Inapproximability proof of DSTLB and USTLB in planar graphs

Dimitri Watel, Marc-Antoine Weisser, Cédric Bentz

Inapproximability proof of DSTLB and USTLB in planar graphs

Dimitri Watel   Marc-Antoine Weisser   Cédric Bentz

February 25, 2013

This document proves the problem of finding a minimum cost Steiner Tree covering k terminals with at most p branching nodes (with outdegree greater than 1), in a directed or an undirected planar graph with n nodes, is hard to approximate within a better ratio than n, even when the parameter p is fixed.

1 Theorem

Definition 1. In a undirected (resp. directed) tree, a *branching node* is a node whose degree (resp. outdegree) is strictly greater than 2 (resp. 1).

Problem 1. min-$(*, p)$-USTLB: Given an undirected graph $G = (V, E)$ with n nodes and a non negative cost function ω on its edges, an integer k and a set $X \subset V$ of k terminals, determine, if it exists, a minimum cost tree T^* spanning all the nodes of X and containing at most p branching nodes.

Problem 2. min-$(*, p)$-DSTLB: Given a directed graph $G = (V, E)$ with n nodes and a non negative cost function ω on its arcs, a node r, an integer k and a set $X \subset V$ of k terminals, determine, if it exists, a minimum cost directed tree T^* rooted at r, spanning all the nodes of X and containing at most p branching nodes.

Theorem 1. *Let $\epsilon < 1$ be a real number. If P $\neq$ NP, the min-$(*, p)$-DSTLB and the min-$(*, p)$-USTLB problems in planar graphs with unit costs cannot be approximated within a factor of $\mathcal{N}^\epsilon$ where $\mathcal{N}$ is the number of nodes in the instance, even if there is a trivial feasible solution.*

2 Proof of the theorem

2.1 Reduction

We prove the theorem in the directed case. The proof is similar in the undirected case.

Finding a hamiltonian path starting at a specified node v in a 3-connected directed planar graph is a NP-Complete problem [1].

Let $\mathcal{I} = (G = (V, A), v)$ be an instance of the hamiltonian path problem in a 3-connected directed planar graph G. The 3-connected property is used in Section 2.3. We construct a min-$(*, p)$-DSTLB instance $\mathcal{I}'_v = (G'_v, r, X, \omega)$ where G'_v is a directed planar graph.

The main idea is that G'_v is divided in three parts. An example is shown in Figure 1. Firstly, a graph $G' = (V' = V \cup W, A')$ built from G where each arc of A is divided in two or more arcs. Secondly, a binary tree $\mathcal{B}$ rooted at r with p branching nodes and $p + 1$ leaves. We link one of the leaves of $\mathcal{B}$ to v with an arc a_v. We define X as the leaves of $\mathcal{B}$ and V. Finally, a graph H and an integer h which ensures the three following properties:

Property 1. *Let n, $n_{G'}$ and n_H be the number of nodes in G, G' and H. $n_{G'} - n$ is no more than n^3 and $n_{G'} - n + n_H$ is no more than $4 \cdot n^3 \cdot h$.*

Property 2. *There exists an elementary path P in $G' \cup H$ going through each node of G starting at v.*

Property 3. *Any elementary path in $G' \cup H$ going through each node of G starting at v using a node of H not as endpoint contains at least h nodes of H.*

Property 2 ensures the existence of a feasible solution. Properties 1, and 3 ensure an inapproximability gap, described in section 2.2. If h is long enough, Property 3 ensures that any node of H will not be allowed in any approximated solution. We will fix G', H and the value of the parameter h later.

