points in the variable cycle for variable i. Then the total number of variable points is $\Sigma_{i=1}^{n}(3L + 1 + d_i)$. Which is composed of exactly n knee points, $\Sigma_{i=1}^{n} 3L$ cycle points, and $\Sigma_{i=1}^{n} d_i$ joints. Since $\Sigma_{i=1}^{n} d_i = 3m$, the total number of variable points is $3m + n + 3L \cdot n$.

And if each edge e has $3e_j$ points, then the total number of edge points is $\Sigma_{e \in G_\Phi} 3e_j = 3\Sigma_{e \in G_\Phi} e_j$.

There are also m clause points.

So the total number of points is $4m + n(3L + 1) + 3\Sigma_{e \in G_\Phi} e_j$.

We will see that we require exactly $k = \Sigma_{i=1}^{n} n_i + \Sigma_{e \in G_\Phi} e_j$ centers to cover the reduction if and only if it is satisfiable.

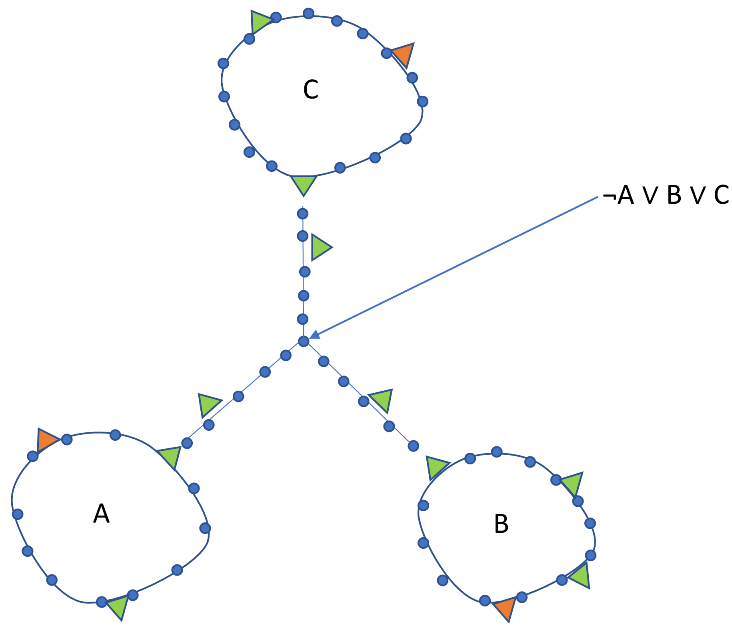

Figure 8: Reduction for $A \vee \neg B \vee \neg C$ from Figure 1

3 Structural Claims

Let C be an assignment of k-centers that cover X_Φ with radius $< \sqrt{2+\sqrt{3}}$.

Lemma 3.1. *For any variable cycle i, $|C \cap i| \geq L$. In particular, if $|C \cap i| = L$, need to open exactly L number of T centers or exactly L number of F centers.*

Proof. Note that this set of cycle points and joints are the only centers that can cover points in the variable cycles, by Observation 2.4. By Observation 2.5, because each center covers at most 3 non-knee cycle points, one must open at least L centers to cover these points.

In order to cover the $3n_i$ non-knee cycle points using just L centers, each center must cover exactly 3 non-knee cycle points. This means that the L centers must be all type N, all type F, or all type T points. Otherwise, there exists some overlap between centers, so one center covers at most 2 unique non-knee cycle points. By Observation 2.5, these centers must then be all T or all F, since these are the only centers that can cover the knee point. □

If a variable i is covered by exactly L centers in the reduction, then by Lemma 3.1 it corresponds to a consistent assignment of T or F. Let this correspond to the assignment of T or F to the variable in the $P3SAT$ instance. Reciprocally, if a variable in the $P3SAT$ instance has a true or false assignment, then it corresponds to a consistent assignment of centers of the same truth value.

Lemma 3.2. *For any variable cycle v_i, if $|C \cap v_i| = L$, then joints are covered by variable centers if and only if their negation is consistent with the variable assignment in Φ.*

Proof. If $|C \cap v_i| = L$ then by Lemma 3.1, they must all be labeled T or all labeled F. If they are all labeled with T, then edge joints are covered if and only if they are unnegated, since negated joints are exactly $\sqrt{2+\sqrt{3}}$ from the nearest T center. Similarly, if they are all labeled F, edge joints are covered if and only if they are negated. □

Lemma 3.3. *For an edge E_e with $3e_j$ points, $|C \cap E| \geq e_j$.*

Proof. The following proof will be for an unnegated edge, analagous results hold for a negated edge.

