On Maximum-Weight Minimum Spanning Tree Color-Spanning Set

Homa Ataei Kachooei* Mansoor Davoodi† Dena Tayebi ‡

Abstract

Given n points with m colors in the plane, we aim at finding m points with distinct color such that their minimum spanning tree is maximized. In this paper, we study a very simple case of this problem, that is, there are at most two points of each color whose distance is unit and have the same $x-$ or $y-$ coordinate. We prove that this problem even under these restrictions is NP-hard and does not have an FPTAS unless $P = NP$. Also, we present an approximation algorithm for an special case of this problem.

Keywords: Minimum Spanning Tree, Color-spanning Set, Uncertainty, NP-hardness, Approximation Algorithm.

1 Introduction

In addition to the theoretical aspect of multi-colored problems, they have real world applications such as modeling uncertain data, e.g., an uncertain point can be modeled by a discrete set of candidates with a special color. In this setting, each of the candidates, but exactly one of them, may consider as the certain point. As a result, a family of challenging problems arise aim at finding a particular structurer. For example, for a given set of n points with m colors, one can select m colored points (exactly one point from each colored set) such that the minimum spanning tree obtained from them is maximized. We call this problem Max-MST and show that it is NP-hard even for a simple case, that is, each colored set contains at most two points with coordinates $p = (x, y)$ and $q = (x, y + 1)$, or $p = (x, y)$ and $q = (x + 1, y)$, for some x and y. We prove the NP-hardness of this simple problem, and present an approximation algorithm for its general cases when the instances are *well-separated*.

Related work. For a set of n points with m colors, *Planar Smallest Perimeter Convex Hull Color-spanning Set problem* ($PSPCHCS$) is finding m points with different colors which convex hull of them is minimized. It was proved that the PSPCHCS problem is NP-complete and two efficient constant factor approximation algorithms were proposed to solve it [7]. *Maximum Diameter Color-spanning Set* ($MaxDCS$) problem is finding m distinct color points which diameter of them is maximized. An $O(n^{1+\epsilon})$ time algorithm was proposed to solve it, where ϵ is an arbitrary small positive constant [7]. Also, an $O(n \log n)$ time algorithm was presented for MaxDCS [3]. The *Largest Closest Pair Color-spanning Set problem* ($LCPCS$) also was studied in [7]. This problem is finding the m different color points such that the distance between the closest pair of them is maximized. It is proved that the LCPCS problem is NP-complete even in one dimension [7]. The $Minimum\ Diameter\ Color-spanning\ Set$ ($MinDCS$) problem is finding m points with distinct colors such that the diameter of them is minimized. In [4], it was shown this problem is NP-hard for any L_p metric, except L_1 and L_∞ which admit polynomial time algorithms. In d dimensions, Ghodsi et al. [5] presented a $(1 + \epsilon)$-approximation algorithm in $O(2^{\frac{1}{\epsilon^d}}.\epsilon^{-2d}.n^3)$ time for the MinDCS problem. Kazemi et al. [9] presented also a $(1 + \epsilon)$-approximation algorithm and improved the running time to $2^{O(\lambda log \lambda) \times O(nlogn)}$. Further, the problem of *Smallest Color-Spanning Ball* ($SCSB$), which is finding the smallest ball containing at least one point of each color, was studied by Khanteimouri et al. [10]. They presented a 3-approximation and a $(1+\epsilon)$-approximation algorithm for solving SCSB problem.

Two other problems related to the Minimum Spanning Tree (MST) under uncertainty, are *Planar Smallest Minimum Spanning Tree Color spanning Set* ($PSMSTCS$) and *Planar Largest Minimum Spanning Tree Color-spanning Set* ($PLMSTCS$). In these problems, the goal is finding m points with distinct colors such that their MST is minimized in PSMSTCS problem and is maximized in PLMSTCS problem. The both problems are NP-complete [7]. The similar problem to PSMSTCS is *Generalized Minimum Spanning Tree* ($GMST$) problem. The GMST problem has two variations. Let n points are clustered in k clusters. First variation of GMST, is finding an MST consisting of at least one point in each cluster while second variation is finding exactly one point in each cluster. The first problem was proved NP-complete including the case where each cluster contains three points [6]. However, the later one is the same as PSMSTCS problem. It was proved that this variation of the GMST problem is NP-complete even if every cluster contains

*ataei.homa@gmail.com
†Institute for Advanced Studies in Basic Sciences (IASBS), Zanjan, Iran, mdmonfared@iasbs.ac.ir
‡denatayebi@yahoo.com

two points with equal y coordinate and also did not have an FPTAS unless $P = NP$ [8].