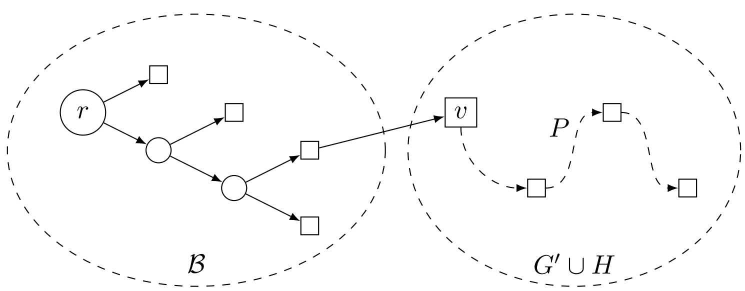

Figure 1: Example of reduction from a graph G with 4 nodes, and $p = 3$. Nodes of W $(= V' \setminus V)$ and H, and arcs of G' and H do not appear on that figure.

The number of nodes $\mathcal{N}$ in G'_v is $n_{G'} + n_H + p + (p + 1)$.

2.2 Inapproximability gap

In this part, we fix the parameter h and show the approximability hardness of $(*, p)$-DSTLB.

Let T^* be an optimal solution of $\mathcal{I}'_v$. It exists because $\mathcal{B} \cup P \cup a_v$ is a feasible solution by Property 2. Let $\epsilon < 1$, and suppose it exists a polynomial $\mathcal{N}^\epsilon$-approximation algorithm for min-$(*, p)$-DSTLB in a planar graph. We will show that in that case, we could use this algorithm to decide whether G has a hamiltonian path starting at v.

If there exists a hamiltonian path in $\mathcal{I}$ starting at v, T^* contains at most $n_{G'} + 2p + 1$ nodes (the $n_{G'}$ nodes of G' and the $2p + 1$ nodes of $\mathcal{B}$), thus it contains at most $n_{G'} + 2p$ arcs. So the approximate solution has a cost $c_{\text{YES}} \leq (n_{G'} + 2p) \cdot \mathcal{N}^{\epsilon}$.

We now discuss the case where there is no hamiltonian path starting at v in $\mathcal{I}$. Then, without H, we cannot build an elementary path going through each node of G.

Lemma 1. *Any feasible solution of $\mathcal{I}'_v$ contains an elementary path going through each node of G starting at v.*

Proof. Let T be a feasible solution. T covers every leaf of $\mathcal{B}$, as a consequence it covers $\mathcal{B}$ entirely. Because $\mathcal{B}$ contains p branching nodes, all other terminals are covered with elementary paths connected to $\mathcal{B}$. T covers every nodes of G and a_v is the only arc linking $\mathcal{B}$ to a node G. So T contains an elementary path going through each node of G starting at v. □

By Lemma 1, without H, we cannot build a feasible solution of $\mathcal{I}'_v$. So the approximate solution uses at least one node of H. On of those node is not an endpoint. Indeed, in this case, we can remove them to get a hamiltonian path in G. By Property 3, it uses at least h nodes of H. So it has a cost $c_{\text{NO}} > h$.

If $c_{\text{NO}} > h > c_{\text{YES}}$, then the approximation algorithm can decide whether there is a hamiltonian path starting at v.

Lemma 2. *Let h satisfies $h = 5^{\frac{\epsilon}{1-\epsilon}}(2n^3+2\cdot p+1)^{\frac{1+\epsilon}{1-\epsilon}}+1$. Then $c_{\text{NO}} > h > c_{\text{YES}}$.*

Proof. Notice that $h > 1$ for all $\epsilon < 1$ and $n \geq 1$. Lines 9 and 13 are proven by Property 1.