From Observation 2.5, the only points in the edge E that can be covered by external centers are the two outermost points, F_1, T_e. Let us suppose that in the worst case, these are covered by the variable joint N_1 and clause point C centers. Then we know that $C \cap E$ must contain at least two points from the set $\{T_1, N_2, F_2, T_2, N_K\}$ in order to cover T_1 and N_K. These 2 centers cover at most 6 points total.

In which case there remain $3e_j - 8$ points to be covered, and each remaining center can cover at most 3 points. So it takes at least $e_j - 2$ centers to cover the remaining points. Thus, requiring at least e_j centers to cover the edge. □

Lemma 3.4. *If the joint of an edge E_e is covered by the variable cycle, then opening e_j centers in the edge will cover the clause point. Otherwise, opening e_j centers will cover the joint but not the clause point.*

Proof. Once again, the proof will be for an unnegated edge, analagous results hold for a negated edge. If the joint is covered by the variable cycle, open e_j centers are T labeled points. Otherwise, there exists a center which only covers at most two non-knee points. Then the clause will be covered by T_e.

Otherwise, if the joint is not covered, open e centers at F labeled points, since a center needs to cover the joint. In which case, the clause point is not covered. □

4 Main Theorem

Theorem 1. *If (Φ, G_Φ) is satisfiable then OPT(X) $= 1$. Otherwise, if (Φ, G_Φ) is unsatisfiable then OPT(X) $\geq \sqrt{2+\sqrt{3}}$.*

Theorem 2. *If a P3SAT instance is satisfiable then one can cover its reduction in exactly k centers.*

Proof. Assign T and F labels to variables according to their assignments in Φ. So each variable cycle is covered in L centers. And by Lemma 3.4 for any joint that is not covered by a variable cycle, its edge E_e can be covered by e_j centers such that they cover E_e and the joint, but not the clause. Because it is satisfiable, for each clause, there exists a variable cycle who covers the corresponding joint point. By Lemma 3.4, the clause point is then covered by that edge. So for each variable cycle v_i, $|C \cap v_i| = L$, and for each edge E_e, $|C \cap E_e| = e_j$. And every knee, joint and clause point is covered by these centers. So $|C| = k$. □

Theorem 3. *If a reduction can be covered in exactly k centers for radius $< \sqrt{2+\sqrt{3}}$, then the P3SAT instance is satisfiable.*

Proof. Assign a budget of N to each variable cycle v_i and e to each edge E_e. Let our optimal assignment be C. By Lemma 3.1, $|C \cap v_i| = L$, and $|C \cap E_e| = e_j$. So $|C| \geq k$. If $|C| = k$, then it must be true that no cycle or edge can exceed its budget.

So by Lemma 3.1, there must exist a consistent assignment of T or F to each variable cycle v_i, let us assign these to our P3SAT instance Φ. By Lemma 3.2, joints are covered by the variable cycle if and only if their truth value is consistent with the variable assignment. So by Lemma 3.4, for every clause, there must exist a joint which is covered by its variable cycle. Which is true if and only if the variable cycle assignment is consistent with the clause in the P3SAT instance. Thus, for this to hold for all clauses, it must be true that Φ is satisfied by the assignment. $\square$

5 Acknowledgements

Thanks to the James O. Freedman Presidential Scholarship for financial support. Also thanks to Professor Deeparnab Chakrabarty for his help with this paper and advising me as part of the Dartmouth Undergraduate Advising and Research.

References

Feder, T., & Greene, D. (1988). Optimal algorithms for approximate clustering, In *Proceedings of the twentieth annual acm symposium on theory of computing*, 434–444. Chicago, Illinois, USA, Association for Computing Machinery. https://doi.org/10.1145/62212.62255

Gonzalez, T. F. (1985). Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, *38*, 293–306. https://doi.org/https://doi.org/10.1016/0304-3975(85)90224-5

Lichtenstein, D. (1982). Planar formulae and their uses. *SIAM Journal on Computing*, *11*(2), 329–343. https://doi.org/10.1137/0211025

Mentzer, S. (1988). Approximability of metric clustering problems. *Manuscript*. https://www.academia.edu/23251714/Approximability_of_Metric_Clustering_Problems

Nagarajan, V., Schieber, B., & Shachnai, H. (2013). The Euclidean k-supplier problem (M. Goemans & J. Correa, Eds.). In M. Goemans & J. Correa (Eds.), *Integer programming and combinatorial optimization*, Berlin, Heidelberg, 290 - 301. Springer Berlin Heidelberg.