In addition to the discrete regions for modeling uncertainty, continuous regions have been also applied for this purpose. The *Minimum Spanning Tree with Neighborhood* ($MSTN$) problem was introduced in 2007 [13]. In this problem given a set of regions and the goal is placing a point on each region such that MST obtained from them is minimized. Yang et al. [13] studied the MSTN problem when the regions of uncertainty are a set of disks. They presented two approximation algorithms, two lower bounds and a PTAS for this problem. Löffler and van Kreveld proved that the MSTN problem is NP-hard where the regions are disks or squares [12]. Dorrigive et al. [2] proved the NP-hardness of the problem for disjoint disks. They also introduced the *Max-MSTN* problem which its goal is placing a point on each region whose MST is maximized. They proved that the Max-MSTN problem where neighborhoods are disjoint disks is also NP-hard, and proposed a parameterized approximation algorithm for the Max-MSTN problem.

2 NP-hardness of the PLMSTCS-2 problem

We are given a set of n points with m colors and the goal is selecting m points with distinct colors such that their MST is maximized. In this section, we study this problem where there are at most two points with the same color whose coordinates are $p = (x, y)$ and $q = (x, y+1)$, or $p = (x, y)$ and $q = (x+1, y)$, for some x or y. We denote this problem by *PLMSTCS-2* and prove its NP-hardness by a reduction from planar 3SAT.

Planar 3SAT is a special 3SAT variation whose corresponding graph is planar. This graph is constructed as follows. Correspond to each variable and each clause in the 3SAT instance, there is a node in this graph –calle the *variable node* or *clause node*. There is also an edge between a variable node and a clauses node if the corresponding variable appears in the corresponding clause. Further, all variable nodes are connected by a path. Precisely, for a 3SAT instance, let $C = \{c_1, c_2, ..., c_m\}$ be the set of clauses and $V = \{v_1, v_2, ..., v_n\}$ be the set of variables. The corresponding graph $G = (U, E)$ is constructed as follows:

$$U = C \cup V, \tag{1}$$

and

$$E = E_1 \cup E_2, \tag{2}$$

where E_1 and E_2 are:

$$E_1 = \{(c_i, v_j) \mid v_j \in c_i \ \ or \ \ \overline{v}_j \in c_i\} \tag{3}$$

and

$$E_2 = \{(v_j, v_{j+1}) \mid 1 \leq j < n\} \cup \{(v_n, v_1)\}. \tag{4}$$

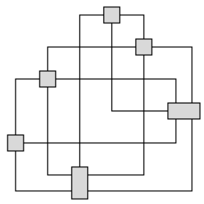


Figure 1: An example of the orthogonally drawing [1]

The set of all edges in E_2 is called *spinal path* [2]. It is proved that the planar 3SAT problem is NP-hard [11].

Theorem 1 *The PLMSTCS-2 problem is NP-hard and it does not admit an FPTAS unless $P = NP$.*

Proof. We prove this theorem by a reduction from the planar 3SAT problem. Let ϕ be an instance of the planar 3SAT problem. We design two types of gadgets –called *variable gadget* and *clause gadget*. We convert ϕ to an instance of PLMSTCS-2 by replacing the nodes with these gadgets. Since the number of colors used in this reduction should be polynomially bounded in the size of ϕ, we use a special drawing graph, called *Orthogonally Drawing* [1], to achieve this bound. In orthogonally drawing, each node is considered as a box and each edge is considered as a sequence of vertical and horizontal line segments. Figure 1 shows an example.

Theorem 2 *[1, Theorem 4] Let H be a simple graph without nodes of degree ≤ 1, where n is the number of nodes and m is the number of edges. Then H has an orthogonally drawing in an $(\frac{m+n}{2} \times \frac{m+n}{2})$-grid with one bend per edge. The box size of each node v is at most $\frac{deg(v)}{2} \times \frac{deg(v)}{2}$. It can be found in $O(m)$ time.*

In Theorem 2, $deg(v)$ is the degree of node v. The bounds are presented in Theorem 2 are established for planar triconnected graphs [1]. As planar 3SAT graph is at most triconnected, we can convert it to the orthogonally drawing in polynomially time bounded in the size of ϕ.

Now we explain variable and clause gadgets.

2.1 Variable Gadgets

We design a gadget for each variable of ϕ. This gadget is constructed by some points with k colors such that there is exactly two points with the same color. k is an even number where $4 \leq k \leq 6c - 2$ and c is the number of clauses in which variable exists. Suppose that we construct a gadget for a variable such as x_i. We consider a structure shown in Figure 2. This structure is part of the variable gadget and we called it *StructureA*.