$$h > 5^{\frac{\epsilon}{1-\epsilon}}(2n^3+2p+1)^{\frac{1+\epsilon}{1-\epsilon}} \quad (1)$$
$$h^{1-\epsilon} > 5^{\epsilon}(2n^3+2p+1)^{1+\epsilon} \quad (2)$$
$$h > 5^{\epsilon}(2n^3+2p+1)^{1+\epsilon}h^{\epsilon} \quad (3)$$
$$h > (2n^3+2p+1)^{1+\epsilon}(5h)^{\epsilon} \quad (4)$$
$$h > (2n^3+2p+1)^{1+\epsilon}(1+4h)^{\epsilon} \quad (5)$$
$$h > (2n^3+2p+1)\cdot(1+4h)^{\epsilon}(2n^3+2p+1)^{\epsilon} \quad (6)$$
$$h > (2n^3+2p+1)\cdot(1+4h)^{\epsilon}(n^3+2p+1)^{\epsilon} \quad (7)$$
$$h > (n^3+n+2p+1)\cdot(1+4h)^{\epsilon}(n^3+2p+1)^{\epsilon} \quad (8)$$
$$h > (n_{G'}+2p)\cdot(1+4h)^{\epsilon}(n^3+2p+1)^{\epsilon} \quad (9)$$
$$h > (n_{G'}+2p)\cdot((n^3+2p+1)+4\cdot(n^3+2p+1)\cdot h)^{\epsilon} \quad (10)$$
$$h > (n_{G'}+2p)\cdot((n+2p+1)+4\cdot(n^3+2p+1)\cdot h)^{\epsilon} \quad (11)$$
$$h > (n_{G'}+2p)\cdot((n+2p+1)+4\cdot n^3\cdot h)^{\epsilon} \quad (12)$$
$$h > (n_{G'}+2p)\cdot((n+2p+1)+n_{G'}+n_H-n)^{\epsilon} \quad (13)$$
$$c_{\text{NO}} > h > c_{\text{YES}} \quad (14)$$

□

As a consequence, if $P \neq NP$, such an algorithm does not exist.

2.3 Existence of G' and H

In this section, we explain how to build the graphs G' and H.

2.3.1 Construction of $G' = (V \cup W, A')$

G' is built from G where each arc of a is divided into several arcs of A' and nodes of W.

G is a 3-connected planar graph. As a consequence, we can embed it in $\mathbb{R}^2$ as a convex polygon such that v is on the outer face of G, using for instance the technique of [2]. For a node $w \in V$, we define its coordinates as x_w and y_w.

Lemma 3. *It exists an angle α such that the rotation $r_\alpha(G)$, of angle α and center v, rotates G so that each node $w \in V$ has a unique x-coordinate x_w with $x_v \leq x_w$ (v is 'on the left').*

Proof. Let α_m and α_M be two angles in $[0; 2\pi]$ such that for each $\alpha \in [\alpha_m; \alpha_M]$, $r_\alpha(G)$ places v on the left.

If there is no angle where, after G rotates, each node $w \in V$ has a unique x-coordinate x_w, for each $\alpha \in [\alpha_m; \alpha_M]$, there are two nodes (u, w) with $x_u = x_w$ and $y_u < y_w$. There are at most n^2 such couples. Let α_i, $i \in [1..(n^2+1)]$, be distinct angles in $[\alpha_m; \alpha_M]$, there are two distinct angles for which the same couple of nodes (u, w) verified, after G rotates, $x_u = x_w$ and $y_u < y_w$, which implies a contradiction. □

We then sort the list of nodes v_i by its x coordinate : $x_v = x_{v_1} < x_{v_2} < x_{v_3} < ... < x_{v_n}$.

We define D_i for $i \in [2..n]$ as the vertical strait lines of abscissa $x_i = \frac{x_{v_{i-1}} + x_{v_i}}{2}$. For each arc $a = (t, u)$ of G crossing a line D_i, we add a node w to W at the intersection of a and D_i and replace a in A' by the two arcs (t, w) and (w, u). An example is shown in Figure 2.

As no new arc cross, G' is planar.

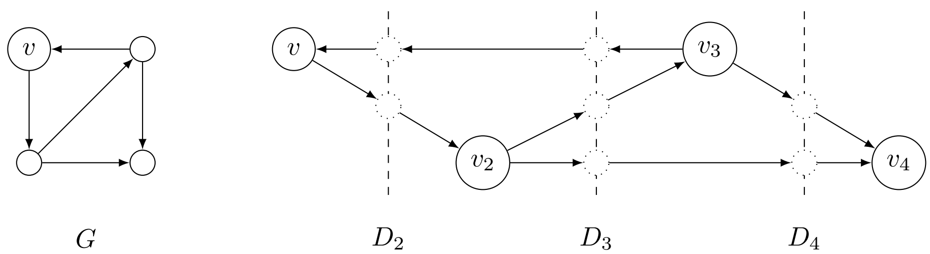

Figure 2: Example of graph $G' = (V \cup W, A')$ built from a graph G with 4 nodes. W is the set of dashed nodes.

2.3.2 Construction of H

We first prove three intermediate lemmas :

Lemma 4. *Any arc of G' starting at a vertical strait line D_i goes to the left, or goes above, below, from or to v_i.*

Proof. Let a be an arc of G' crossing a vertical strait line D_i at a node u. If a goes to the left, the lemma is verified. Else, if a do not go above, below, from and to v_i, there is a node $t \in V'$ with $a = (u, t)$ or $a = (t, u)$ and $x_t \in [x_i, x_{v_i}[$. If $t \in V$, by definition of D_i, $x_i > x_t$ which implies a contradiction. If $t \in W$, there is a strait line D_j with $x_t = x_j \in [x_i, x_{v_i}[$ which also implies a contradiction. □

We can similarly prove the following lemma :

Lemma 5. *Any arc of G' going above, below, from or to v_i goes to the right of v_i or cross D_i.*

Lemma 6. *For each node $v_i \in V$, $i \in [2..n]$, we can add to H a node $v_{i,l}$ on D_i and an arc $(v_{i,l}, v_i)$ such that the graph $G' \cup H$ remains planar.*

Proof. Let a_m and a_M be respectively the lowest arc of G' going above v_i and the highest arc going below or to v_i, going from or to the left of v_i. An example is shown in figure 3.

If a_m and a_M do not exist, by Lemma 4, there is no arc crossing D_i going to or from the right (the graph is then disconnected). We can add $v_{i,l}$ on D_i anywhere there is no node of W.

If only a_m exists, the arc cross D_i at a node t_m by Lemma 5. We can add $v_{i,l}$ on D_i anywhere below t_m where there is no node of W.

If only a_M exists, the arc cross D_i at a node t_M by Lemma 5. We can add $v_{i,l}$ on D_i anywhere above t_M where there is no node of W.

If a_m and a_M exists, they cannot cross at a point of abscissa $x \in [x_i; x_{v_i}]$. If they do, either G' is not planar which is not true, or G' contains a node $t \in [x_i; x_{v_i}]$. Like in the proof of lemma 1, this would imply a contradiction. So we can add $v_{i,l}$ on D_i anywhere above t_M and below t_m where there is no node of W. □

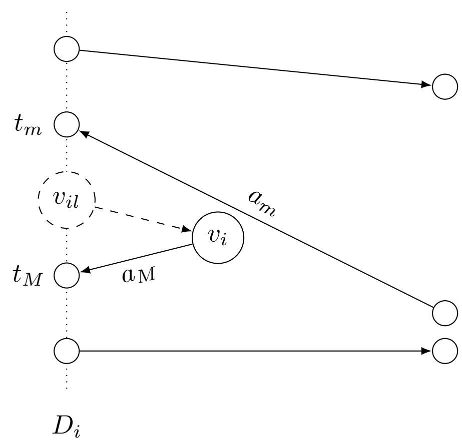

Figure 3: Example of insertion of v_{il}

Similarly, for each node $v_i \in V$, $i \in [2..n]$we can add to H a node $v_{i,r}$ on D_{i+1} and an arc $(v_i, v_{i,r})$ such that the graph $G' \cup H$ remains planar.

Finally, for $i \in [2..n]$, we sort the nodes of abscissa x_i by increasing y-coordinate (those nodes are nodes of G' or nodes of H). For each couple (u, t) of consecutive nodes we add to H a path of h nodes going from u to t and a path from t to u through the same h nodes. An example is shown in Figure 4.