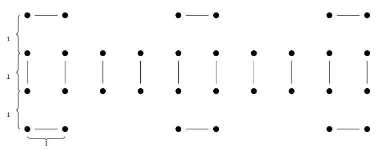
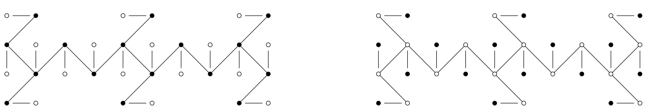


Figure 2: Structure A. The segment between two points illustrates them have the same color. That means there is 16 colors in this structure

Lemma 3 *The PLMSTCS-2 with k colors for the structure A admits two optimal solutions with weight of $\sqrt{2}(k-1)$ (See Figure 3). Further, the weight of MST for any other (non-optimal) solution is at least 0.41 less than the optimal solutions.*

Proof : See the Appendix.

Now we should design some points for connection of spinal path to variable gadgets. These points should be located such that do not affect on the selection of points in the optimal solutions. So, we locate these points in two sides of the gadget such that have equal distance from the nearest top and bottom points with the same color. These points are shown in Figure 4. Also, we should design some points for connection of the variable gadget to the edges come from the clauses. According to Lemma 3, there are two optimal solutions for structure A. One of the optimal solutions leads to select the solid points and the other one leads to select the hollow points. We called these solutions the *Solid Solution* and the *Hollow Solution*, and consider them to be correspond with the variable x_i is TRUE and FALSE, respectively. For each clause containing x_i, we set a point at the distance of 2 units to one of the solid points in the direction of y. Also, for each clause which contains $\overline{x}_i$, we set a point at the distance of 2 units to one of the hollow points in the direction of y. Figure 4 shows a variable gadget which is correspond with a variable x_i appeared in three clauses such that x_i appears in one clause and $\overline{x}_i$ appears in two clauses.



(a) Solid Solution. An optimal solution of the PLMSTCS-2 problem for structure A.

(b) Hollow Solution. An optimal solution of the PLMSTCS-2 problem for structure A.

Figure 3: Optimal solutions of the structure A. The solid and hollow points which is connected to each other have the same color

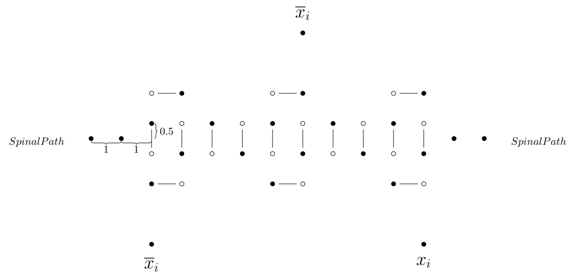


Figure 4: An example of the variable gadget

2.2 Clause Gadgets

A clause gadget is constructed by three sequence of points in the length of the edges which meet in a point. Figure 5a illustrates a clause node in the orthogonally drawing and figure 5b illustrates the clause gadget replaced with the clause node shown in Figure 5a.

2.3 Reduction

We designed two gadgets correspond with the variables and clauses. If they are replaced with the variables and clauses of the planar 3SAT graph, we have a PLMSTCS-2 instance. We scale up the orthogonally drawing of the planar 3SAT graph with a proper constant factor, e.g., 2, and then replace the graph nodes with the gadgets. The graph edges also should be exchange with a sequence of points which have unit distances along the edges. Maximum size of the variable gadget is $(4(c-1)+3)\times 7$ which is equal to $(4(deg(v)-9))\times 7$ and it is polynomially bounded in the size of the box in the orthogonally drawing. Because of the size of orthogonally drawing is at most $(\frac{m+n}{2} \times \frac{m+n}{2})$, the number of fixed points used in this reduction is polynomially bounded in the size of ϕ.

We convert every Planar 3SAT instance to a PLMSTCS-2 instance in polynomially time. Now we show that every PLMSTCS-2 solution determines whether the planar 3SAT problem has a TRUE assignment or not. We have:

$$W_T = W_E + W_G. \tag{5}$$

Where W_T is the total weight of the MST in the optimal solution of the PLMSTCS-2, W_E is the weight of

(a) A clause node

(b) A clause gadget

Figure 5: Replacing a clause node in the orthogonally drawing with a clause gadget

the MST obtained from the fixed points and W_G is the weight of the MST obtained from variable gadgets. The fixed points which is used in this reduction have a MST with a unique constant weight, so W_E has a constant value and it is enough to consider W_G. If we select either solid points or hollow points in all the variable gadgets,

$$W_G = (R - n)\sqrt{2} + 2m, \tag{6}$$

where R is the number of colors, n is the number of variables and m is the number of clauses. If W_G is equal to equation 6, then each variable gadget connects to either the clauses in which variable appears or the clauses in which variable negation appears. This means there exists a TRUE assignment in the planar 3SAT problem. If W_G is less than the equation 6, there exists at least one variable that connects to the clause in which variable appears and a clause in which variable negation appears. This means does not exist a TRUE assignment for the corresponding planar 3SAT instance.