Lemma 7. *G', H and h verify Properties 1, 2 and 3.*

Proof. $n'_G - n$ and n_H are the number of arcs in W and H, in other words, the nodes of all the lines D_i. For each vertical line D_i, we create at most m nodes

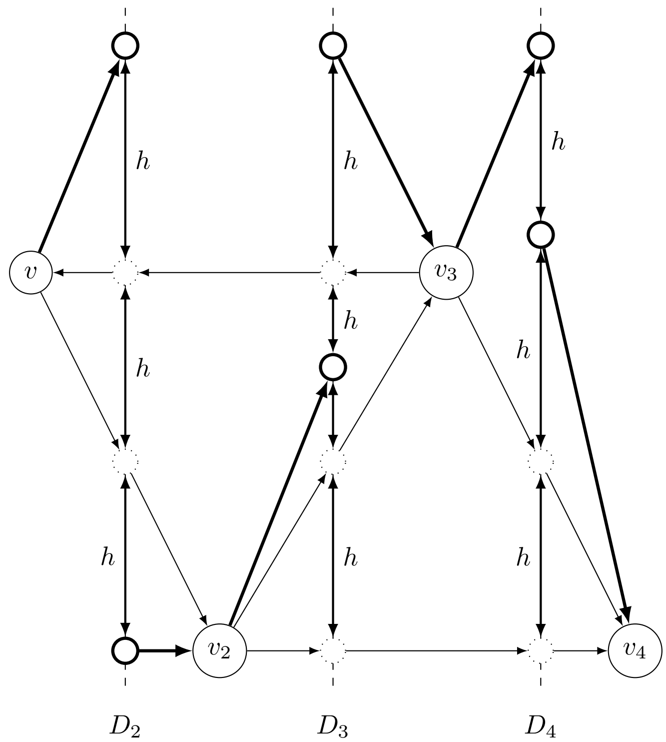

Figure 4: Example of graph $G' \cup H$ built from a graph G with 4 nodes. Thick nodes are v_{ir} and v_{il}. Dashed nodes are W. Each vertical *arc* is actually a path with h nodes.

of G' and $2 + h \cdot (1 + m)$ nodes of H.

$$n'_G \leq (n-1) \cdot m \quad (15)$$

$$n'_G \leq n^3 \quad (16)$$

$$n'_G + n_H - n \leq (n-1) \cdot ((1+m)(h+1) + 1) \quad (17)$$

$$n'_G + n_H - n \leq n(1+m)(h+1) + (n-1) - (1+m)(h+1) \quad (18)$$

.

Because $n \leq m$ in a connected graph and $h \geq 0$, we now that $(n-1) < (1+m)(h+1)$. Thus $n'_G + n_H - n \leq n(1+m)(h+1) < n \cdot (2m) \cdot (2h) < 4n^3h$. Property 1 is verified. If the graph G is not connected, there is no solution to the hamiltonian path problem.

The path P starting at v, going to $v_{1,r}$, from $v_{i-1,r}$ to $v_{i,l}$ through D_i and to v_i for $i \in [2..n]$ goes through each node of G. Property 2 is verified.

Let P be an elementary path going through every nodes of G and one node of H not as endpoint. As only the nodes $v_{i-1,r}$ and $v_{i,l}$, $i \in [2..n]$ are linked to a node of G. If P contains a node of H, it exists a node t and $i \in [2..n]$ such that $t = v_{i-1,r}$ or $t = v_{i,l}$ is in P. As t is linked to only one node not in D_i, P goes out of D_i (or enters D_i) through an other node of D_i and P contains at least h nodes of D_i. Property 3 is verified. □

References

[1] Garey, M. R., Johnson, D. S., and Tarjan, R. E. The planar hamiltonian circuit problem is np-complete. *SIAM Journal on Computing 5*, 4 (1976), 704–714.

[2] Plestenjak, B. An algorithm for drawing planar graphs. *Software-Practice and Experience 29*, 11 (1999), 973–984.