Now we show that the PLMSTCS-2 problem does not admits an FPTAS unless $P = NP$. Suppose that there is an FPTAS for the PLMSTCS-2. Consider an instance of the planar 3SAT problem and construct the corresponding instance of the PLMSTCS-2 and compute W_T. If we set $\epsilon \leq \frac{0.41}{W_T}$, using a $(1-\epsilon)-$solution for the PLMSTCS-2, it is possible to decide whether one can exist a TRUE assignment for the planar 3SAT or not. So, PLMSTCS-2 problem does not have an FPTAS unless $P = NP$. □

3 Approximation Algorithm

In this section we present a $\frac{1}{2}$-approximation algorithm for a special case of the PLMSTCS problem. Given n points with m colors such that the minimum distance between the points with different colors is twice greater than that of the maximum distance between the points with the same color. This means the points of each color are separated from the points of other colors. Let denote this restricted problem by *well-separated PLMSTCS* problem.

Let c_i be the point from color i, whose distance from the furthest point with color i is minimum. Consider $C = \{c_1, c_2, ..., c_m\}$ as a solution for PLMSTCS.

Theorem 4 *The solution $C = \{c_1, c_2, ..., c_m\}$ described above is a $\frac{1}{2}$–approximation solution for PLMSTCS problem.*

Proof. See the Appendix.

4 Conclusion

In this paper we study the problem of *Planar Largest Minimum Spanning Tree Color-spanning Set* ($PLMSTCS$). We prove that this problem is NP-hard and dose not have an FPTAS even if there are two points from each color such that their distance is unit and have the same horizontal or vertical coordinates. We guess this variation of the PLMSTCS problem is the simplest case of this problem which is NP-hard. We also present a $\frac{1}{2}$-approximation algorithm for an special case of PLMSTCS that the points with different colors are well-separable.

References

[1] T. C. Biedl and M. Kaufmann. Area-efficient static and incremental graph drawings. In *European Symposium on Algorithms*, pages 37–52. Springer, 1997.

[2] R. Dorrigiv, R. Fraser, M. He, S. Kamali, A. Kawamura, A. López-Ortiz, and D. Seco. On minimum-and maximum-weight minimum spanning trees with neighborhoods. *Theory of Computing Systems*, 56(1):220–250, 2015.

[3] C.-L. Fan, J. Luo, W.-C. Wang, F.-R. Zhong, and B. Zhu. On some proximity problems of colored sets. *Journal of Computer Science and Technology*, 29(5):879–886, 2014.

[4] R. Fleischer and X. Xu. Computing minimum diameter color-spanning sets. In *International Workshop on Frontiers in Algorithmics*, pages 285–292. Springer, 2010.

[5] M. Ghodsi, H. Homapour, and M. Seddighin. Approximate minimum diameter. In *International Computing and Combinatorics Conference*, pages 237–249. Springer, 2017.

[6] E. Ihler, G. Reich, and P. Widmayer. On shortest networks for classes of points in the plane. In *Workshop on Computational Geometry*, pages 103–111. Springer, 1991.

[7] W. Ju, C. Fan, J. Luo, B. Zhu, and O. Daescu. On some geometric problems of color-spanning sets. *Journal of Combinatorial Optimization*, 26(2):266–283, 2013.

[8] H. A. Kachooei, M. Davoodi, and D. Tayebi. On the generalized minimum spanning tree in the euclidean plane. *1st Iranian Conference on Computational Geometry*, 2018.

[9] M. R. Kazemi, A. Mohades, and P. Khanteimouri. Approximation algorithms for color spanning diameter. *Information Processing Letters*, 135:53–56, 2018.

[10] P. Khanteimouri, A. Mohades, M. A. Abam, and M. R. Kazemi. Efficiently approximating color-spanning balls. *Theoretical Computer Science*, 634:120–126, 2016.

[11] D. Lichtenstein. Planar formulae and their uses. *SIAM journal on computing*, 11(2):329–343, 1982.

[12] M. Löffler and M. van Kreveld. Largest and smallest convex hulls for imprecise points. *Algorithmica*, 56(2):235, 2010.

[13] Y. Yang, M. Lin, J. Xu, and Y. Xie. Minimum spanning tree with neighborhoods. In *International Conference on Algorithmic Applications in Management*, pages 306–316. Springer, 2007.

Appendix

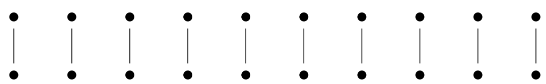
Lemma 3 *The PLMSTCS-2 with k colors for the structure A admits two optimal solutions with weight of $\sqrt{2}(k-1)$ (See Figure 3). Further, the weight of MST for any other (non-optimal) solution is at least 0.41 less than the optimal solutions.*

Proof. In Figures 3a and 3b, two selections of the points which lead to the optimal solutions are shown. First, we consider the structure B which is shown in Figure 6a. Consider the symbol U which is equivalent to the selection of the top point between two same color points, and the symbol D which is equivalent to the selection of the bottom point between two same color points.

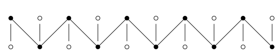
Consider a sequence of U and D symbols for each selection of the points for the structure B. We claim that the solution of the PLMSTCS-2 problem for B is $UDUDUD...UD$ or $DUDUDU...DU$ sequence which is shown in Figure 6b and 6c. The weight of MST in these solutions is $\sqrt{2}(p-1)$, where p is the number of colors in B. If we have UU or DD in the sequence of the optimal solution, the weight of MST is $\sqrt{2}(p-2)+1$ which is $\sqrt{2}-1$ less than $\sqrt{2}(p-1)$. So in the sequence leads to optimal solution we could not have UU or DD. Now we consider the structure C which is shown in Figure 7a. Structure C is a part of the structure A and repeats $\lceil \frac{k}{6} \rceil$ times. Structure C has two optimal solutions can be obtained by verifying all 2^8 possible solutions. These two optimal solutions of C are shown in Figures 7b and 7c. Clearly, in all repeats of the structure C in A either solid points or hollow points are selected, otherwise, in the solution of PLMSTCS-2 problem for structure B, we have at least one UU or DD which is contradict with the optimality of the solution. So, the PLMSTCS-2 problem for structure A has two optimal solutions which are shown in Figure 3. We call the optimal solution leads to selection of the solid points as *solid solution* and the optimal solution leads to selection of the hollow points as *hollow solution*. The weight of the MST in the solid solution and hollow solution is $\sqrt{2}(k-1)$, where k is the number of colors. Also, the weight of MST in other solutions is $(\sqrt{2}-1) \approx 0.41$ less than the weight of the optimal solutions. □

Theorem 4 *The solution $C = \{c_1, c_2, ..., c_m\}$ described above is a $\frac{1}{2}$-approximation solution for well-separated PLMSTCS problem.*

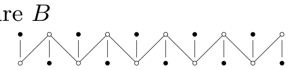
Proof. We originate three trees $Tree_{opt}$, $Tree_{center}$ and $Tree_m$ from [2]. $Tree_{center}$ is an MST obtained from C,



(a) Structure B

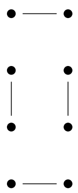


(b) The solid optimal solution of the structure B.

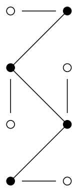


(c) The hollow optimal solution of the structure B.

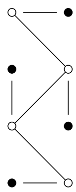
Figure 6: Structure B and its two optimal solution.



(a) Structure C



(b) The solid optimal solution of the structure C.



(c) The Hollow optimal solution of the structure C.

Figure 7: Structure C and it's two optimal solutions.

$Tree_{opt}$ is an optimal solution of PLMSTCS problem and $Tree_m$ is a spanning tree whose nodes are vertices of the $Tree_{opt}$ and it's topology is similar to the topology of the $Tree_{center}$.

Since $Tree_{opt}$ and $Tree_m$ have the same vertices and $Tree_{opt}$ is an MST of them, we have:

$$Weight(Tree_{opt}) \leq Weight(Tree_m). \tag{7}$$

Since, any edge of $Tree_{center}$ connects two centers and $Tree_m$ and $Tree_{center}$ have the same topology,

$$Weight(Tree_m) \leq 2Weight(Tree_{center}). \tag{8}$$

According to inequalities 7 and 8 :

$$Weight(Tree_{opt}) \leq 2Weight(Tree_{center}). \tag{9}$$

Consequently, C is a $\frac{1}{2}$ approximation ratio for the well-separated PLMSTCS problem. $\square$